Controlling Payload Heterogeneity in Lipid Nanoparticles for RNA-Based Therapeutics

Turash Haque Pial[1*], Sixuan Li[2], Jinghan Lin[1,3,5], Tza-Huei Wang[2,3,4], Hai-Quan Mao[1,3,4,5], Tine Curk[1,3*]

[1]Department of Materials Science and Engineering, Johns Hopkins University, Baltimore

[2]Department of Mechanical Engineering, Johns Hopkins University, Baltimore

[3]Institute for NanoBioTechnology, Johns Hopkins University, Baltimore

[4]Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore

[5]Translational Tissue Engineering Center, Johns Hopkins University School of Medicine, Baltimore

Abstract:

Lipid nanoparticles (LNPs) are a leading platform for nucleic acid delivery, yet conventional assembly by mixing lipids with RNA yields particles with heterogeneous, bimodal payload distribution. Our transfection experiments demonstrate that heterogeneous siRNA distribution within LNPs significantly reduces gene knockdown efficiency. To systematically investigate the origin and extent of payload heterogeneity, we integrate coarse-grained molecular dynamics and kinetic Monte Carlo simulations with single-particle characterization via cylindrical illumination confocal spectroscopy and machine learning analysis. We find that the balance between RNA diffusion kinetics and lipid self-assembly dynamics is the dominant driver of payload heterogeneity. Leveraging this mechanistic insight, we show that (i) finely controlled turbulent mixing minimizes payload variance and increases the uniformity of RNA distribution without altering LNP size, and (ii) systematic adjustment of salt concentration and PEG-lipid content tunes RNA loading in a volume-dependent manner. Together, these results elucidate the self-assembly landscape of RNA-LNPs and provide actionable design principles for crafting more uniform, potent, and safer LNP-based nucleic acid therapies.

Keywords: Lipid nanoparticles, RNA therapeutics, RNA encapsulation, kinetic Monte Carlo, RNA payload distribution, single-particle characterization

*Corresponding author. Email: mpial1@jh.edu, tcurk@jhu.edu

Introduction:

Nucleic acid-based therapeutics have revolutionized disease treatment, with lipid nanoparticles (LNPs) established as an effective delivery platform.[1–6] The success of LNPs in FDA-approved vaccines, which deliver messenger RNA (mRNA), underscores their immense potential as a versatile delivery platform. Beyond vaccines, LNPs are being developed for various therapies, including small interfering RNA (siRNA)-mediated gene silencing,[7,8] mRNA-based protein replacement,[9] gene editing using mRNA and guide RNA,[10] circular RNA for long-lasting protein expression in genetic disorder treatments,[11] and DNA/RNA-based induction of immune responses against cancers and chronic infections.[12–15] Despite this progress, a key gap lies in achieving precise control over how therapeutic payloads are distributed among individual LNPs as unwanted heterogeneous payload distribution can compromise therapeutic performance.[16–19] Therefore, establishing a mechanistic understanding of payload heterogeneity and its effect on drug delivery is needed for optimizing nucleic acid-based delivery systems, and thereby de-risk and broaden the applicability of the LNP delivery platform.

RNA-LNPs are formed through a self-assembly process, where an aqueous solution of RNA is mixed with an alcoholic lipid solution containing ionizable lipids, PEGylated lipids, cholesterols, and helper lipids. Two key interactions primarily drive the self-assembly: the amphiphilic nature of the lipids, and electrostatic interactions between the negatively charged RNAs and the positively charged ionizable lipids. However, this assembly process often produces LNPs with heterogeneous compositions, including LNPs without RNA (“empty LNPs") and LNPs containing multiple RNA copies, characterized by a bimodal RNA loading distribution[20]. This payload heterogeneity can significantly impact therapeutic outcomes, leading to toxicity or reactogenicity.[16,17] A recent study has linked empty LNPs containing YSK13 lipids to liver toxicity, suggesting that the lipid component itself contributes to adverse effects.[18] This highlights the importance of controlling the dose of administered RNA-LNPs to minimize toxicity, which necessitates precise RNA distribution within the LNPs. Researchers also found that the higher mRNA loading reduces transfection potency, likely due to their deviation from the optimal lipid/mRNA ratio and mRNA-mediated bleb-like structure formation.[21] However, according to our knowledge, no study about potency is available for shorter or non-bleb forming RNAs like siRNA. Although less critical in vaccines, minimizing undesirable inflammation, improving potency, and enhancing drug delivery consistency are critical in long-term therapies such as gene delivery.

Formulating LNPs with desired physical properties, such as average size and RNA encapsulation efficiency, is straightforward, especially with techniques like microfluidics.[22–24] However, controlling and measuring payload heterogeneity remains a significant challenge. Recently, we developed Cylindrical Illumination Confocal Spectroscopy (CICS), enabling analysis of mRNA, siRNA, or other payload characteristics at the single-particle level.[20,25] This method uncovered substantial payload heterogeneity and a bimodal loading distribution, with up to 50% of LNPs being empty. While bimodal distribution is typically caused by thermodynamic phase separation, it may also stem from kinetic factors. The key question is whether thermodynamic or kinetic effects dominate. If the cause is kinetic, enhancing mixing could reduce heterogeneity. On the other hand, if the cause is thermodynamic, allowing the LNP configurations to equilibrate via enhanced mixing would likely

increase heterogeneity. Thereby, identifying the precise cause of heterogeneity is crucial for optimizing strategies to obtain the desired RNA payload distribution.

In this study, we systematically explore the origin and extent of RNA payload heterogeneity and its effect on delivery efficiency. We investigate the formation of empty LNPs and demonstrate how to control their population. To this end, we first develop advanced multiscale computational techniques, employing molecular dynamics (MD) to model the initial stage of the RNA-LNP assembly process and kinetic Monte Carlo (kMC) to model the later stages, thus overcoming the challenge of limited simulation timescales. We investigate the effect of mixing rate, salt concentration (c_{salt}), and PEGylation on the LNP size and payload distribution, and use a random-forest machine-learning method to analyze the effect of these features. Findings are validated using the CICS single

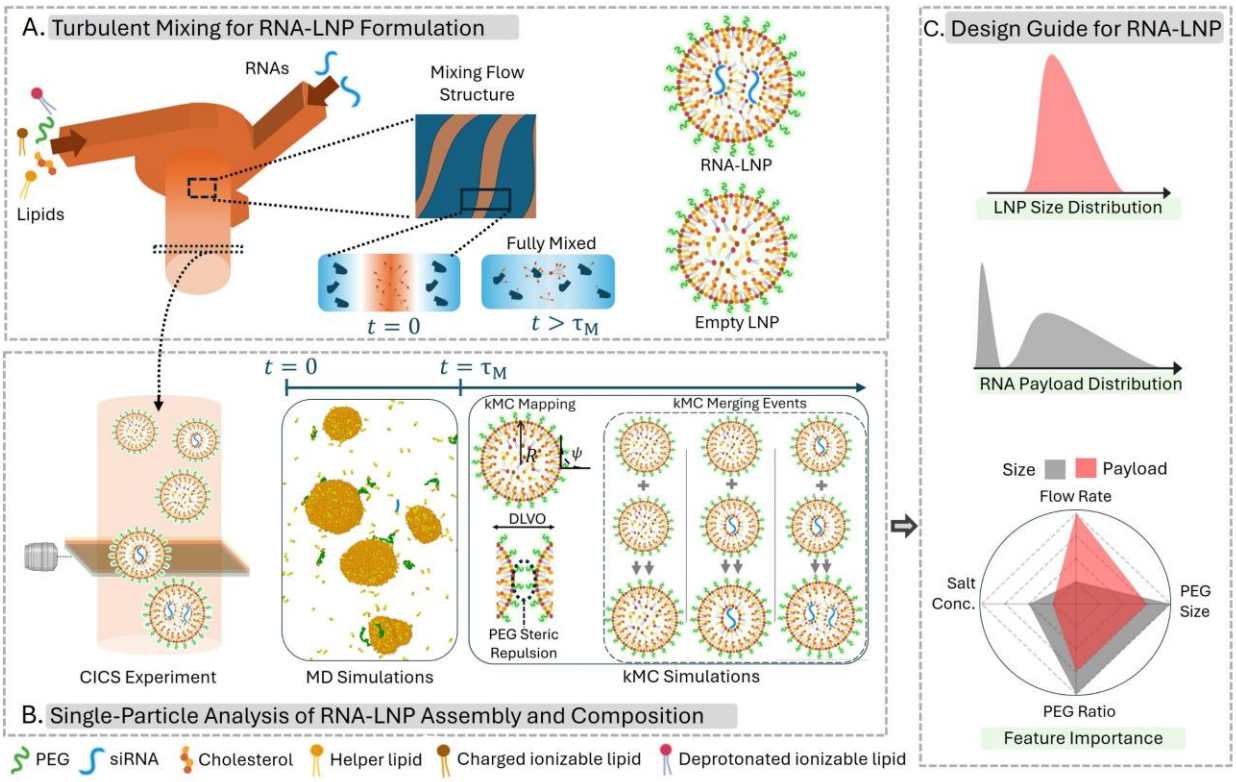

Figure 1: ***Overview of the RNA-LNP Self-Assembly Process: Experimental and Modeling Framework.*** *(A) Schematic illustrating the self-assembly process of RNA-LNPs, where lipid-ethanol and RNA-aqueous solutions are mixed using a vortex mixer (left). Inlet flows generate turbulent inter-shearing layers, keeping lipids and RNAs separate until full mixing occurs due to diffusion on timescale τ_M. This process resulted in RNA-loaded and RNA-free LNPs (empty-LNPs). (B) Cylindrical illumination confocal spectroscopy (left) is employed to analyze particle composition. Molecular dynamics (MD) simulations capture the initial stages of the assembly process (middle), whereas kinetic Monte Carlo (kMC) simulations are used for later stages (right). kMC incorporate charge regulation, Derjaguin-Landau-Verwey-Overbeek (DLVO) interactions, and PEG-PEG repulsion. Upon merging, LNPs update their composition and associated interactions. (C) kMC simulations provide detailed insights into RNA-LNP size distribution, payload distribution, and assembly dynamics. Coupling these simulations with machine learning enables the development of design rules for optimizing RNA-LNP formulations.*

nanoparticle detection experiments (Figure 1). Finally, we establish design principles to obtain a preferred LNP size and payload distribution.

Results:

We utilized a benchmark lipid formulation based on the FDA-approved siRNA drug Patisiran.[26] Typically, RNA-LNPs are formed by mixing an RNA-containing aqueous solution with a lipid-containing alcohol solution in turbulent setups such as T-junctions,[27] confined impinging jets,[28] or multi-inlet vortex mixers.[29] We used a 3-inlet vortex mixer at flow rates of 10, 20, and 30 mL/min, with a siRNA-to-lipid flow rate ratio of 3:1. After mixing and subsequent dialysis (conducted 1-hour post-mixing), the siRNA-LNPs were analyzed using CICS (at 18 hours post-mixing), where each particle passes through a detector individually. Fluorescence signals from the detector enabled differentiation among LNP populations, categorizing them into siRNA-encapsulated LNPs, empty LNPs, and free unencapsulated siRNAs (Figure 1A-1B). This experimental system served as a reference for parameter selection and validation in our modeling approach.

MD simulations reveal the origin of empty LNPs

To investigate the initial stages of the siRNA-LNP self-assembly process, we employed coarse-grained (CG) MD simulations. In turbulent mixing setups, the kinetic energy from the inlet flows breaks the flow jets into inter-shearing layers or turbulent eddies as the solutions enter the mixing chamber.[30] To replicate these inter-shearing layers in simulations, we initially divided the system into distinct siRNA and lipid layers, maintaining a volume ratio 3:1 (Figure 2A(i)). Simulations reveal that during the mixing process, most lipids quickly aggregate to form small nanoparticles before significantly interacting with siRNA (Figure 2A, 2B). This early lipid aggregation leads to the formation of primarily empty LNPs. To our knowledge, this represents the first MD simulation that captures the RNA-LNP mixing process in a manner consistent with experimental mixing conditions.

The thickness of inter-shearing layers can be controlled by adjusting the flow rate during siRNA-LNP formation.[30] At higher flow rates, the increased kinetic energy leads to more intense breakdown of the flow and smaller turbulent eddies.[31] In contrast, smaller flow rates may result in laminar flow, without any turbulent eddies. Following Ref. 30, we define the characteristic mixing length scale, l, as half the lipid layer thickness (Figure 2A). The distribution of siRNA on LNPs was analyzed at a simulation time of 100,000 τ (simulation time unit $\tau = 2$ ns, yielding a total assembly time of 0.2 ms) at varying l. The results show that thicker inter-shearing layers (correspond to slower mixing rates in experiments) lead to a higher proportion of empty LNPs compared to thinner layers (Figure 2C). To quantitatively explain this behavior, we calculate the mixing timescale, τ_{M}, for siRNA and lipids to diffuse across the mixing length scale, $\tau_{\mathrm{M}} = \frac{l^2}{2D}$ where D is the inter-diffusion constant (noting that in these simulations, τ_{M} values are about 25 µs, 57 µs, and 100 µs, respectively). Since individual siRNA molecules are much smaller than the LNPs, we approximate $D \approx D_{\mathrm{siRNA}}$ (see supporting information SI, Figure S3). We find that at τ_{M} the loading distribution for different l collapse into a master curve (Figure 2C, inset). This demonstrates that the distribution of siRNA on LNPs is primarily governed by the diffusion timescale τ_{M}, which is influenced by the thickness of the inter-shearing layers.

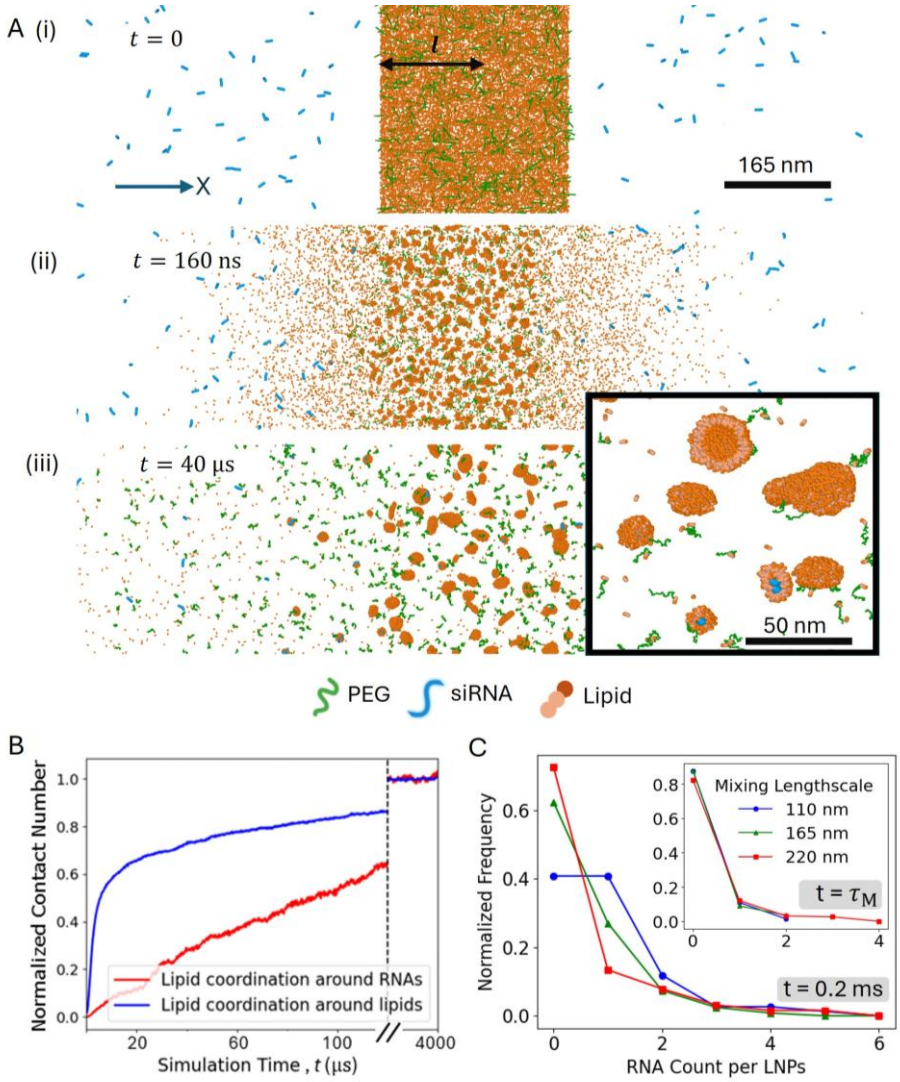

Figure 2: ***Molecular Dynamics (MD) Simulation of siRNA-LNP Assembly****. (A) CG-MD simulation depicting the initial distribution of siRNA (blue), lipids (orange), and PEG (green) at (i) t=0, (ii) t=1600 τ, and (iii) t=40000 τ (τ= 2 ns). The characteristic mixing length scale, l (165 nm for these snapshots), is shown in black. Over time, lipids begin to aggregate and form small nanoparticles. Inset shows a zoom-in of the configuration, showing empty and siRNA containing LNPs. (B) Number of RNA–lipid, and lipid–lipid contacts normalized with the coordination number. (C) RNA payload distribution for different mixing length, l, at t=$10^5\tau$ (or 0.2 ms). RNA distribution at mixing time scale $\tau_M = l^2/2D$, collapses to a master curve (inset). The τ_M values are about 25 μs, 57 μs, and 100 μs for l=110, 165, and 220 nm, respectively.*

kMC simulations and CICS experiments show the effect of mixing flow rates

Since MD simulations cannot capture the full growth dynamics of LNPs due to time scale limitations, we employ kMC simulations to model the aggregation of LNPs on timescales larger than the mixing time τ_{M}. kMC simulations enable dynamic evolution of LNP growth[32] and allow for the investigation of payload and size distributions. To initialize kMC simulations, we calculate the LNP radius R_0 at mixing time τ_{M}, using a mean-field coalescence theory of lipids (Figure 3A; see SI for derivation). We validated this approach by comparing the predicted sizes to those obtained from MD simulations. Based on MD simulation results (Figure 2C, inset), we set that at τ_{M} the fraction of empty LNPs $\theta_0 = 0.83$. In the turbulent mixing regime (flow rate $Q \geq 10\ \mathrm{ml/min}$), the characteristic mixing time depends on the flow rate as $\tau_{\mathrm{M}} = b_Q/1.75 \times Q^{-1.5}$, where $b_Q = 1.3 \times 10^3\ \mathrm{ms}$ [30,33,34] which allows us to compare experimental observations with kMC simulations (the factor 1.75 accounts for diffusivity of siRNA relative to conditions in Ref. 30). A faster flow rate results in more turbulent mixing conditions with smaller l and shorter τ_{M} (Figure 2), and thus smaller size R_0 (Figure 3A).

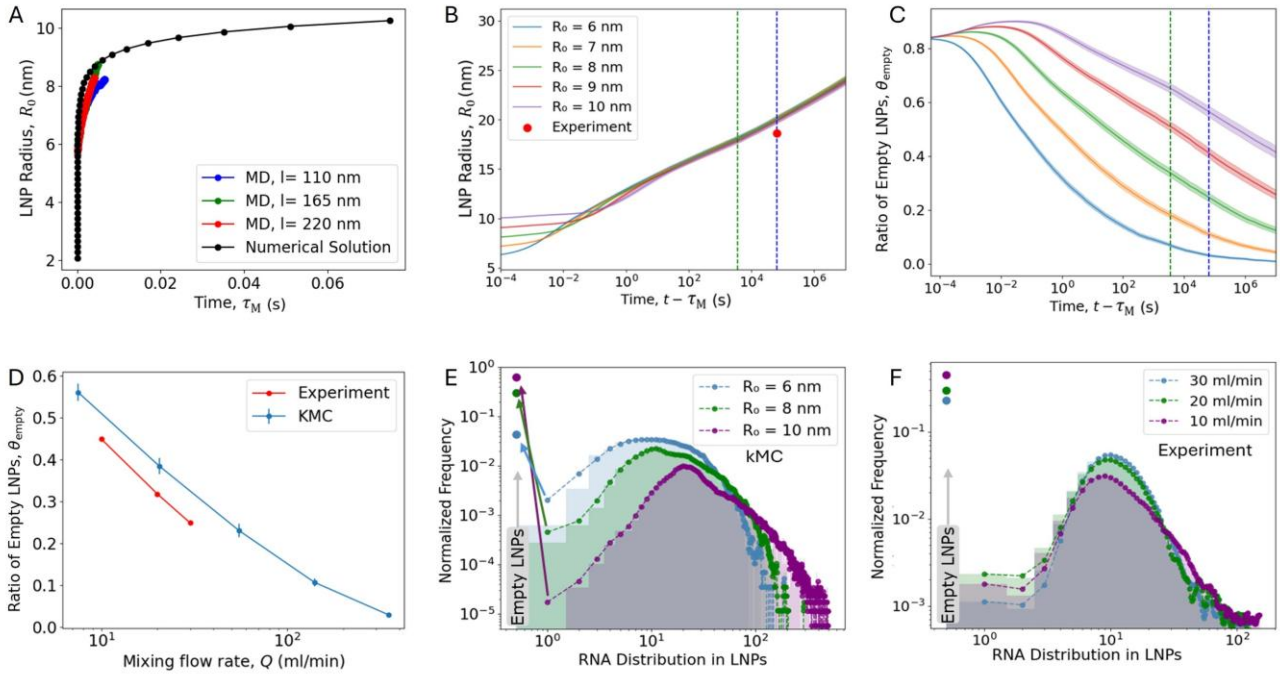

Figure 3: ***Effect of Mixing Flow Rate on LNP Size and Payload Distribution.*** *(A) Theory prediction for LNP growth and MD simulation results providing the initial LNP radius input for kMC simulations. (B) Time evolution of the average LNP radius from kMC simulations. (C) Time evolution of the ratio of empty LNPs, defined as the fraction of LNPs without siRNA. The green vertical lines denote t = 1 h (post-mixing dialysis), and the blue lines denote t = 18 h (to match experimental measurement via CICS). (D) Ratio of empty LNPs as a function of mixing flow rate Q at t =18 h. (E-F) siRNA loading distribution at t =18 hours from kMC (E) and experiment (F). In both (E) and (F), normalized frequency distributions show a bimodal profile, with empty LNPs indicated on the left. The shaded regions show the raw histogram data, while the dashed lines show a moving average for visual clarity. Arrows in (E) help to show the bimodal nature of the loading distributions. Parameters: PEG MW is fixed at 2000 Da (R_F=3.6 nm), PEG ratio 1.5% of total lipids, and the initial salt concentration c_{salt} = 0.025 M.*

143 Figure 3B illustrates the evolution of the average LNP radius *R* (note *R* is the bare LNP radius that does
144 not include the PEG chains) in the kMC simulation. Surprisingly, particles reach an average radius of
145 approximately 20 nm at 18 hours regardless of the initial size R_0. This similarity in particle size for
146 different R_0 is also observed in experiment. Dynamic light scattering (DLS) intensity-weighted data
147 reveals weak dependance of particle size on the mixing rate, with radii of 50.3 nm, 51.9 nm, and
148 52.5 nm at flow rates of $Q = 10\ \mathrm{ml/min}$, $Q = 20\ \mathrm{ml/min}$, and $Q = 30\ \mathrm{ml/min}$, respectively. This
149 weak correlation is consistent with previous studies, which demonstrated that nanoparticle size
150 decreases with increasing flow rates in laminar mixing regime but plateaus in the turbulent flow
151 regime.[30,35] Finally, according to our prior LNP experiments, DLS-derived average radii are about 2.8
152 times larger than those measured by cryo-TEM[20] because DLS is skewed towards large particles, and
153 we find that rescaling the DLS data agrees very well with the kMC simulations (Figure 3B).

154 To investigate the effects on LNP loading distribution, we analyzed the time-dependent changes in
155 the ratio of empty LNPs θ_{empty} (Figure 3C), which is defined as the fraction of LNPs that contain no
156 RNA, $\theta_{\mathrm{empty}} = N_{\mathrm{empty}}/N_{\mathrm{total}}$, with N_{total} the total number of LNPs and N_{empty} the number of LNPs
157 containing no RNA. As LNPs merged into larger particles, the ratio of empty LNPs decreased overall.
158 However, unexpectedly, the ratio increased in the early stages, which can be explained by the
159 stronger electrostatic repulsion between empty LNPs, leading to LNPs containing siRNA to merge
160 more quickly. Figure 3C demonstrates that reducing the initial LNP size significantly reduces the ratio
161 of empty LNPs. Quantitative comparison between kMC results and experimental data at different
162 mixing flow rates shows excellent agreement (Figure 3D).

163 While previous studies have shown the controllability of average nanoparticle size and drug loading
164 efficiency, the distribution of internal components has remained largely unexplored. Our findings
165 show that rapid mixing improves the uniformity/homogeneity of the internal composition within the
166 nanoparticles. Surprisingly, this implies that the bimodal loading distribution results from mixing
167 kinetics rather than thermodynamic phase separation. This is further supported by analyzing the
168 entire siRNA loading distribution within LNPs (kMC data in Figure 3E, CICS data in Figure 3F). A larger
169 initial LNP radius (smaller flow rate) results not only in more empty LNPs but also increases the
170 number of LNPs with excessive siRNA (longer distribution tail), indicating highly heterogeneous
171 encapsulation and a broader bimodal distribution. The agreement between kMC and experimental
172 data (Figure 3B, D–F) is remarkable given that the kMC model contains no fitting parameters. This
173 strongly suggests that the kMC model captures the relevant physical processes and can be used as
174 an independent tool to predict LNP assembly.

175 *A more uniform payload distribution improves delivery efficiency in vitro*

176 To investigate the impact of payload distribution uniformity on delivery efficiency, we performed in
177 vitro transfection assays using LNPs generated at flow rates of 1 mL/min (not covered in the kMC
178 calculations), 10 mL/min, and 30 mL/min. Increasing the flow rate reduced both the average siRNA
179 payload and the ratio of empty LNPs, indicating that higher flow rates yield LNP populations with
180 more uniform and efficient siRNA encapsulation (as shown in Figure 3 E, F; high flow rate leads to
181 narrower payload distribution, see also Figure 6D). Note that LNPs formed at the lowest flow rate (1

mL/min, corresponding to the laminar mixing regime) were larger in average size (62.0 nm in radius, compared to 51.0 nm and 52.5 nm for 10 ml/min and 30 ml/min, respectively), which resulted in LNPs with a higher average siRNA payload per particle.

At a high siRNA dose (100 ng), all formulations achieved near-complete knockdown of GFP expression in GFP+ cells (Figure 4A). This indicates that at saturating concentrations, the influence of payload distribution becomes negligible. However, at lower doses, the differences in knockdown efficiency diverged: LNPs prepared at 30 ml/min consistently outperformed those produced at lower flow rates. These particles, with a more uniform payload distribution and fewer empty LNPs, achieved greater GFP silencing and lower mean fluorescence intensity (MFI, Figure 4B) than the more heterogeneous populations from 1 and 10 ml/min.

These results demonstrate that improved knockdown efficiency at lower doses correlates strongly with the uniformity of siRNA payload distribution within the LNP population. At lower flow rates, the broader payload distribution negatively affected knockdown efficiency, particularly at subsaturating doses. Importantly, the encapsulation efficiency was greater than 90% across all conditions, ensuring a consistent total number of delivered siRNA across equivalent doses. Therefore, the observed differences in knockdown efficiency can be attributed exclusively to variations in payload distribution among the LNP populations rather than total siRNA content.

LNP Quality

Flow rate (mL/min)	1	10	30
Average payload per LNPs	46.4	20.9	15.5
Ratio of empty LNPs (θ_{empty})	0.46	0.44	0.23

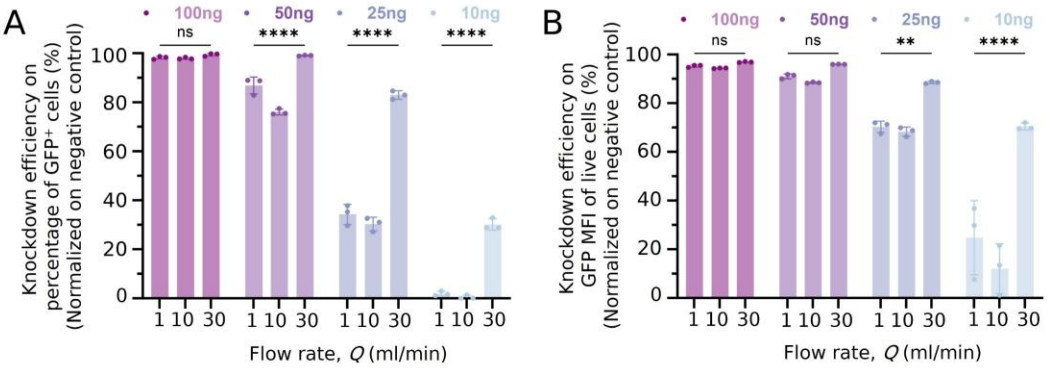

*Figure 4: **Effect of payload heterogeneity in transfection efficiency**. The top table summarizes siRNA-LNP quality metrics, illustrating improved payload uniformity and reduced fraction of empty LNPs at higher flow rates. (A, B) Knockdown efficiency of GFP expression in vitro (normalized to negative control) as a function of flow rate used during LNP formulation and siRNA dose. Higher flow rates produce more uniform LNPs with a reduced ratio of empty LNPs (top table), enhancing knockdown efficiency at low siRNA doses. At high doses, all formulations achieve near-complete GFP silencing.*

kMC simulations show the effect of PEG size and salt concentration

Along with mixing setup, PEG molecular weight (MW) is crucial in LNP assembly, stability, and delivery efficiency.[36–38] PEG forms a hydration shell that prevents LNP aggregation and extends circulation time. LNPs with shorter PEG (e.g., 750 Da) show little difference from PEG-free LNPs. However, longer PEG chains (over 5000 Da) can hinder cellular uptake and endosomal escape, key steps in delivering therapeutic agents.[38] Therefore, understanding the intermediate range (1000-5000 Da) is critical for optimizing stability and RNA delivery.

We investigated the effect of PEG MW by adjusting the Flory radius, $R_{\mathrm{F}} = \mathrm{a}n^{3/5}$ in PEG mediated repulsion calculation (a=0.37 nm is the monomer size and n the degree of polymerization).[39,40] Figure 5A shows that increasing the PEG MW slows LNP growth, creating an inverse relationship between PEG MW and LNP size. This well-known effect[25,37] is explained by the increased steric barrier E_{PEG} resulting from a larger R_{F} (see Eq. 2 in the Methods section). On the other hand, increasing PEG monotonically increases the ratio of empty LNPs (Figure 5B), reduces the normalized frequency of RNA loading (Figure 3C) and yields a broader distribution. A similar effect is expected when increasing the PEG fraction, as both variables influence the PEG mediated repulsions(see Eq. 2 and Ref. 20).

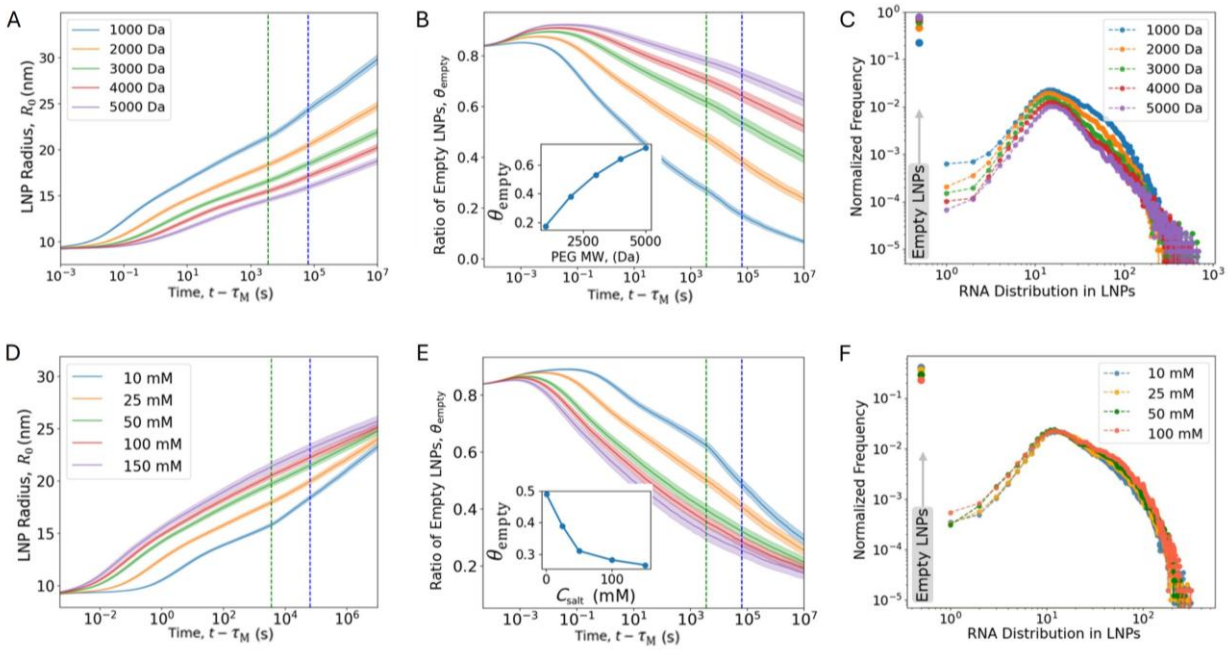

Figure 5: ***Effect of PEG Size and Salt Concentration on LNP Size and Payload Distribution.*** *(A) LNP growth and (B) ratio of empty LNPs from for varying PEG MW. The green vertical line denotes t = 1 h (post-mixing dialysis), and the blue line denotes t = 18 h (to match experimental measurement via CICS). Inset in (B) shows the ratio of empty LNPs at 18 hours. (C) siRNA loading distribution at t = 18 h. (D) LNP growth and (E) ratio of empty LNPs for different c_{salt}. Inset in (E) shows the ratio of empty LNPs for varying salt concentration at 18 h. (F) siRNA loading distribution at t = 18 h. Parameters: Initial LNP size is fixed at 9 nm, PEG ratio is 1.5% of total lipids, in (A, B, C) initial salt concentration c_{salt} = 0.025 M, while in (D, E, F) PEG MW = 2000 Da (R_F = 3.6 nm).*

In addition to PEG, salt concentration c_{salt} plays a key role in the self-assembly of LNPs by modulating electrostatic interactions. We explore an experimentally relevant range of c_{salt}, from 10 mM to 150 mM. kMC simulations reveal that increased c_{salt} accelerates aggregation by reducing electrostatic repulsion between cationic LNPs, as shown in Figure 5D. Beyond 50 mM, c_{salt} yielded similar aggregation kinetics, due to neutralization of surface charge and reduction of electrostatic surface potential. Additionally, the faster aggregation at higher c_{salt} reduces the proportion of empty LNPs (Figure 5E and inset). Changes in c_{salt} yield a similar effect on siRNA distribution (Figure 5F) as changes in the PEG MW (Figure 5C), although the effect of salt is less pronounced compared to PEG.

Analysis of LNP polydispersity and payload distribution

The above findings, combining kMC simulations and experimental data, highlight the complex interplay between mixing behavior, steric repulsion, and electrostatic interactions in LNP coalescence and growth. To further unravel precisely what factors determine the LNP heterogeneity, we isolate and analyze how different parameters affect LNP size and payload distribution by selectively deactivating the contributions from PEG (Eq. 2) and electrostatic and van der Waals

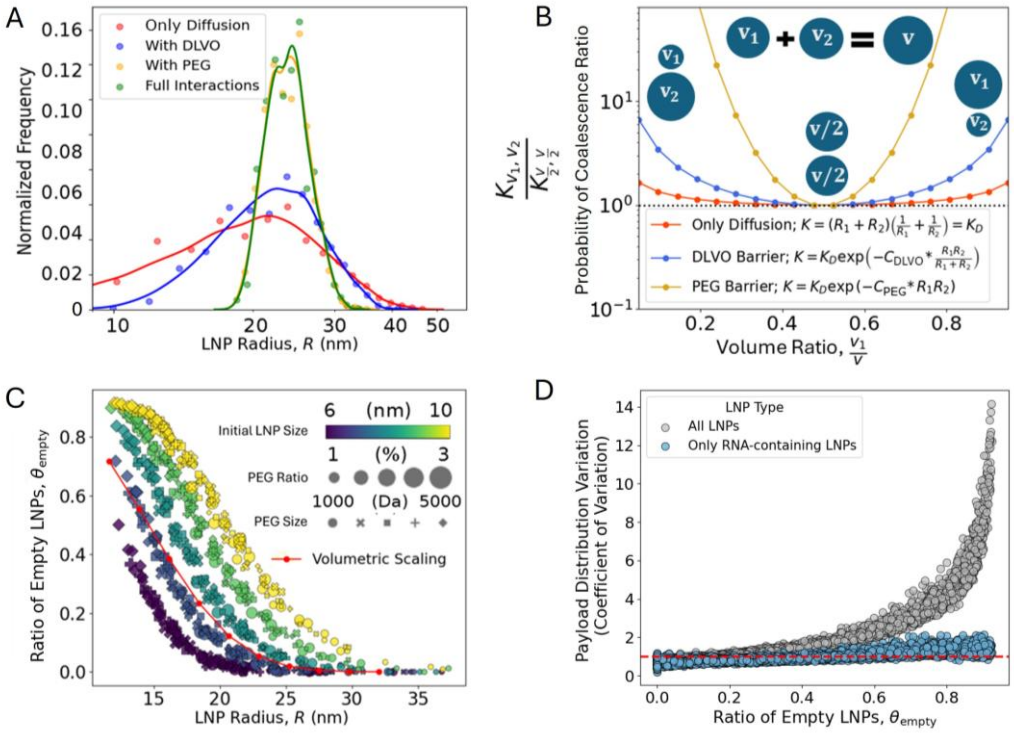

Figure 6: ***Analysis of Size Polydispersity and Payload Distribution Heterogeneity.*** *(A) LNP size distributions from kMC simulations using different aggregation kernels. Kernel density estimation (KDE) lines are overlaid to highlight the distribution profiles. (B) Aggregation bias expressed as the ratio* $K_{v_1,v_2} / K_{v_{/2},v_{/2}}$*, i.e., the probability of coalescence between an LNP of volume* v_1 *and one of volume* v_2*, relative to the probability of coalescing two LNPs of volume* $v/2$*, to form a larger LNP of volume* v*. This illustrates size-selective merging behavior. (C) Correlation between the ratio of empty LNPs and the average LNP radius at t = 18 h from kMC simulations. The red line indicates a volumetric scaling, with* $R_0 = 10$ *nm and* $\theta_0 = 0.83$*. (D) Correlation between the ratio of empty LNPs and the pooled coefficient of variation (CV) of siRNA payload distribution across LNP populations. The red dashed line shows the reference value CV = 1.*

(DLVO) interactions (Eq. 3) in kMC. Excluding both PEG and DLVO interactions results in LNP coalescence driven solely by diffusion, leading to a very broad LNP size distribution (Figure 6A). Minimizing size polydispersity is crucial for improving the performance, stability, and safety of LNPs.[1] Reintroducing either DLVO or PEG interactions narrows the distribution, with PEG having a significantly stronger effect.

To understand why PEG exerts greater influence, we examined the aggregation kernel *K* and the associated fusion energy barriers which reveal size-selective merging behavior. The PEG contribution to the fusion barrier between two spheres with radii R_1 and R_2 is $E_{\mathrm{PEG}} = C_{\mathrm{PEG}} R_1 R_2$, and for DLVO interactions it is $W = C_{\mathrm{DLVO}} \frac{R_1 R_2}{R_1 + R_2}$ (Eqs. 2 and 3). Here, C_{PEG} and C_{DLVO} are simplified multiplicative factors, derived using typical values in Eqs. 2 and 3 (see Figures S7 and S8). In Figure 6B, we plot the aggregation bias ratio $\frac{K_{v_1,v_2}}{K_{v/2,v/2}}$, representing the probability of coalescence between an LNP of volume v_1 and with another LNP of volume v_2, relative to the probability of merging two LNPs of volume $v/2$ to form a larger LNP of volume v. This ratio reveals that the probability of merging a smaller LNP with a larger one is higher than the probability of merging two LNPs of similar sizes. This behavior promotes a narrower size distribution by: (1) reducing the number of smaller LNPs and (2) limiting the formation of very large LNPs. We find that this effect is stronger for PEG compared to DLVO, as the DLVO fusion barrier increases exponentially with LNP radius, whereas the PEG fusion barrier increases as a squared exponential. Thus, PEG steric repulsion is the main factor that determines the LNP size and size polydispersity, with DLVO electrostatic repulsion having a smaller effect.

Having analyzed the size polydispersity, we now turn to payload distribution. Investigating the correlation between LNP size and the payload distribution, and whether these properties can be independently controlled, is crucial for optimizing LNP formulations. Surprisingly, we find that the PEG properties have no influence on the ratio of empty LNPs θ_{empty} (see overlapping symbols in Figure 5C), instead θ_{empty} is determined solely by the initial and final LNP size. The ratio of empty LNPs θ_{empty} and the average LNP radius *R* behave in a manner close to volumetric scaling: $\theta_{\mathrm{empty}}(R) = \theta_0^m$, where $\theta_{\mathrm{empty}}(R)$ is the fraction of empty LNPs at radius R, θ_0 the fraction of empty LNPs at R_0, and $m = (R/R_0)^3$ the average number of merging events to grow an average particle size from R_0 to R. The slight deviation of kMC results compared to volumetric scaling (Figure 6C) is attributed to the non-constant aggregation kernel, as the probability of merging is not equal for all LNPs. Thereby, mechanistically, changes in PEG or salt concentration directly affect the LNP size and size polydispersity, but do not directly affect the payload distribution (see overlapping symbols in Figure 6C), instead their effect on the payload distribution is indirect, a secondary effect due to the change in the average size *R*.

In addition to understanding the fraction of empty particles, it is also important to predict the shape of the full loading distribution (Figures 3E,F and 5C,F). We calculate the pooled coefficient of variation (CV) of siRNA loading distributions across LNP populations (Figure 6D). The CV, defined as the standard deviation divided by the mean, quantifies the payload variability in a population. Two metrics were computed: (1) the total CV across all LNPs, including empty particles, and (2) the CV

restricted to non-empty LNPs only. We observed that the heterogeneity in payload distribution increases monotonically with the fraction of empty LNPs, even for non-empty LNPs. This suggests that inefficient mixing not only result in more empty particles but also lead to greater variability in payload among loaded LNPs. For instance, this effect is clearly illustrated in Figure 3E,F where the peak frequency decreases and the distribution tail broadens, indicating higher payload variation. All data points for all different parameters fall on two master curves in Figure 6D, implying that the shape of the full loading distribution is highly correlated to the fraction of empty particles. Together, the results in Figure 6 establish a foundational understanding of the underlying correlations, offering a framework for decoupling and independently tuning LNP size and RNA loading distributions.

Machine learning elucidated design rules

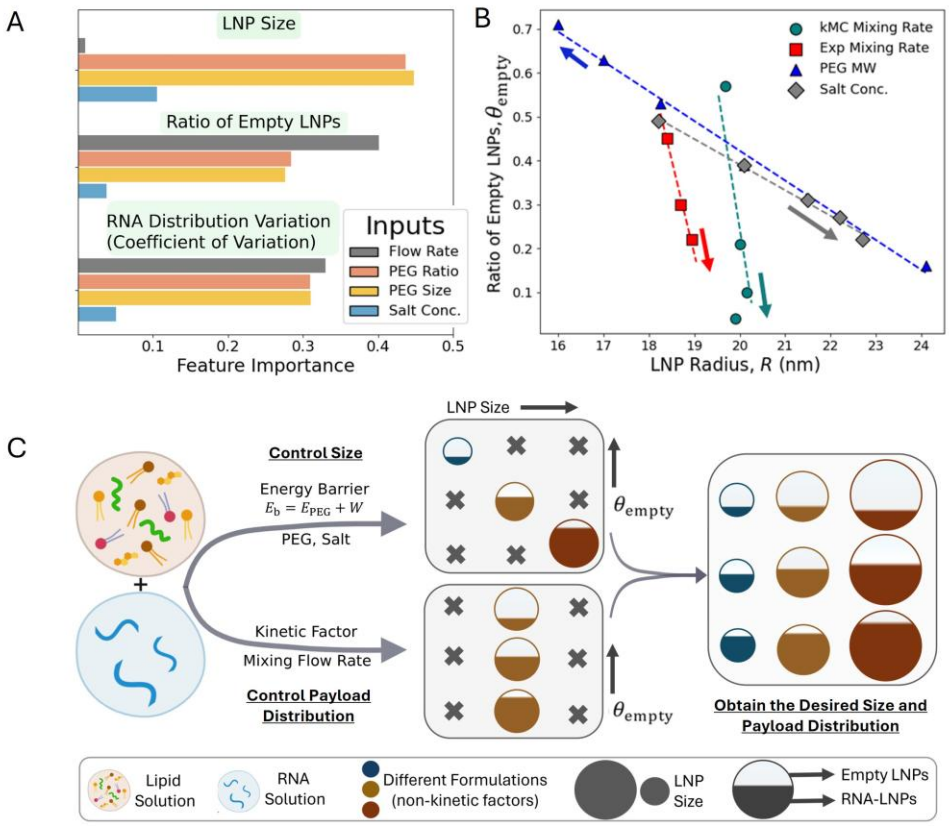

Figure 7. ***Design Guide for LNP Size and Payload Distribution.*** *(A) Random-forest feature importance analysis highlights the relative importance of different parameters in controlling LNP size and payload. The corresponding SHapley Additive exPlanations (SHAP), which shows positive and negative correlation of features with model predictions is shown in SI Figure S9. (B) Mixing rate strongly affects the ratio of empty LNPs θ_{empty}, but only weakly affects the LNP size R, in both kMC and experiment. Volumetric trend is observable for PEG and salt concentration (Figure 5). (C) Schematic demonstrating how the non-kinetic (energy barrier) and kinetic factors influence the LNP size and payload distribution. LNP energy barrier affects both the LNP size and payload distribution, whereas kinetic measures (e.g. mixing flow rate) independently control only the payload distribution. Crosses indicate inaccessible regions that cannot be achieved using either kinetic or non-kinetic tuning alone. By integrating both kinetic and non-kinetic factors, a wide range of LNP sizes and payload distributions can be achieved.*

To identify and quantify the most influential factors affecting LNP size and payload distribution, we performed a feature importance analysis using a random forest-based ensemble machine learning approach[41]. Feature importance quantifies how much each input variable (feature) contributes to output predictions. Our input parameters space includes: R_0 (6 to 10 nm), PEG MW (1000 to 5000 Da), PEG ratio of lipids (1 to 3%), and c_{salt} (10 to 150 mM). The analysis reveals that the initial LNP size or mixing flow rate has a negligible effect on the final average LNP size (Figure 7A), whereas the PEG molecular weight (MW) and the PEG ratio play the most significant roles in determining LNP size. c_{salt} can also influence LNP size, but its effect is constrained by the upper limit (charge neutralization for c_{salt} > 50 mM), as well as a relatively weaker effect compared to PEG (as seen in Figure 6B). Interestingly, the flow rate emerges as the most critical factor controlling the ratio of empty LNPs and the payload distribution, despite having minimal influence on the final LNP size (Figure 7A).

The LNP size and loading distribution data from Figures 3 and 5 are summarized in Figure 7B. Changes in mixing rate yield a nearly vertical slope on the θ_{empty}–R parametric plot, observed in both computational and experimental data, highlighting the ability to independently control payload distribution and LNP size. In contrast, the gradual slopes for PEG and salt indicate a strong correlation between LNP size and payload distribution, showing a volumetric trend.

Discussion:

Kinetic Monte Carlo simulations and experimental results together demonstrate that the size and payload distribution of LNPs can be controlled by utilizing energetic and kinetic factors (Figure 7A, B). These insights can be used to produce LNPs with controlled size and payload distribution (Figure 7C). The molecular weight of PEG, the PEG ratio, and salt concentration all contribute to the LNP fusion barrier (Figure 6), thereby determining the final LNP size. Adjusting these non-kinetic parameters allows for precise control of the LNP size, while their effect on the fraction of empty particles θ_{empty} and the loading distribution is indirect; it is a secondary effect due to the change in LNP size. In contrast, the mixing flow rate serves as a kinetic control parameter that dictates the relative assembly of lipid-lipid and RNA-lipid, enabling regulation of the RNA payload distribution independently of particle size. By integrating both types of control parameters, it becomes possible to produce LNPs with desired combination of size and payload distribution. Such precision is critical because LNP size strongly affects therapeutic delivery and cellular uptake, while homogeneity in RNA loading is important for maximizing potency and ensuring safety.

Our integrative computational framework combining molecular dynamics with kinetic Monte Carlo and a surrogate model for electrostatic interactions offers new mechanistic insights and practical design rules. This computational framework is generalizable to a broad class of self-assembly systems in solution. Within the kMC model, the chemical characteristics of lipid components can be systematically tuned by adjusting parameters such as pKa and molecular mass. Moreover, key experimental variables, including the duration of dialysis, total assembly time, and the ratio or amount of different molecular components, are explicitly controllable in the simulation. The framework is also adaptable to diverse payloads, provided their mass and charge properties are

properly defined. The simulation results show very good agreement with experimental observations, supporting the validity of the approach.

Despite the broad applicability of our LNP design rules, opportunities remain to extend the model and will shape future work. First, all simulations and experiments use a 22-bp siRNA payload. Larger cargos, such as >3 kb mRNA or plasmid DNA diffuse far more slowly. Extending the model to capture the interdiffusion and encapsulation of larger nucleic acids is an important next step, although we anticipate that the core design principles will remain the same. Second, the current parameter space; PEG 1–3 mol%, PEG MW 1–5 kDa, and 10–150 mM salt concentration covers common but not exhaustive formulation conditions. Systematically exploring higher PEG fractions, alternative buffer systems, broader flow-rate ratios, and a range of stoichiometry ratios will enhance the utility and generality of the proposed design map. This enhancement will extend the framework to other clinically realistic LNP formulations, accelerating the rational design of safer, more potent nucleic-acid therapeutics.

Conclusion:

In summary, through the integration of CG-MD simulations, kMC simulations, single-particle spectroscopy, machine learning, and transfections experiments, we have developed a comprehensive understanding of the siRNA payload distribution and its effect on delivery efficiency. We found that increasing homogeneity of siRNA distribution in LNPs is significantly improves drug delivery efficiency, especially at subsaturating doses. Our simulation model spans a wide range of time scales from microseconds to days, sufficient to capture experimentally relevant dynamics and heterogeneities. We found that the origin of empty LNPs and bimodal payload distribution is kinetic rather than thermodynamic. Bimodal RNA loading is caused by fast self-assembly of empty LNPs compared to the relatively slow diffusive mixing timescale. Rapid turbulent mixing of lipids and RNA significantly reduces empty LNPs and narrows the siRNA payload distribution without altering LNP size. While PEGylation and salt concentration also affect payload distribution, they do so only indirectly by changing the average LNP size. Additionally, kMC data shows that controlling the fusion barrier through PEG modification, rather than electrostatic stabilization alone, is crucial for achieving a narrow size distribution. These findings provide fundamental understanding of the LNP assembly process and offer essential design principles for optimizing LNP-based RNA therapeutics, aiming to enhance both their efficacy and safety. Additionally, the computational framework that we developed is freely accessible and can be extended to other therapeutic components, such as mRNA, and other nanoparticles, offering a powerful tool for efficiently screening nanoparticle formulation parameters.

Methods:

Single particle analysis of siRNA LNP on CICS

For the single nanoparticle analysis of siRNA LNP, the LNPs use the same formulation stoichiometry as used in simulations, with DLin-MC3-DMA (MedKoo Biosciences, Cat.#555308), DSPC (Avanti Polar Lipids, Cat.#850365), cholesterol (Sigma-Aldrich, Cat. #C8667), DMG-PEG2000 (NOF America,

356 Cat# GM020) at a molar ratio of 50:10:38.5:1.5 dissolved in 100% ethanol and RNA in 25 mM sodium
357 acetate buffer at pH 4.0. For the fluorescence detection, the LNP sample was prepared with
358 fluorescently labeled DSPC-TopFluor (Avanti Polar Lipids, Cat. #810281) substituting 10% DSPC
359 lipid, Cy5 labeled DSPE PEG2000-N-Cy5 (Avanti Polar Lipids, Cat. #810891) substituting 33.3% PEG
360 lipid and 100% siRNA- Cy3 (Sigma-Aldrich, Cat. #SIC003). The LNP samples were formulated by a 3-
361 inlet vortex mixer[42] at a total flow rate of 1, 10, 20, and 30 mL/min with a lipid to RNA flow rate ratio of
362 1:3. The final siRNA concentration after LNP formulation was 20 µg/mL. The sample was then (at 1 h)
363 dialyzed against 1x phosphate buffered saline (PBS) buffer at pH 7.4 for 12 h under 4 °C using dialysis
364 tubes with a molecular weight cutoff (MWCO) of 3,500 Pur-A-Lyzer dialysis kit, Sigma-Aldrich, Cat.
365 #PURD35050). Following dialysis, the LNPs were analyzed by Cylindrical Illumination Confocal
366 Spectroscopy (CICS).[43] The instrumentation and single-nanoparticle analysis have been developed
367 in our previous work and are described in detail.[20,25] Briefly, LNP samples were diluted sufficiently to
368 allow for single particles or siRNA molecules to go through the detector one at a time. The 3-color
369 fluorescence signals of each event were used to differentiate LNP populations into siRNA-
370 encapsulated LNPs, empty LNPs, and free unencapsulated siRNAs. By applying the fluorescence
371 deconvolution to the fluorescence distributions of siRNA encapsulated LNPs against a baseline of
372 single siRNA run separately, the siRNA payload distribution on the single LNP level was obtained (see
373 Figure S1).

374 *In vitro transfection*

375 LNPs were prepared as described above, and Silencer™ GFP (eGFP) siRNA (Thermo Fisher Scientific,
376 MA, USA) was encapsulated into the particles. The resulting LNPs were dialyzed against 1× PBS (pH
377 7.4) overnight at 4 °C. The final siRNA concentration was 20 µg/mL. B16F10-GFP cells were seeded
378 in 96-well plates at 10,000 cells per well. On the following day, cells were treated with LNPs at doses
379 of 100 ng, 50 ng, 25 ng, or 10 ng siRNA per well. After 48 h, cells were harvested and analyzed by flow
380 cytometry using an Attune NxT instrument (Thermo Fisher Scientific, MA, USA).

381

382 *Molecular dynamics simulations*

383 We modified Cooke and Deserno 3-bead lipid model[44] by introducing attractive head-to-head
384 interactions to promote nanoparticle formation (see Figure S2). Strength and cut-off values for these
385 attractive head-to-head interactions are provided in the SI. siRNA was included at a nitrogen-to-
386 phosphate (N/P) ratio of 6, and 1.5% of the lipid was PEGylated, using the same stoichiometry as the
387 experimental setup. The 22 base-pair siRNA and 2000 Da PEG were represented with a bead-spring
388 model. To match experimental conditions, 50% of the lipids carried a positive charge, interacting with
389 the negatively charged siRNA via Debye–Hückel interactions at a salt concentration of 0.025 M.[45]
390 Since, solvent mixing (e.g., ethanol dilution into the aqueous solution) occurs much faster (as
391 ethanol diffusion coefficient is 2 orders of magnitude greater than siRNA) relative to lipid and RNA
392 mixing, we consider that ethanol and water are fully mixed. Simulations were conducted using a
393 Langevin thermostat ensuring accurate diffusion behavior of RNA and lipids (see SI Figure S3). All
394 simulations were performed using LAMMPS,[46] and OVITO was used for visualization.[47] RNA–lipid and

lipid–lipid contacts (Figure 2B) were counted within the first valley of the radial distribution functions (see SI, Figure S4).

Kinetic Monte Carlo simulations

We consider that during mixing time τ_{M}, most lipids form spherical nanoparticles and that after τ_{M}, the solution is well-mixed. Consequently, the rate per unit concentration at which two spherical nanoparticles with radii R_i and R_j collide due to diffusion is $K_{\mathrm{collide}} = \frac{2k_{\mathrm{B}}T}{3\eta}\frac{(R_i+R_j)^2}{R_iR_j}$,[48] with k_{B} the Boltzmann constant, η the viscosity, and T the absolute temperature. Assuming Arrhenius behavior, the rate at which two specific LNPs in a volume V_{system} merge is

$$k_{ij}(R_i, R_j, \Phi_i, \Phi_j) = K_{\mathrm{collide}}(R_i, R_j)e^{-\frac{E_{\mathrm{b}}(R_i,R_j,\Phi_i,\Phi_j)}{k_BT}} / V_{\mathrm{system}} \,. \quad (1)$$

Here E_b is the fusion barrier between two LNPs with compositions Φ_i, Φ_j (volume fractions of different components in LNPs); E_{b} is modeled by the summation of PEG repulsion (E_{PEG}) and Derjaguin-Landau-Verwey-Overbeek (DLVO) interaction (W),[49] $E_{\mathrm{b}} = E_{\mathrm{PEG}} + W$. We consider that PEG chains are mobile on the LNP and for the two particles to merge the PEG must be excluded from the interaction area, thus, the free-energy barrier can be calculated as an entropic penalty to exclude PEG chains from the interaction area,

$$E_{\mathrm{PEG}}\left(R_i, R_j, \phi_i^{\mathrm{PEG}}, \phi_j^{\mathrm{PEG}}\right) = \frac{2\pi k_{\mathrm{B}}TR_{\mathrm{F}}}{3v_0(1-f_w)} R_iR_j \frac{R_i\phi_i^{\mathrm{PEG}} + R_j\phi_j^{\mathrm{PEG}}}{R_i + R_j} \,. \quad (2)$$

Here f_w=0.2 is the water volume fraction in LNPs[20,50], v_0 the volume per lipid, R_{F} the Flory radius of PEG determined by the PEG size,[39,40] and ϕ_i^{PEG} the PEG fraction in particle *i*. The DLVO interaction energy is a sum of van der Waals and electrostatic interactions[49]:

$$W\left(d,\ R_i, R_j\right) = -\frac{AR_iR_j}{6d\left(R_i + R_j\right)} + \epsilon\frac{R_iR_j\left(\psi_i^2 + \psi_j^2\right)}{4(R_i + R_j)}\left[\frac{2\psi_i\psi_j}{\psi_i^2 + \psi_j^2}\log\left(\frac{1 + e^{-\frac{d}{l_D}}}{1 - e^{-\frac{d}{l_D}}}\right) + \log\left(1 - e^{-\frac{d}{l_D}}\right)\right], \quad (3)$$

with A is the Hamaker constant, l_{D} the Debye screening length, ψ the surface electrostatic potential, ϵ the permittivity of solvent, and d the surface-to-surface distance between LNPs. The average time for a fusion event to occur is $\tau = {}^{1}/_{\sum k_{ij}}$, where the sum proceeds over all LNP pairs in the system. In the kMC implementation, a fusion event is chosen randomly according to the rate k_{ij}, with the time increment determined by τ. Each simulation was initialized with 2500 LNPs, expect for those in Figure 5A where simulations start with N = 2000 LNPs and proceed until only 100 LNPs remain to compare the polydispersity at the same average particle volume.

Charge regulation and surrogate model

The potentials ψ in the DLVO equation (Eq. 3) depend on the exact composition and size of the LNP and must be calculated by solving the charge regulation relations[51]. Using the Donnan

approximation[52] we approximate the potential ψ anywhere inside the LNP is a constant and equal to the surface potential ψ_0. Thus, the charge density is,

$$\rho = (1 - f_{\mathrm{w}})[\alpha_{\mathrm{MC3}}\rho_{\mathrm{MC3}}\phi^{\mathrm{MC3}} + \alpha_{\mathrm{RNA}}\rho_{\mathrm{RNA}}\phi^{\mathrm{RNA}}] + 2f_{\mathrm{w}}c_{\mathrm{salt}}N_{\mathrm{A}}\sinh\frac{e_0\psi_0}{k_{\mathrm{B}}T}\ , \tag{4}$$

Here f_{w} is the water volume fraction, ρ_{MC3} and ρ_{RNA} are the charge density of cationic lipid and siRNA, $\alpha_{\mathrm{MC3}}$and α_{RNA} are the degree of ionizations, $\phi^{\mathrm{MC3}}$and ϕ^{RNA} are volume fractions, N_{A} is the Avogadro number, and e_0 is the elementary charge. The surface potential depends on the charge density ρ as:

$$\frac{e_0\psi_0}{k_B T} = \frac{4\pi R^2 \rho l_{\mathrm{D}}}{3\left(1 + \frac{R}{l_{\mathrm{D}}}\right)}. \tag{5}$$

A root-finding approach is used to determine the charge density that satisfies the charge regulation equations (Eqs. 4 and 5). To make the model usable for all salt concentration, a surrogate neural network regression model is trained with data varying nanoparticle sizes, RNA fractions, and salt concentrations. The model efficiently predicts surface potential across a broad parameter space (see Figures S5 and S6). This trained data is provided as a part of the FormLNP kMC code and predictive toolbox.

Feature importance calculations

A random forest regression model was implemented to analyze the influence of formulation parameters on LNP size and the ratio of empty LNPs. The input variables included initial particle size, PEG MW, PEG fraction, and salt concentration, while the output variables were LNP size and the ratio of empty LNPs obtained from kMC simulations. Prior to training, all input variables were standardized. Separate random forest models were trained for each output variable (LNP final size, ratio of empty LNPs and payload distribution variance). The models were evaluated using 5-fold cross-validation with root mean squared error (RMSE) as the performance metric. Feature importance was determined using the mean decrease in impurity (Gini importance) from the trained models. All analyses, including data preprocessing, model training, cross-validation, and visualization, were conducted in Python using the scikit-learn machine learning package.

Associated content

Data Availability Statement: The data supporting the findings of this study are available within the Article and its Supplementary Information. All data are available from the corresponding author upon request.

Code availability:

The "FormLNP" computational framework that predicts LNP size and RNA payload distribution can be found on this website: https://sites.google.com/view/formlnp/home

All other custom scripts and tools used in this study are available from the corresponding author upon request.

Supporting Information: The Supporting Information includes detailed experimental procedures, molecular dynamics (MD) simulation parameters, kinetic Monte Carlo (kMC) simulation methodology[34,53], and machine learning approaches for charge regulation modeling. It also contains details of the predictive toolbox for LNP size and siRNA payload distribution.

Acknowledgement

This work is supported by start-up funds provided by the Whiting School of Engineering at JHU to TC, a grant from the National Cancer Institute to HQM (R01CA293906-01A1), and the National Institute of Allergy and Infectious Diseases (R01AI183336, R01AI181217) to THW. Computational work was carried out at the Advanced Research Computing at Hopkins (ARCH) core facility (rockfish.jhu.edu), which is supported by the National Science Foundation (NSF) grant number OAC 1920103.

References:

1. Hou, X., Zaks, T., Langer, R. & Dong, Y. Lipid nanoparticles for mRNA delivery. *Nat. Rev. Mater.* **6**, 1078–1094 (2021).
2. Kulkarni, J. A. *et al.* The current landscape of nucleic acid therapeutics. *Nat. Nanotechnol.* **16**, 630–643 (2021).
3. Brader, M. L. *et al.* Encapsulation state of messenger RNA inside lipid nanoparticles. *Biophys. J.* **120**, 2766–2770 (2021).
4. Barbier, A. J., Jiang, A. Y., Zhang, P., Wooster, R. & Anderson, D. G. The clinical progress of mRNA vaccines and immunotherapies. *Nat. Biotechnol.* **40**, 840–854 (2022).
5. Ferhan, A. R. *et al.* Lipid Nanoparticle Technologies for Nucleic Acid Delivery: A Nanoarchitectonics Perspective. *Adv. Funct. Mater.* **32**, 2203669 (2022).

486 6. Hejdankova, Z. *et al.* Lipid Nanoparticles for Broad-Spectrum Nucleic Acid Delivery. *Adv.*
487 *Funct. Mater.* **31**, 2101391 (2021).
488 7. Kang, H. *et al.* Small interfering RNA (siRNA)-based therapeutic applications against
489 viruses: principles, potential, and challenges. *J. Biomed. Sci.* **30**, 88 (2023).
490 8. Wang, Q. *et al.* Bispecific, Exosome-Mimetic Lipid Nanoparticles Facilitate Dual siRNAs
491 for Synergistic Therapy against Pancreatic Cancer. *Adv. Funct. Mater.* **34**, 2400485
492 (2024).
493 9. Vavilis, T. *et al.* mRNA in the Context of Protein Replacement Therapy. *Pharmaceutics* **15**,
494 166 (2023).
495 10. Qiu, M., Li, Y., Bloomer, H. & Xu, Q. Developing Biodegradable Lipid Nanoparticles
496 for Intracellular mRNA Delivery and Genome Editing. *Acc. Chem. Res.* **54**, 4001–4011
497 (2021).
498 11. Chen, R. *et al.* Engineering circular RNA for enhanced protein production. *Nat.*
499 *Biotechnol.* **41**, 262–272 (2023).
500 12. Huang, X. *et al.* Nanotechnology-based strategies against SARS-CoV-2 variants. *Nat.*
501 *Nanotechnol.* **17**, 1027–1037 (2022).
502 13. Pandya, A. *et al.* The future of cancer immunotherapy: DNA vaccines leading the
503 way. *Med. Oncol. Northwood Lond. Engl.* **40**, 200 (2023).
504 14. Liu, M. A. DNA and mRNA Vaccines for Chronic Viral Infections and Cancer:
505 Rationale, Mechanisms, and Progress. *Cancers* **14**, 5874 (2022).
506 15. Kim, S.-J. *et al.* Artificial Immunogenic Cell Death Lipid Nanoparticle Functions as a
507 Therapeutic Vaccine for Cancer. *Adv. Funct. Mater.* **33**, 2302825 (2023).

508 16. Bitounis, D., Jacquinet, E., Rogers, M. A. & Amiji, M. M. Strategies to reduce the risks
509 of mRNA drug and vaccine toxicity. *Nat. Rev. Drug Discov.* **23**, 281–300 (2024).
510 17. Ndeupen, S. *et al.* The mRNA-LNP platform's lipid nanoparticle component used in
511 preclinical vaccine studies is highly inflammatory. *iScience* **24**, 103479 (2021).
512 18. Sato, Y. *et al.* Highly specific delivery of siRNA to hepatocytes circumvents
513 endothelial cell-mediated lipid nanoparticle-associated toxicity leading to the safe and
514 efficacious decrease in the hepatitis B virus. *J. Control. Release Off. J. Control. Release*
515 *Soc.* **266**, 216–225 (2017).
516 19. Simonsen, J. B. A perspective on bleb and empty LNP structures. *J. Controlled*
517 *Release* **373**, 952–961 (2024).
518 20. Li, S. *et al.* Single-Particle Spectroscopic Chromatography Reveals Heterogeneous
519 RNA Loading and Size Correlations in Lipid Nanoparticles. *ACS Nano* **18**, 15729–15743
520 (2024).
521 21. Liao, S. *et al.* Transfection Potency of Lipid Nanoparticles Containing mRNA
522 Depends on Relative Loading Levels. *ACS Appl. Mater. Interfaces* **17**, 3097–3105 (2025).
523 22. Shepherd, S. J., Issadore, D. & Mitchell, M. J. Microfluidic formulation of
524 nanoparticles for biomedical applications. *Biomaterials* **274**, 120826 (2021).
525 23. Maeki, M., Uno, S., Niwa, A., Okada, Y. & Tokeshi, M. Microfluidic technologies and
526 devices for lipid nanoparticle-based RNA delivery. *J. Controlled Release* **344**, 80–96
527 (2022).
528 24. Prakash, G. *et al.* Microfluidic fabrication of lipid nanoparticles for the delivery of
529 nucleic acids. *Adv. Drug Deliv. Rev.* **184**, 114197 (2022).

530 25. Li, S. *et al*. Payload distribution and capacity of mRNA lipid nanoparticles. *Nat*.
531 *Commun*. **13**, 5561 (2022).
532 26. Akinc, A. *et al*. The Onpattro story and the clinical translation of nanomedicines
533 containing nucleic acid-based drugs. *Nat. Nanotechnol*. **14**, 1084–1087 (2019).
534 27. Kulkarni, J. A. *et al*. Rapid synthesis of lipid nanoparticles containing hydrophobic
535 inorganic nanoparticles. *Nanoscale* **9**, 13600–13609 (2017).
536 28. Zhao, L. *et al*. A review of confined impinging jet reactor (CIJR) with a perspective of
537 mRNA-LNP vaccine production. *Rev. Chem. Eng*. **40**, 887–916 (2024).
538 29. Zheng, H. *et al*. Preparation of Drug-Loaded Liposomes with Multi-Inlet Vortex
539 Mixers. *Pharmaceutics* **14**, 1223 (2022).
540 30. Hu, Y. *et al*. Kinetic Control in Assembly of Plasmid DNA/Polycation Complex
541 Nanoparticles. *ACS Nano* **13**, 10161–10178 (2019).
542 31. A First Course in Turbulence. *MIT Press* https://mitpress.mit.edu/9780262536301/a-
543 first-course-in-turbulence/.
544 32. Voter, A. F. INTRODUCTION TO THE KINETIC MONTE CARLO METHOD. in *Radiation*
545 *Effects in Solids* (eds Sickafus, K. E., Kotomin, E. A. & Uberuaga, B. P.) vol. 235 1–23
546 (Springer Netherlands, Dordrecht, 2007).
547 33. Feng, J., Markwalter, C. E., Tian, C., Armstrong, M. & Prud'homme, R. K. Translational
548 formulation of nanoparticle therapeutics from laboratory discovery to clinical scale. *J.*
549 *Transl. Med*. **17**, 200 (2019).

550 34. Clamme, J. P., Azoulay, J. & Mély, Y. Monitoring of the Formation and Dissociation of
551 Polyethylenimine/DNA Complexes by Two Photon Fluorescence Correlation
552 Spectroscopy. *Biophys. J.* **84**, 1960–1968 (2003).
553 35. Markwalter, C. E. & Prud'homme, R. K. Design of a Small-Scale Multi-Inlet Vortex
554 Mixer for Scalable Nanoparticle Production and Application to the Encapsulation of
555 Biologics by Inverse Flash NanoPrecipitation. *J. Pharm. Sci.* **107**, 2465–2471 (2018).
556 36. Suk, J. S., Xu, Q., Kim, N., Hanes, J. & Ensign, L. M. PEGylation as a strategy for
557 improving nanoparticle-based drug and gene delivery. *Adv. Drug Deliv. Rev.* **99**, 28–51
558 (2016).
559 37. Xia, Y., Tian, J. & Chen, X. Effect of surface properties on liposomal siRNA delivery.
560 *Biomaterials* **79**, 56–68 (2016).
561 38. Influence of the steric barrier activity of amphipathic poly(ethyleneglycol) and
562 ganglioside GM1 on the circulation time of liposomes and on the target binding of
563 immunoliposomes in vivo - Mori - 1991 - FEBS Letters - Wiley Online Library.
564 https://febs.onlinelibrary.wiley.com/doi/10.1016/0014-5793%2891%2980699-4.
565 39. Marsh, D., Bartucci, R. & Sportelli, L. Lipid membranes with grafted polymers:
566 physicochemical aspects. *Biochim. Biophys. Acta BBA - Biomembr.* **1615**, 33–59 (2003).
567 40. Li, M. *et al.* Brush Conformation of Polyethylene Glycol Determines the Stealth
568 Effect of Nanocarriers in the Low Protein Adsorption Regime. *Nano Lett.* **21**, 1591–1598
569 (2021).
570 41. Ho, T. K. Random decision forests. in *Proceedings of 3rd International Conference*
571 *on Document Analysis and Recognition* vol. 1 278–282 vol.1 (1995).

572 42. He, Z. *et al*. Scalable fabrication of size-controlled chitosan nanoparticles for oral
573 delivery of insulin. *Biomaterials* **130**, 28–41 (2017).
574 43. Liu, K. J. & Wang, T.-H. Cylindrical Illumination Confocal Spectroscopy: Rectifying
575 the Limitations of Confocal Single Molecule Spectroscopy through One-Dimensional
576 Beam Shaping. *Biophys. J.* **95**, 2964–2975 (2008).
577 44. Cooke, I. R. & Deserno, M. Solvent-free model for self-assembling fluid bilayer
578 membranes: Stabilization of the fluid phase based on broad attractive tail potentials. *J.*
579 *Chem. Phys.* **123**, 224710 (2005).
580 45. Kontogeorgis, G. M., Maribo-Mogensen, B. & Thomsen, K. The Debye-Hückel theory
581 and its importance in modeling electrolyte solutions. *Fluid Phase Equilibria* **462**, 130–
582 152 (2018).
583 46. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J.*
584 *Comput. Phys.* **117**, 1–19 (1995).
585 47. Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO–
586 the Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2010).
587 48. Roger, K., Botet, R. & Cabane, B. Coalescence of Repelling Colloidal Droplets: A
588 Route to Monodisperse Populations. *Langmuir* **29**, 5689–5700 (2013).
589 49. Hogg, R., Healy, T. W. & Fuerstenau, D. W. Mutual coagulation of colloidal
590 dispersions. *Trans. Faraday Soc.* **62**, 1638–1651 (1966).
591 50. Yanez Arteta, M. *et al*. Successful reprogramming of cellular protein production
592 through mRNA delivered by functionalized lipid nanoparticles. *Proc. Natl. Acad. Sci.* **115**,
593 E3351–E3360 (2018).

51. Curk, T. & Luijten, E. Charge Regulation Effects in Nanoparticle Self-Assembly. *Phys. Rev. Lett.* **126**, 138003 (2021).

52. Doi, M. *Soft Matter Physics*. (oxford university press, 2013).

53. L. Fliervoet, L. A. *et al.* Local release of siRNA using polyplex-loaded thermosensitive hydrogels. (2020) doi:10.1039/D0NR03147J.