Challenges and opportunities in computational studies for lipid nanoparticle development

Check for updates

Younghoon Oh[1] ✉, Sean K. Bedingfield[1], Severin T. Schneebeli[2], Jianing Li[3], Arezoo M. Ardekani[4], Kyle J. Colston[2] & Scott P. Brown[1]

Lipid nanoparticles (LNPs) are essential carriers for genetic medicines, yet optimizing their design remains challenging due to numerous parameters. Computational methods—including molecular dynamics (MD), computational fluid dynamics (CFD), and machine learning (ML)—offer molecular insights and predictive power. This perspective highlights recent advances, ongoing challenges, and the need for multiscale modeling frameworks and standardized experimental datasets to systematically explore LNP design space and improve the efficacy of next-generation formulations.

Lipid nanoparticles (LNPs) are nanoscale delivery vehicles composed of amphiphilic lipid components that self-assemble into colloidally stabilized structures in aqueous environments. They can be designed to encapsulate and protect genetic cargo such as RNA or DNA until delivery into target cells. LNPs represent an extremely complex system with nearly infinite design variables, making traditional experimental approaches alone insufficient for fully understanding and optimizing their performance. Computational studies provide a powerful complementary tool, allowing researchers to explore vast chemical and physical spaces efficiently. By systematically modeling key interactions and predicting functional outcomes, computational methods can accelerate breakthroughs in LNP design that would be impractical or impossible to achieve solely through experiments.

LNP optimization is plagued by limited design principles, even as the generation of in vivo data becomes increasingly feasible[1]. LNPs are the leading non-viral method for delivering genetic medicines involving mRNA and DNA, highlighted by the global implementation in COVID-19 mRNA vaccines. LNPs are produced as colloidally stabilized nanostructures. Despite being formed by simple oil-water emulsions, a highly complex series of tasks is required for LNPs to be therapeutically relevant. Performance relies on (1) encapsulation of nucleic acids, (2) stable particle formation, (3) stable circulation in the bloodstream, (4) favorable interaction and endosomal uptake in the target cells, and (5) endosomal escape to the cytoplasm for the nucleic acid to access relevant machinery. Each of these tasks is influenced by subtle, interdependent changes to parameters such as lipid structure[2], lipid composition[3], cargo-to-vehicle material ratio, particle fabrication process, and surface surfactants. Elucidating design principles

among so much data will require better data structuring and enable analytical techniques to optimize LNP performance.

Given the multi-scale and multi-parameter complexity of LNPs, leveraging computational power is essential for rational design and optimization. LNP performance is governed by a hierarchy of structural and functional determinants, spanning molecular lipid chemistry, self-assembly mechanisms, particle morphology, and in vivo pharmacokinetics. Each level presents unique challenges, requiring different computational approaches to extract meaningful insights. As illustrated in Fig. 1, these hierarchical length scales capture the intricacies of LNP behavior, emphasizing the need for integrative computational strategies. By systematically modeling key interactions at each scale, computational methods help bridge the gaps between fundamental molecular properties and therapeutic efficacy, enabling more precise control over LNP design.

Broadly, computational approaches in LNP research can be categorized into physics-based modeling and knowledge-based data science, both of which play crucial roles. Physics-based modeling—including computational quantum chemistry, all-atom and Coarse-grained molecular dynamics (CG-MD) simulations, and computational fluid dynamics (CFD)—offers unparalleled molecular and submolecular insights into LNP behavior[4–6]. These methods enable researchers to investigate structural dynamics, lipid-RNA interactions, and endosomal escape mechanisms at a level of detail inaccessible to experiments. Unlike traditional computer-aided drug discovery (CADD), which models small-molecule-protein interactions, physics-based modeling for LNPs must capture the complexities of self-assembly. Lipids are highly flexible molecules with rich phase behavior, requiring insights from soft-matter physics to understand the

[1]Eli Lilly and Company, Lilly Seaport Innovation Center, Boston, MA, USA. [2]Department of Industrial & Molecular Pharmaceutics and James Tarpo Jr. and Margaret Tarpo Department of Chemistry, Purdue University, West Lafayette, IN, USA. [3]Borch Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN, USA. [4]Department of Mathematics, School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA.
✉e-mail: oh_younghoon@lilly.com
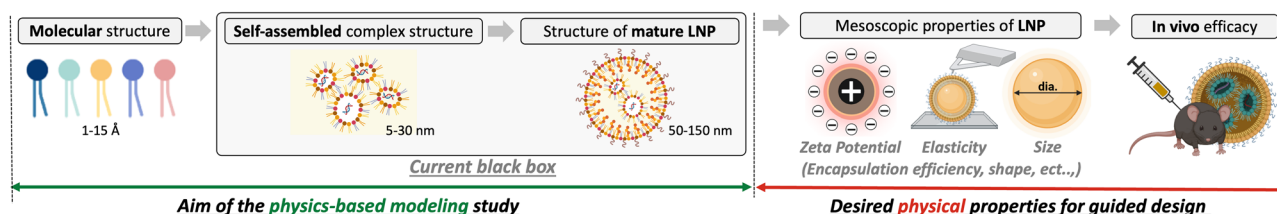
Fig. 1 | Hierarchical length scales in lipid nanoparticle (LNP) research, illustrate the complexity of LNP design and performance. Each level—from molecular structure to in vivo efficacy—captures key determinants of LNP function. Computational approaches, spanning physics-based modeling and machine learning-driven data science, offer essential tools for navigating this vast design space, enabling rational optimization of LNPs for improved therapeutic outcomes. Created in BioRender https://BioRender.com/ubts4t1.

thermodynamic and kinetic factors that govern LNP formation, stability, and function. Meanwhile, knowledge-based data science, particularly ML-driven approaches, has recently emerged as a promising tool for uncovering complex patterns in LNP formulation and performance. While early ML applications have shown encouraging results, their full potential remains untapped due to the scarcity of high-quality experimental datasets needed for robust model training.

In this perspective, we discuss how these computational approaches—physics-based modeling and ML-powered data science—can collectively drive breakthroughs in LNP research. By integrating mechanistic insights with predictive data-driven models, computational studies hold the potential to guide rational LNP design, improve therapeutic efficacy, and ultimately expand the possibilities of RNA-based medicines.
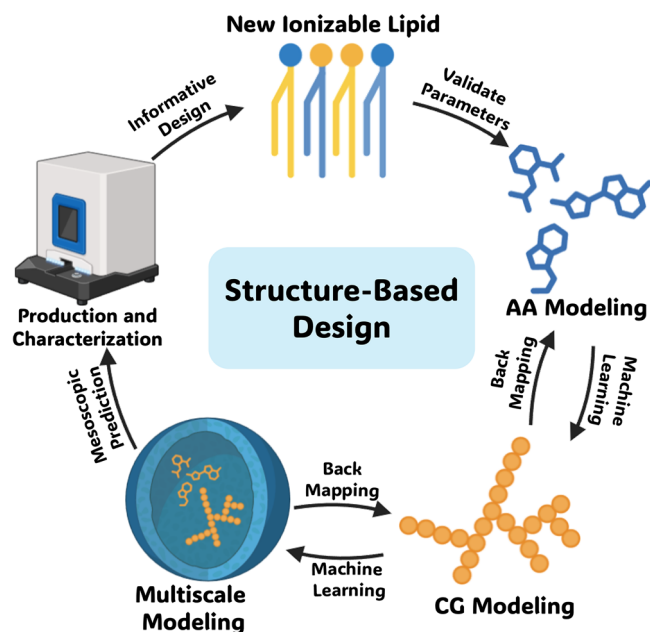
Physics-based modeling

Physics-based modeling refers to the use of molecular-level simulation techniques grounded in physical laws (e.g., Newtonian or statistical mechanics) to investigate the behavior, structure, and dynamics of biomolecular systems such as lipid nanoparticles. Physics-based modeling of lipid nanoparticles is a rapidly developing field, especially driven by recent advances in multiscale modeling and high-performance computing techniques. Complementary to experimental efforts for LNP formulation and characterization, physics-based modeling is expected to offer molecular-level insight into the LNP structure and interactions, essential to connect LNP composition to their activities, which ultimately provides predictive power to guide LNP design. An increasing number of publications have begun to demonstrate the effectiveness of physics-based modeling in explaining experimental observations, the self-assembly process of LNPs, and interactions with various biomolecules under different conditions. The goal of LNP physics-based modeling will be to provide accurate, high-throughput, structure-based virtual screening for LNP development and, hopefully, reduce the experimental time and cost and the need for extensive tests of composition variations. Herein, we provide a brief review of current approaches and their limitations in the physics-based modeling of LNPs, including all-atom and CG-MD, and CFD simulations, along with forward-seeing perspectives on future directions for advancement.

All-atom MD simulation

MD is a family of computational techniques that model the time-dependent behavior of atoms and molecules by numerically solving Newton's equations of motion. It has been widely used in physics, chemistry, biochemistry, and related areas to connect the microscopic structures of molecules to their collective or macroscopic properties, which enables the computational investigation of systems ranging from simple argon liquid[7] to complex biological systems like coronaviruses[8]. A primer text is available for readers who are new to MD[9]. More specifically, all-atom (AA) MD is a well-established technology for simulating lipid membranes and membrane-protein interactions, with numerous applications primarily aimed at enhancing our understanding of membrane dynamics[10], membrane remodeling processes[11–13], and membrane

proteins[12,14–16]. Recently, AA-MD models have also been used to examine the structure and dynamics of LNPs[17–19], although accurately modeling the protonation states of ionizable lipids in various membrane environments relevant to LNPs remains challenging[20–23]. Importantly, the protonation states of ionizable lipids in LNPs—factors that affect the overall charge of the LNPs as well as their interactions with biological systems—are often environment-dependent when the pK_a values of ionizable sites are near the pH of the solution. This can significantly influence the overall charge and interactions of an LNP with cells and surrounding biological media (e.g., proteins binding to an LNP as part of the biocorona). Due to this environment-dependent nature of ionizable sites, the protonation states can also be affected by specific manufacturing conditions (such as the type of dialyzing buffer used during LNP production, which is known to influence the transfection efficiency of LNPs) and the types and concentrations of helper lipids surrounding a particular ionizable lipid[24,25]. To address these challenges, it is essential to utilize more precise, constant pH molecular dynamics (CpHMD) models[26–29]. Notably, a scalable CpHMD model has been reported, which performs at comparable speeds to standard MD models[30]. This method implements l-dynamics based on the linear interpolation of partial charges between protonated and deprotonated states of appropriately parameterized ionizable sites. The additional computational cost associated with parameterization is offset by the substantial increase in performance, which allows for hundreds of ionizable sites to be modeled simultaneously. We anticipate that these models will effectively capture environment-dependent effects within LNPs, similar to how they can model the protonation states of peptides and permeation enhancers integrating into membranes during oral peptide absorption[31]. Very recently, scalable CpHMD models have been implemented for LNP modeling and were shown to accurately reproduce the apparent pK_a values for different LNP formulations (mean average error (MAE) = 0.5 pK_a units) in which pH-dependent structures are observed[32].

Overall, a key strength of atomistic adaptive membrane models is their accuracy in capturing complex supramolecular interactions, such as the hydrophobic effect, which dictates membrane self-assembly. Entropy plays a significant role in these molecular interactions among various lipid components within the membrane, as well as in the interactions at the membrane-solvent interface. However, a major challenge associated with AA-MD models is their relatively high computational cost due to the need to treat all the atoms in the system explicitly, particularly the solvent molecules, which often represent more than 70% of the total atoms present. Some of these challenges can be addressed by establishing reduced model systems, such as bilayer or multilamellar membrane models combined with periodic boundary conditions to approximate larger lipid nanoparticle (LNP) structures. Furthermore, enhanced sampling techniques—including umbrella sampling[33], metadynamics[34], replica exchange MD[35], steered MD[36,37], and biased MD[38]—can be employed to model events occurring on timescales that exceed the current capabilities of AA models. These advanced sampling techniques are specifically designed to improve the sampling of rare events during MD simulations, which would otherwise be

Fig. 2 | Structure-based design of new ionizable lipids and LNP formulations can be guided by hierarchical physics-based modeling systems. Results from highly accurate fine-grain systems can be used to improve the quality of less detailed simulations with the help of machine learning. Improved coarse-grain simulations, combined with the incorporation of multiscale models, serve to increase the quality of theoretical systems and improve their ability to predict mesoscopic properties. The production and characterization of a new LNP formulation can then inform the design of the next generation of formulations as theoretical methods are further refined to increase their predictive capabilities. Created in BioRender https://BioRender.com/po4kp29.

extremely difficult to observe within the limited timeframes that can be simulated with classical MD. We anticipate that this enhancement will ultimately allow AA-MD simulations to model rare events crucial for LNP function. This includes but is not limited to membrane reorganization processes that occur during LNP manufacturing or the endosomal release of LNP-encapsulated RNA from endosomes[6,39,40].

Nevertheless, each collective variable (CV) sampled using enhanced sampling methods incurs significant additional computational costs. This limitation restricts the number of CVs that can be efficiently sampled. Furthermore, defining reasonable CVs for enhanced sampling often requires a hypothesis about a molecular mechanism, which makes the simulation outcomes dependent on these initial assumptions. This dependency can hinder the exploration of the potential energy surface for CVs that aren't well-represented in the selected set for enhanced sampling. To address this issue, it is essential to develop new multiscale computational techniques that can better bridge models at different resolutions hierarchically, enabling the exploration of systems over larger time and spatial scales without sacrificing the accuracy of all-atom models. Machine learning (ML) and artificial intelligence (AI) will be crucial in these efforts, facilitating effective feature representation and linking various models for coarse-graining and back-mapping tasks.

Coarse-grained molecular dynamics simulation

CG-MD is a simulation approach in which groups of atoms are represented by simplified interaction sites, allowing for the modeling of larger systems and longer timescales compared to all-atom MD simulations. MD simulations of coarse-grained (CG) models help understand the detailed molecular structures and mechanisms of LNPs, which are often difficult to characterize experimentally[41]. Unlike AA models, there is a variety of CG models, ranging from the highly CG/low-resolution ones (e.g., 1 to 3 CG sites per lipid) to relatively fine-grained/high-resolution ones (e.g., over

6 sites per lipid). In the popular Martini-CG model[42–45], a typical lipid is represented by around 10–15 CG sites per lipid, with the key principle being a "four-to-one mapping" where ~4 heavy atoms are represented by a single CG site. The number of CG sites per lipid can vary slightly depending on the lipid structure, which can result in heterogeneity in the CG model and the resulting dynamics. The fine-grained CG models like Martini-CG retain essential chemical details of LNP and greatly facilitate parameterization and back mapping to AA models, which are useful to simulate LNPs with different lipid and nucleic acid compositions[45–47]. Further reducing the model resolution, the highly CG models are useful to simulate LNPs on more relevant temporal and spatial scales, and thus suitable to study the LNP self-assembly, size dependence, mechanical properties, etc. The highly CG models are also limited in the chemical details and complexities, and their parameters are often not transferrable, which requires significant efforts to develop and validate such models. However, many tools have been developed to automate CG model construction and parameterization[46,48–52].

Given the pros and cons of AA and CG models, hierarchical simulations (Fig. 2) that combine multiple models seamlessly may be a way to get the best of both models, allowing for AA accuracy and CG efficiency. Current hierarchical simulations have been categorized by how information is transferred between different resolutions[53]—in serial or in parallel. (i) The serial multiscale method carries out modeling at different resolutions in sequence, which takes advantage of sampling efficiency at lower resolutions and detailed accuracy at higher resolutions[54–56]. For instance, one can start modeling from the least detailed model and ultimately obtain a fully atomic model. This so-called top-down modeling[57] is promising to simulate complex systems like LNPs. (ii) The parallel multiscale methods include two different classes. The "hybrid resolution" methods[58–62] combine AA or united-atom (UA) models of a given subsystem of interest with a CG representation of the environment. New parameters, however, are often needed to account for the cross interactions between two resolutions[62,63]. In short, these hierarchical methods can be useful to study LNPs, but many key issues, such as transformations between multiple resolutions, sampling effectiveness, and simulation protocol optimization, still need to be studied systematically to advance their applications to systematic LNP simulation and, eventually, LNP development.

Computational fluid dynamics (CFD)

In the synthesis of LNPs, achieving rapid and uniform mixing is crucial for producing particles with well-defined sizes and high encapsulation efficiency[64,65]. To produce LNPs with low polydispersity via antisolvent precipitation, the process requires mixing times on the order of 100 ms. Research indicates that confined impinging jet mixers (CIJMs) and multi-inlet vortex mixers (MIVMs) are effective for facilitating rapid solvent exchange and nanoprecipitation[65–68]. CFD simulations can be used to better understrand fluid flow and mixing dynamics in different mixers.

Microfluidic mixing has played a key role in the self-assembly of LNPs at the lab scale[69]. A key challenge in these systems is achieving efficient mixing at low Reynolds numbers, where turbulence is largely absent, making diffusion the dominant transport mechanism[70,71]. Diffusion-based self-assembly is impractical due to its slow timescales, making hydrodynamic mixing essential for rapid nucleation and controlled growth[72]. Staggered herringbone mixers have been shown to produce monodisperse LNPs, but their low throughput presents a challenge[72]. While parallelization can increase throughput, it also adds complexity and cost to the system[69]. Higher throughput LNP production can occur using inertial micromixers at higher flow rates[69]. Among these microfluidic mixers, Dean vortex-based micromixers are suggested for LNP manufacturing due to their ability to maintain efficient mixing at high throughput[64]. Dean vortex-based micromixers use curved microchannels to generate transverse rotational flows, known as Dean vortices[73]. These vortices arise due to flow instabilities in curved geometries and actively moving fluid between different regions of the channel, enhancing mixing even at low Reynolds numbers. This passive design offers effective mixing without complex structures. There is a critical transition regime in these devices, which influences the optimal flow

Fig. 3 | Maturity of ML platforms across areas of active research in the pharmaceutical industry. As one traverses from left to right the maturity of ML method development declines, while at the same time the inherent complexity of research data being generated increases. ML methods are just beginning to be explored in nanomedicine. We expect the impact of ML methods to be greater when data complexity increases.
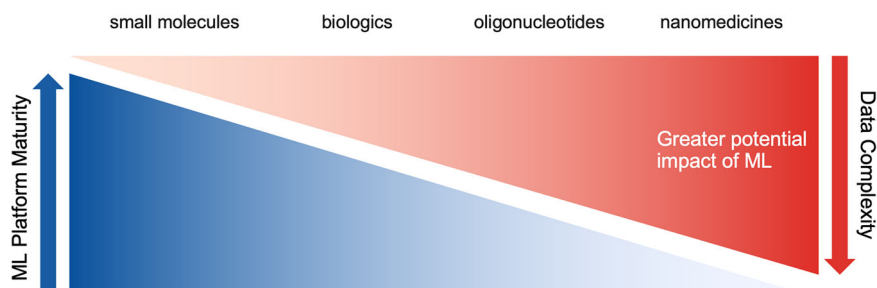
conditions for LNP formation[64]. For achieving LNPs with optimal encapsulation efficiency, charge, and monodispersity, it is crucial to operate above this transition regime, as performance is compromised when operating within or below it. These insights highlights the importance of computational fluid dynamics to define the physical parameters necessary for consistent LNP quality.

CFD has been instrumental in analyzing and optimizing mixing, providing insights into flow behavior and mixing efficiency [74,75]. Various passive micromixer designs have been developed to enhance mixing performance, including split-and-recombine (SAR) micromixers[76,77], staggered herringbone mixers[78], and Dean vortex-based mixers[79]. These designs enhance mixing by stretching and folding fluid layers, thereby increasing the interfacial surface area available for diffusion.

Large Eddy Simulations (LES) and Direct Numerical Simulations (DNS) have been extensively used to investigate turbulence-driven mixing in these systems, understanding the role of self-sustained oscillations and flow structures on mixing uniformity[80,81]. Studies on confined impinging jet mixers (CIJMs) suggest that turbulent structures impact mixing and encapsulation efficiency[82].

Computational studies can be used to evaluate mixing dynamics for different micromixer designs. High-fidelity CFD simulations provide a detailed understanding of fluid dynamics, mixing efficiency, and nanoparticle size, complementing experimental measurements. Computational approaches enable researchers to investigate a broad range of design parameters, flow conditions, and geometric modifications saving time and reducing costs. By systematically examining the effects of flow regimes, e.g., Reynolds number (the ratio inertial to viscous forces), chaotic flow structures, and turbulence-driven mixing, these studies can help optimize mixing platforms for enhanced nanoparticle properties, encapsulation efficiency, and scalability[74,75,78,79,81,82].

Knowledge-based data science
Recent progress and limitations of current machine learning (ML)-based approaches

ML refers to data-driven computational methods that identify patterns and make predictions based on large datasets. In drug development, ML methods present opportunities to reduce R&D burden and improve design success rates. To successfully bring a new drug to market requires substantial investment of time and resources[83]. Methods in ML are opportunities for systematic reduction in the investment burden required for drug discovery, with the potential to improve probabilities of success as well as reduce design cycle times. However, ML methods require as input existing data sets that are representative of the research problem of interest. ML methods are unable to overcome problems caused by irrelevant or erroneous research data.

In small-molecule drug discovery, ML methods are mature platforms with wide deployment and routine use. This is perhaps not surprising as ML methods in small-molecule drug discovery have access to very large data sets. Additionally, method development has been a focus of intense research for well over 30 years.

The situation can be very different when one examines more recent paradigms in drug discovery. For example, the use of ML methods in support of biologics research is still relatively recent and under intense active

development. Shown in Fig. 3 is a depiction of platform maturity over four different paradigms in drug discovery: small molecule, biologics, oligonucleotides, and nanomedicine. In moving from left to right in Fig. 3 we see decreasing platform maturity, while we also observe increasing complexity in the data that is generated in the course of research operations. For research data with high complexity, we expect a greater benefit from ML methods compared to situations with lower complexity research data.

ML methods for use in nanomedicine research are in their infancy. Despite this fact, there has been noteworthy progress reported in recent literature. For example, image-based classification of LNP experimental readouts, allowing detection of subtle features corresponding to differences in internal composition[84]. Another noteworthy advancement is seen in the recent report for pooled in vitro activity and cell viability data for on 6454 LNP formulations reported across 21 independent studies. This study examined 11 different molecular featurization techniques (e.g., descriptors, fingerprints, and graph-based representations), alongside six ML algorithms. The resulting accuracy of >90% was reported[85]. The authors also implemented transfer learning to bridge the gap between in vitro and in vivo predictions by integrating base model outputs with LNP size, polydispersity index, and zeta potential. Despite the limited size and class imbalance of the in vivo dataset, the transfer learning models achieved accuracy >82%[85].

Additional reports appear in the literature with the primary objective of exploring optimization of the ionizable-lipid component, as it is considered to be a key variable in LNP property optimization and in vivo tissue distribution[86–88]. Another publication reports results for multiparameter optimization of LNP properties[89]. The above methods show promise for acceleration of nanomedicine research. However, it is too early to tell how transferable these methods will be to other research contexts in nanoparticle design.

Inherent challenges of nanomedicine research data

The rational design of nanomedicines represents a relatively new research paradigm for the pharmaceutical industry. Examination of published data in nanomedicine literature reveals a predominance of sparse data sets that are not representative of the breadth of the research problem. As described in the introduction, the parameter space for LNP design is inherently high-dimensional and is not well understood or even well characterized. Additional layers of nuance and complexity can be added to the problem for research projects that require in vivo readouts as the primary assay for hypothesis evaluation, data interpretation, and design prioritization.

Some noteworthy attempts to develop new approaches for systematic exploration of LNP design space have been reported recently in scientific literature[90,91]. However, progress to date has been limited to custom solutions designed and deployed in-house, which, by necessity, focus on immediate needs and near-term deliverables with limited impact on the field.

The LNP design problem has created new challenges for computational methods, due to the unprecedented underlying complexity of the problem. Contributing to the challenge is the lack of established scientific standards for the reporting of nanomedicine research data. A large number of experimental parameters must be captured during formulation of LNPs in order to provide adequate detail for the procedure used to prepare just one

LNP sample. As described in the introduction, we have encountered quite a few situations in which seemingly minor changes to one process variable can produce LNP samples with profoundly different readouts from in vivo experimental assay. These results are robust in that they persist across replicate preparations and replicate experimental measurements. For a subset of these cases, the LNP property characteristics in the samples are measured to be identical by experiment (e.g., size, encapsulation efficiency, etc.). The implications of this are subtle but significant: the measured properties of LNPs are not sufficient to distinguish between samples for in vivo experiment. In order for ML methods to be relevant to in vivo design, the process variables must be captured.

Thus, there is a real need for the development of new data models that are capable of supporting and even driving advances in the field of nano-medicine research. A successful data model should provide sufficient detail to adequately capture the parameter space required for rational design of LNPs. Proposals for new data models should derive from critical discussion in the nanomedicine experimental and theoretical communities. Solution implementation should be driven by community consensus and adopted as editorial standards for publication of nanomedicine research. A collective push toward the common goal of advancing our understanding of nanoparticle design and enabling the successful development of novel therapeutics.

Lipid nanoparticles (LNPs) have revolutionized the delivery of genetic medicines, yet their rational design remains an unsolved challenge due to the immense complexity of their structure-function relationships. Computational approaches—including physics-based modeling and ML—offer powerful tools to navigate this complexity by enabling molecular-level insight, multiscale simulation, and predictive optimization of LNP formulations.

In this perspective, we outlined the current landscape of computational strategies in LNP research. All-atom and CG-MD simulations provide a mechanistic understanding of lipid-lipid and lipid-cargo interactions, while CFD supports the rational design of scalable mixing systems. ML-based data science offers new ways to mine experimental data, accelerate formulation screening, and uncover latent design rules—though such efforts remain limited by the quality and structure of available datasets.

Integration across modeling scales and data modalities is essential to fully realize the potential of computational tools in LNP development. A community-wide push toward standardized data reporting, improved data models, and interdisciplinary collaboration will be critical for building reliable in-silico platforms that can inform real-world design decisions. With these advances, computational studies will not only complement experimental workflows but also drive a new paradigm of rational, predictive, and efficient LNP engineering for next-generation therapeutics.

Data availability
No datasets were generated or analysed during the current study.

References

1. Dahlman, J. E. et al. Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics. *Proc. Natl. Acad. Sci. USA* **114**, 2060–2065 (2017).
2. Kulkarni, J. A. et al. The current landscape of nucleic acid therapeutics. *Nat. Nanotechnol.* **16**, 630–643 (2021).
3. Gilleron, J. et al. Image-based analysis of lipid nanoparticle-mediated siRNA delivery, intracellular trafficking and endosomal escape. *Nat. Biotechnol.* **31**, 638–646 (2013).
4. Chan, C., Du, S., Dong, Y. & Cheng, X. Computational and experimental approaches to investigate lipid nanoparticles as drug and gene delivery systems. *Curr. Top. Med. Chem.* **21**, 92–114 (2021).
5. Mageed, H. A. E., Mohamed, S. & Saleh, A. The tiny big world of solid lipid nanoparticles and nanostructured lipid carriers: an updated review. *J. Microencapsul.* **39**, 1–42 (2021).
6. Settanni, G. Computational approaches to lipid-based nucleic acid delivery systems. *Eur. Phys. J. E* **46**, 127 (2023).
7. Rahman, A. Correlations in the motion of atoms in liquid argon. *Phys. Rev.* **136**, A405 (1964).
8. Yu, A. et al. A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys. J.* **120**, 1097–1104 (2021).
9. Li, J. Molecular dynamics. *ACS Public.* https://doi.org/10.1021/acsinfocus.7e9008 (2025).
10. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143 (2018).
11. Hsieh, M.-K., Yu, Y. & Klauda, J. B. All-atom modeling of complex cellular membranes. *Langmuir* **38**, 3–17 (2022).
12. Li, J. Biomolecular simulation to elucidate small-molecule modulation of mechanosensor protein. *Proc. Natl. Acad. Sci. USA* **121**, e2319968121 (2024).
13. Sodt, A. J., Sandar, M. L., Gawrisch, K., Pastor, R. W. & Lyman, E. The molecular structure of the liquid-ordered phase of lipid bilayers. *J. Am. Chem. Soc.* **136**, 725–732 (2014).
14. Goossens, K. & De Winter, H. Molecular dynamics simulations of membrane proteins: an overview. *J. Chem. Inf. Model.* **58**, 2193–2202 (2018).
15. Itaya, H. et al. All-atom molecular dynamics elucidating molecular mechanisms of single-transmembrane model peptide dimerization in a lipid bilayer. *ACS Omega* **6**, 11458–11465 (2021).
16. Remington, J. M., Ferrell, J. B., Schneebeli, S. T. & Li, J. Concerted rolling and penetration of peptides during membrane binding. *J. Chem. Theory Comput.* **18**, 3921–3929 (2022).
17. Trollmann, M. F. W. & Boeckmann, R. A. Mitochondrial RNA lipid nanoparticle phase transition. *Biophys. J.* **121**, 3927–3939 (2022).
18. Zhang, X., Ma, G. & Wei, W. Simulation of nanoparticles interacting with a cell membrane: probing the structural basis and potential biomedical application. *NPG Asia Mater* **13**, 52 (2021).
19. Zhang, Z. et al. Molecular dynamics simulation of lipid nanoparticles encapsulating mRNA. *Molecules* **29**, 4409 (2024).
20. Dehghani-Ghahnaviyeh, S. et al. Ionizable amino lipids distribution and effects on DSPC/cholesterol membranes: implications for lipid nanoparticle structure. *J. Phys. Chem. B* **127**, 6928–6939 (2023).
21. Ibrahim, M., Gibert, J., Heinz, M., Nylander, T. & Schwierz, N. Structural insights on ionizable Dlin-MC3-DMA lipids in DOPC layers by combining accurate atomistic force fields, molecular dynamics simulations and neutron reflectivity. *Nanoscale* **15**, 11647–11656 (2023).
22. Paloncyova, M., Cechova, P., Srejber, M., Kuhrova, P. & Otyepka, M. Role of Ionizable Lipids in SARS-CoV-2 vaccines as revealed by molecular dynamics simulations: from membrane structure to interaction with mRNA fragments. *J. Phys. Chem. Lett.* **12**, 11199–11205 (2021).
23. Rissanou, A. N., Ouranidis, A. & Karatasos, K. Complexation of single stranded RNA with an ionizable lipid: an all-atom molecular dynamics simulation study. *Soft Matter* **16**, 6993–7005 (2020).
24. Hald Albertsen, C. et al. The role of lipid components in lipid nanoparticles for vaccines and gene therapy. *Adv. Drug Deliv. Rev.* **188**, 114416 (2022).
25. Zhang, J., Fan, H., Levorse, D. A. & Crocker, L. S. Ionization behavior of amino lipids for siRNA delivery: determination of ionization constants, SAR, and the impact of lipid pKa on cationic lipid-biomembrane interactions. *Langmuir* **27**, 1907–1914 (2011).
26. Buslaev, P. et al. Best practices in constant pH MD simulations: accuracy and sampling. *J. Chem. Theory Comput.* **18**, 6134–6147 (2022).
27. Henderson, J. A. et al. A guide to the continuous constant pH molecular dynamics methods in Amber and CHARMM. *Living J. Comput. Mol. Sci.* **4**, 1563 (2022).
28. Jansen, A., Aho, N., Groenhof, G., Buslaev, P. & Hess, B. Phbuilder: a tool for efficiently setting up constant pH molecular dynamics simulations in GROMACS. *J. Chem. Inf. Model.* **64**, 567–574 (2024).

29. Martins de Oliveira, V., Liu, R. & Shen, J. Constant pH molecular dynamics simulations: Current status and recent applications. *Curr. Opin. Struct. Biol.* **77**, 102498 (2022).

30. Aho, N. et al. Scalable constant pH molecular dynamics in GROMACS. *J. Chem. Theory Comput.* **18**, 6148–6160 (2022).

31. Colston, K. J., Faivre, K. T. & Schneebeli, S. T. Permeation enhancer-induced membrane defects assist the oral absorption of peptide drugs. *ChemRxiv* https://doi.org/10.26434/chemrxiv-22025-n26424f26438 (2025).

32. Colston, K. J. M., Santiago, C. & Schneebeli, S. T. Structure-based modeling of environment-dependent protonation states across LNP formulations with atomistic CpHMD. *ChemRxiv* https://doi.org/10.26434/chemrxiv-22025-psmd26433 (2025).

33. Ngo, S. T. & Pham, M. Q. Umbrella sampling-based method to compute ligand-binding affinity. *Methods Mol. Biol.* **2385**, 313–323 (2022).

34. Ray, D. & Parrinello, M. Kinetics from metadynamics: principles, applications, and outlook. *J. Chem. Theory Comput.* **19**, 5649–5670 (2023).

35. Lei, H. & Duan, Y. Improved sampling methods for molecular simulation. *Curr. Opin. Struct. Biol.* **17**, 187–191 (2007).

36. Ramírez, C. L., Martí, M. A. & Roitberg, A. E. Steered molecular dynamics methods applied to enzyme mechanism and energetics. *Methods Enzymol.* **578**, 123–143 (2016).

37. Do, P.-C., Lee, E. H. & Le, L. Steered molecular dynamics simulation in rational drug design. *J. Chem. Inf. Model.* **58**, 1473–1482 (2018).

38. Campbell, A. J., Lamb, M. L. & Joseph-McCarthy, D. Ensemble-based docking using biased molecular dynamics. *J. Chem. Inf. Model.* **54**, 2127–2138 (2014).

39. Garaizar, A. et al. Toward understanding lipid reorganization in RNA lipid nanoparticles in acidic environments. *Proc. Natl. Acad. Sci. USA* **121**, e2404555121 (2024).

40. Mendonca, M. C. P., Kont, A., Kowalski, P. S. & O'Driscoll, C. M. Design of lipid-based nanoparticles for delivery of therapeutic nucleic acids. *Drug Discov. Today* **28**, 103505 (2023).

41. Kjølbye, L. R. et al. Towards design of drugs and delivery systems with the Martini coarse-grained model. *QRB Discov.* **3**, e19 (2022).

42. de Jong, D. H. et al. Improved parameters for the Martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9**, 687–697 (2013).

43. Monticelli, L. et al. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).

44. Souza, P. C. T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388 (2021).

45. Kjoelbye, L. R. et al. Martini 3 building blocks for lipid nanoparticle design. *ChemRxiv* https://doi.org/10.26434/chemrxiv-2024-bf4n8 (2024).

46. Grzetic, D. J., Hamilton, N. B. & Shelley, J. C. Coarse-grained simulation of mRNA-loaded lipid nanoparticle self-assembly. *Mol. Pharm.* **21**, 4747–4753 (2024).

47. Leung, A. K. K. et al. Lipid nanoparticles containing siRNA synthesized by microfluidic mixing exhibit an electron-dense nanostructured core. *J. Phys. Chem. C* **116**, 18440–18450 (2012).

48. Empereur-Mot, C. et al. Automatic multi-objective optimization of coarse-grained lipid force fields using SwarmCG. *J. Chem. Phys.* **156**, 024801 (2022).

49. Empereur-Mot, C. et al. Swarm-CG: automatic parametrization of bonded terms in MARTINI-based coarse-grained models of simple to complex molecules via fuzzy self-tuning particle swarm optimization. *ACS Omega* **5**, 32823–32843 (2020).

50. Jin, J., Pak, A. J., Durumeric, A. E. P., Loose, T. D. & Voth, G. A. Bottom-up coarse-graining: principles and perspectives. *J. Chem. Theory Comput.* **18**, 5759–5791 (2022).

51. Souza, P. C. T. et al. Protein-ligand binding with the coarse-grained Martini model. *Nat. Commun.* **11**, 3714 (2020).

52. Empereur-Mot, C. et al. Automatic optimization of lipid models in the martini force field using swarmCG. *J. Chem. Inf. Model.* **63**, 3827–3838 (2023).

53. Ayton, G. S., Noid, W. G. & Voth, G. A. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* **17**, 192–198 (2007).

54. Perlmutter, J. D. et al. All-atom and coarse-grained molecular dynamics simulations of a membrane protein stabilizing polymer. *Langmuir* **27**, 10523–10537 (2011).

55. Rohrdanz, M. A., Zheng, W., Lambeth, B., Vreede, J. & Clementi, C. Multiscale approach to the determination of the photoactive yellow protein signaling state ensemble. *PLoS Comput. Biol.* **10**, e1003797 (2014).

56. Rzepiela, A. J., Sengupta, D., Goga, N. & Marrink, S. J. Membrane poration by antimicrobial peptides combining atomistic and coarse-grained descriptions. *Faraday Discuss* **144**, 431–443 (2010).

57. Zhao, X., Liao, C., Ferrell, J. B., Schneebeli, S. T. & Li, J. Top-down multiscale approach to simulate peptide self-assembly from monomers. *J. Chem. Theory Comput.* **15**, 1514–1522 (2019).

58. Han, W., Wan, C.-K., Jiang, F. & Wu, Y.-D. PACE force field for protein simulations. 1. full parameterization of version 1 and verification. *J. Chem. Theory Comput.* **6**, 3373–3389 (2010).

59. Han, W., Wan, C.-K. & Wu, Y.-D. PACE force field for protein simulations. 2. folding simulations of peptides. *J. Chem. Theory Comput.* **6**, 3390–3402 (2010).

60. Kar, P. & Feig, M. Hybrid all-atom/coarse-grained simulations of proteins by direct coupling of CHARMM and PRIMO force fields. *J. Chem. Theory Comput.* **13**, 5753–5765 (2017).

61. Liao, C. et al. Capturing the multiscale dynamics of membrane protein complexes with all-atom, mixed-resolution, and coarse-grained models. *Phys. Chem. Chem. Phys.* **19**, 9181–9188 (2017).

62. Shelley, M. Y. et al. A new mixed all-atom/coarse-grained model: application to melittin aggregation in aqueous solution. *J. Chem. Theory Comput.* **13**, 3881–3897 (2017).

63. Shi, Q., Izvekov, S. & Voth, G. A. Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane-bound ion channel. *J. Phys. Chem. B* **110**, 15045–15048 (2006).

64. Ripoll, M. et al. Optimal self-assembly of lipid nanoparticles (LNP) in a ring micromixer. *Sci. Rep.* **12**, 9483 (2022).

65. Hourdel, L. et al. Overview on LNP-mRNA encapsulation unit operation: Mixing technologies, scalability, and influence of formulation & process parameters on physico-chemical characteristics. *Int. J. Pharm.* **672**, 125297 (2025).

66. Johnson, B. K. & Prud'homme, R. K. Chemical processing and micromixing in confined impinging, jets. *AIChE J* **49**, 2264–2282 (2003).

67. Santos, J. L. et al. Continuous production of discrete plasmid DNA-polycation nanoparticles using flash nanocomplexation. *Small* **12**, 6214–6222 (2016).

68. Johnson, B. K. & Prud'homme, R. K. Mechanism for rapid self-assembly of block copolymer nanoparticles. *Phys. Rev. Lett.* **91**, 118302 (2003).

69. Maeki, M. et al. Understanding the formation mechanism of lipid nanoparticles in microfluidic devices with chaotic micromixers. *PLoS ONE* **12**, e0187962 (2017).

70. Jeffs, L. B. et al. A scalable, extrusion-free method for efficient liposomal encapsulation of plasmid DNA. *Pharm. Res.* **22**, 362–372 (2005).

71. Stone, H. A., Stroock, A. D. & Ajdari, A. Engineering flows in small devices: microfluidics toward a lab-on-a-chip. *Annu. Rev. Fluid Mech.* **36**, 381–411 (2004).

72. Shepherd, S. J. et al. Scalable mRNA and siRNA lipid nanoparticle production using a parallelized microfluidic device. *Nano Lett.* **21**, 5671–5680 (2021).

73. Tabeling, P. *Introduction to Microfluidics* (Oxford University Press, 2005).

74. Hao, Y., Seo, J. H., Hu, Y., Mao, H. Q. & Mittal, R. Flow physics and mixing quality in a confined impinging jet mixer. *AIP Adv.* **10**, 045105 (2020).

75. Liu, Y., Cheng, C., Prud'homme, R. K. & Fox, R. O. Mixing in a multi-inlet vortex mixer (MIVM) for flash nano-precipitation. *Chem. Eng. Sci.* **63**, 2829–2842 (2008).

76. Kim, D. S., Lee, S. H., Kwon, T. H. & Ahn, C. H. A serpentine laminating micromixer combining splitting/recombination and advection. *Lab Chip* **5**, 739–747 (2005).

77. Raza, W., Hossain, S. & Kim, K.-Y. A review of passive micromixers with a comparative analysis. *Micromachines* **11**, 455 (2020).

78. Belliveau, N. M. et al. Microfluidic synthesis of highly potent limit-size lipid nanoparticles for in vivo delivery of siRNA. *Mol. Ther. Nucleic Acids* **1**, e37 (2012).

79. Na, G. S. et al. Full-cycle study on developing a novel structured micromixer and evaluating the nanoparticle products as mRNA. *delivery carriers. J. Contr. Release* **373**, 161–171 (2024).

80. Icardi, M. et al. Investigation of the flow field in a three-dimensional Confined Impinging Jets Reactor by means of microPIV and DNS. *Chem. Eng. J.* **166**, 294–305 (2011).

81. Liu, Z., Passalacqua, A., Olsen, M. G., Fox, R. O. & Hill, J. C. Dynamic delayed detached eddy simulation of a multi-inlet vortex reactor. *AIChE J.* **62**, 2570–2578 (2016).

82. Devos, C. et al. Impinging jet mixers: A review of their mixing characteristics, performance considerations, and applications. *AIChE J.* **71**, e18595 (2025).

83. Wouters, O. J., McKee, M. & Luyten, J. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA* **323**, 844–853 (2020).

84. Lim, K. & Ardekani, A. M. Hyperspectral enhanced imaging analysis of nanoparticles using machine learning methods. *Nanoscale Adv.* https://doi.org/10.1039/d4na00205a (2024).

85. Kumar, G. & Ardekani, A. M. Machine-learning framework to predict the performance of lipid nanoparticles for nucleic acid delivery. *ACS Appl. Biol. Mater.* https://doi.org/10.1021/acsabm.4c01716 (2024).

86. Metwally, A. A., Nayel, A. A. & Hathout, R. M. In silico prediction of siRNA ionizable-lipid nanoparticles in vivo efficacy: machine learning modeling based on formulation and molecular descriptors. *Front. Mol. Biosci.* **9**, 1042720 (2022).

87. Wang, W. et al. Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery. *Nat. Commun.* **15**, 10804 (2024).

88. YuT, Y. LipidBERT: a lipid language model pre-trained on METiS de novo lipid library. *arXiv* https://doi.org/10.48550/arXiv.2408.061502024 (2024).

89. Maharjan, R., Kim, K. H., Lee, K., Han, H. & Jeong, S. H. Machine learning-driven optimization of mRNA-lipid nanoparticle vaccine quality with XGBoost/Bayesian method and ensemble model approaches. *J. Pharm. Anal.* **14**, 100996 (2024).

90. Cui, H. et al. LUMI-lab: a foundation model-driven autonomous platform enabling discovery of new ionizable lipid designs for mRNA delivery. *bioRxiv* https://doi.org/10.1101/2025.02.14.638383 (2025).

91. Xu, Y. et al. AGILE platform: a deep learning powered approach to accelerate LNP development for mRNA delivery. *Nat. Commun.* **15**, 6305 (2024).

Author contributions
Y.O. conceived the topic and planned the overall structure of the manuscript. Y.O. and S.K.B. assisted with integration and editing across sections. S.T.S., K.J.C., J.L., and A.M.A. contributed to the physics-based modeling section: S.T.S. and K.J.C. developed the all-atom molecular dynamics (MD) subsection, J.L. composed the coarse-grained MD subsection, and A.M.A. contributed the computational fluid dynamics subsection. S.P.B. authored the section on knowledge-based data science. All authors reviewed and edited the manuscript.

Competing interests
Prof. Jianing Li is a member of the editorial board of npj Drug Discovery. The author had no involvement in the peer review or editorial decision-making process for this manuscript. The other authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Younghoon Oh.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.