

High-throughput platforms for machine learning-guided lipid nanoparticle design

Andrew R. Hanna¹, David A. Issadore^{1,2,3} & Michael J. Mitchell^{1,4,5,6,7,8,9} ✉

Abstract

To design a lipid nanoparticle (LNP) that effectively delivers nucleic acids to a specific cell or tissue type, multiple lipid components and their relative proportions must be decided on from a large number of options. As there is an incomplete understanding of the relationship between the molecular composition of a delivery vehicle, its structure and its activity, the decision is made by screening many formulations. Emerging technologies have rapidly accelerated the generation of large LNP libraries and the testing of their physicochemical properties and behaviour in vitro and in vivo. These screening tools are being increasingly integrated within artificial intelligence-driven discovery systems, wherein data obtained from the characterization and biological testing of LNPs are fed into machine learning models. These models can provide non-obvious relationships between composition and physical or biological outputs, or predict entirely new lipid structures. In this Perspective, we discuss advancements in the automation and parallelization of chemical synthesis, particle formulation, characterization and pharmacological screening that have improved the throughput of generating and testing large libraries of LNPs for nucleic acid delivery. We notably highlight the short-term potential of coupling these high-throughput platforms with machine learning to accelerate the prediction of optimal nucleic acid LNPs for new therapeutic targets.

Sections

Introduction

Size of the LNP design space

High-throughput synthesis and characterization of lipid libraries

LNP formulation and characterization

In vitro and in vivo characterization

Outlook

¹Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA. ²Department of Chemical Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Electrical and Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA. ⁴Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁵Center for Cellular Immunotherapies, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁶Penn Institute for RNA Innovation, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁷Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁸Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁹Institute for Regenerative Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ✉e-mail: mjmitch@seas.upenn.edu

Introduction

Since the early 2000s, combinatorial synthesis and high-throughput screening have enabled the development of substantially more efficient and less toxic nonviral delivery systems for nucleic acid therapeutics¹. Iterative screening and optimization of polymeric nanoparticles made from polyethylenimine² and poly- β amino esters^{1,3,4} have helped reduce biological side effects. Ultimately, these polymers became a framework for the ionizable lipid, a major component of lipid nanoparticles (LNPs)^{5,6}. From these early advancements, LNPs have emerged as a clinically successful nucleic acid delivery method⁷ as demonstrated by patisiran^{8,9}, a treatment for transthyretin-mediated amyloidosis, and by the Pfizer and Moderna SARS-CoV-2 vaccines¹⁰. As of August 2025, over 100 LNP-based therapeutics are currently in the clinical pipeline and over 170 studies are actively recruiting participants¹¹ ([ClinicalTrials.gov search](https://ClinicalTrials.gov/search)).

Given the complexity of the LNP as a multicomponent delivery vehicle, the screening pipeline generally comprises synthesis of new lipid components, formulation of LNPs with those components, then subsequent physicochemical characterization, in vitro testing and in vivo testing of lead candidates (Fig. 1a). Modifications to any one of the lipid components or their relative ratios, cargo, buffer conditions, or method of mixing can lead to distinct performances across different

biologically relevant models. Across the literature, inconsistencies in formulation and screening methods ultimately obscure concrete design rules. Furthermore, given the approximately billion atom scale of an LNP and the complex relationship between kinetic and thermodynamic interactions during nucleation and subsequent particle stabilization, physics-based in silico modelling strategies have remained too computationally demanding to narrow the potential design space of lipid structures that generate biologically effective LNPs^{12,13}. Although large areas of unexplored chemical space still remain^{14–17}, modular, parallelizable synthetic manifolds for new lipids have enabled the discovery of LNPs with preferable immunogenicity¹⁸, transfection efficiency across cell types of interest^{19,20}, or biodistribution^{21,22}.

In this Perspective, we highlight developments in the automated chemical synthesis of lipid components, high-throughput formulation of LNP libraries^{17,18}, and the screening¹⁹ of these libraries, which have enabled the expanding use of machine learning (ML) approaches to relate LNP composition to behaviour in biological systems, and thus to predict new formulations (Fig. 1b). We discuss how these technologies both accelerate throughput and improve consistency across the many steps necessary to gather screening data on the biological performance of LNPs. More consistency in the techniques used to formulate and screen LNPs, combined with the sharing of data in public

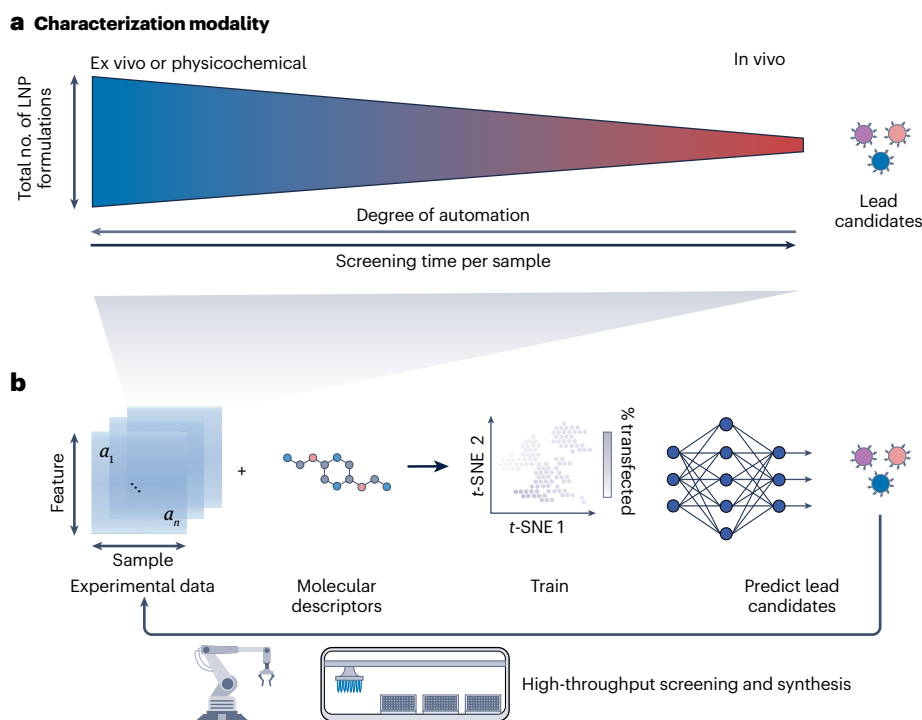


Fig. 1 | Integrating automation and machine learning to accelerate LNP discovery. **a**, The canonical screening pathway for lipid nanoparticles (LNPs) relies on the progression of formulations through a funnel of screening techniques, from physicochemical to in vitro to in vivo screening, wherein candidates that do not meet a set of standards are eliminated. Automation is more widely applied to the measurement of physicochemical or ex vivo readouts, and it is often used to narrow the number of potential candidates for screening in vivo. Although the screening time required per sample to generate in vivo data is higher, the resultant data often provide more information about how the LNP will ultimately perform as a therapeutic. **b**, Automation can accelerate

the collection of screening data, which can train machine learning models to reduce the size of future screening sets. An example of a closed-loop discovery system entails the synthesis and formulation of LNPs, which are tested in biological systems to generate data, iterating through experimental cycles to predict an optimally performing LNP candidate for some therapeutic function. Data gathered to train machine learning models are represented here using a *t*-distributed stochastic neighbour embedding (*t*-SNE) plot, a technique that maps high-dimensional data to a lower-dimensional map for visualization. Training is illustrated using a schematic neural-network architecture.

repositories, could enable the development of LNP foundation models that predict new formulations for unmet, and potentially unrelated, therapeutic needs.

Size of the LNP design space

To understand the potential of ML models for LNP discovery, it is important to understand the size of the design space for any given formulation. When referring to a design space, what is understood is a representation of the entire possible set of lipid structures and their representative combinations. Given the constraint in resources, we can only traverse small representative sections of this design space, which is, for example, what is done in a 'design of experiments' approach^{23,24} that typically relies on holding the stoichiometric ratio of certain lipid components fixed while varying across others to establish correlations across the data set. However, given the size of the design landscape, it is possible to spend considerable time screening without finding the optimal candidate.

Variability in lipid structure

The potential design space for an LNP is generally constrained to a combination of four lipid components – the helper lipid, the steroid, the polyethylene glycol (PEG) lipid and the ionizable lipid – and a nucleic acid cargo²⁵. Considering published lipids, if we conservatively estimate that there exist at least 10^5 ionizable lipids, 10^2 possible steroids²⁶, 10^3 helper lipids and 10^3 PEG lipids, there already exist 10^{13} possible LNPs at a single compositional ratio²⁷ (Fig. 2a). The function of each lipid defines its number of potential structures (Fig. 2b). Helper lipids are commonly phospholipids that largely define LNP curvature^{28,29}. Structural modifications include changes to fatty acid tail length and structure, and modification of headgroup polarity and size³⁰. The steroid component contributes to the internal order of the other lipids. Although cholesterol is commonly used in LNP formulations, many other steroids have been tested given their variable effects on LNP stiffness^{31,32}. PEG lipids are surface-presenting lipids that contribute to particle size, colloidal stability and circulation times^{33,34}. Structural alterations to PEG length and lipophilic tail length markedly change the rate at which PEG desorbs from the LNP surface³³. Finally, ionizable lipids promote escape of the nucleic acid from the endosome and have by far the greatest structural variability^{19,35}. These lipids consist of an amine-modified 'core' and long lipophilic tails. Common structural motifs include piperazine head groups^{36–38}, which react with epoxides¹⁸ or acrylates^{38,39} on lipophilic tails in one-pot syntheses.

Although greater theoretical structural variability might exist among these lipids, our approximation is based on contemporary literature, generating estimates on the number of possible structures by extrapolating from those already tested or those that could be conceived through slight structural modifications – such as changes in the number of carbons in each fatty acid tail. Assuming at least 10^2 compositional ratios of these lipids result in biological differences, there exist at least 10^{15} possible LNPs to screen (Fig. 2a). For each biological indication, the optimal formulation may change entirely. In terms of the nucleic acid cargo, there exist a massive number of potential functional variants of each small interfering RNA⁴⁰, mRNA²⁵, circular RNA⁴¹, self-amplifying mRNA⁴² and gene-editing cargo^{43,44}.

Biological implications of lipid structure

Changing the lipid structures or ratios within this design landscape leads to different outcomes both in and ex vivo. Large screening efforts varying composition and lipid structure have identified formulations capable of cell-specific delivery, as demonstrated in models for both ex

vivo^{37,45} and in vivo T cell⁴⁶ transfection with chimeric antigen receptor (CAR) mRNA. Although lipids often share structural motifs, these motifs do not indicate similarities in biological performance. For instance, although the ionizable lipids C12-200 and C14-494 both centrally feature piperazine moieties, LNPs incorporating these lipids function optimally for different applications, hepatic transfection^{35,47} and ex vivo CAR T cell transfection^{37,48}, respectively (Table 1). Beyond the most structurally diverse lipid – the ionizable lipid – changes to the helper lipid modulate the efficiency of endosomal escape^{12,28}. LNPs can be redirected to specific cell types or organs such as the lung^{20,32,49–51} or the spleen⁵² through the addition of permanently charged lipid components. This strategy promotes extrahepatic tropism yet can result in additional toxicity⁵³.

In vivo behaviour is particularly difficult to predict given the relationship between surface chemistry, topology and the recruitment of different biomolecular corona proteins^{54,55}. As early as 30 s after entering the bloodstream⁵⁶, a series of proteins will adsorb on the surface of LNPs, dynamically binding and interchanging based on charge and topology^{57,58}. This protein corona can have marked effects on particle biodistribution⁴⁹, transport mechanisms both intracellularly and through the bloodstream⁵⁷, cellular metabolism⁵⁶ and immunogenicity^{59,60}.

Thus far, deep learning models have been used to generate virtual lipid libraries based on experimentally validated synthetic pathways. Although some of these candidates have been screened in vitro to predict new ionizable lipid structures⁶¹, the biological performance of these lipids can vary, owing to differences in the structural features that are most relevant to function across cell types^{61,62}. Thus, the generation of new lipids or prediction of formulations that perform better than existing standards for a given biological function probably requires training of the ML model on data collected from the specific biological system being studied, necessitating continued screening in vivo.

High-throughput synthesis and characterization of lipid libraries

The synthesis of new lipids is integral to the generation of new LNP formulations. Driven by growing interest in drug carriers, combinatorial chemistry has experienced a notable resurgence since the 2020s. This revival has led to innovations that have enabled the automation of synthesis and purification of new chemical compounds⁶³. Pairing these high-throughput platforms with ML tools can help in creating a feedback loop^{64,65}, wherein automated systems can generate products, analyse them and optimize reaction schemes with minimal human input⁶⁶.

Synthetic schemes and methods

Although many potential structures for each chemical component of the LNP exist, the ionizable lipid is the most structurally diverse of all lipids^{19,23,37}. From early development, ionizable lipids have been designed to be synthesized combinatorially. These modular combinatorial schemes use a weakly nucleophilic and ionizable amine core that reacts with a number of fatty acid-like 'tail' components bearing an electrophilic head group^{4,6,38,67} in a single step, in one pot, and with minimal work-up⁴⁷ (Fig. 3a). Thus, as chemical diversity scales multiplicatively with the number of each component, the ionizable lipid is the most readily applicable of all LNP components for automation of synthesis^{4,36,47}. Typical examples of the head groups on these fatty acid-like tails include epoxides^{37,68} or enones³⁸, reacting with the amine-functionalized cores in S_N2 -type^{38,39} or aza-Michael-type^{69,70} additions. These combinatorial libraries can then be formulated into LNPs

Perspective

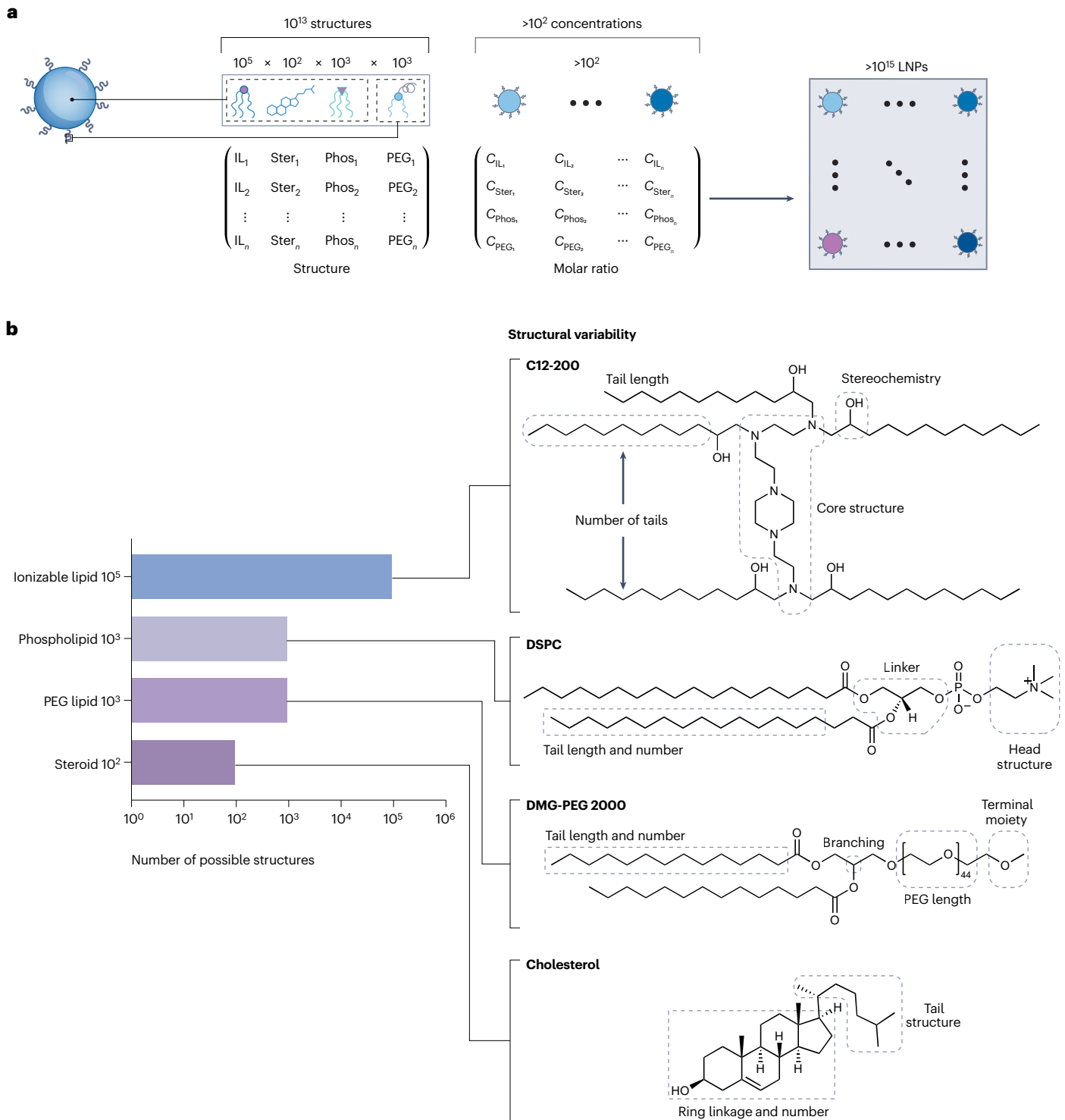


Fig. 2 | The size of the LNP design space. **a**, The potential combinations of lipid structures and compositions result in a massive space to consider when designing lipid nanoparticles (LNPs). Different areas of this design space (that is, different lipid structures and compositional ratios) will result in particles with meaningfully different behaviour in biological systems. **b**, Relative proportions

of each lipid component with respect to the total number of structural combinations (left). Commonly varying substructures for each lipid type (right, top to bottom): ionizable lipid, helper lipid, polyethylene glycol (PEG) lipid and steroid. DMG-PEG 2000, 1,2-dimyristoyl-*rac*-glycero-3-methoxypolyethylene glycol-2000; DSPC, 1,2-distearoyl-*sn*-glycero-3-phosphocholine.

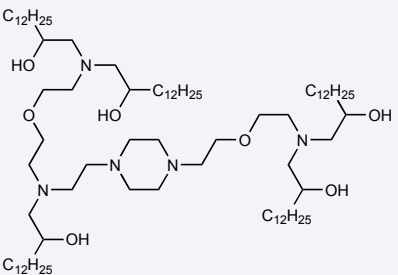
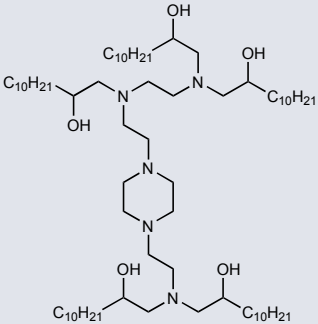
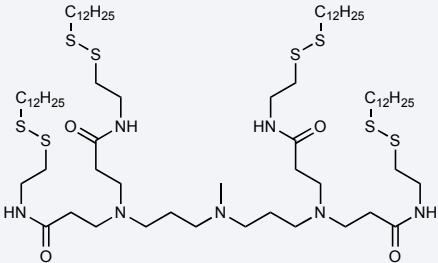
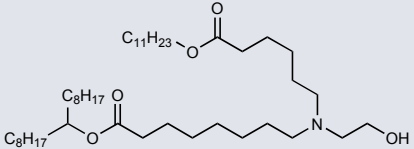
Perspective

wherein relative concentrations of each of the four components are typically initially fixed while varying the ionizable lipid, then screened en masse both in vitro and in vivo to identify lead candidates^{35,39,71,72}.

Although both ionizable and other lipids have been previously synthesized without concern for their stereochemistry^{35,38,72}, differences in cellular uptake and response are seen with variation of stereochemistry^{73,74}. For ionizable lipids generated using epoxide ring openings, the number of potential stereoisomers scales by 2^n with the number of amines n in the core. Generating and purifying particular stereoisomers in these systems would probably require the addition of engineered enzymes⁷⁵ or enantiospecific catalysts⁷⁶, potentially requiring additional work-up steps.

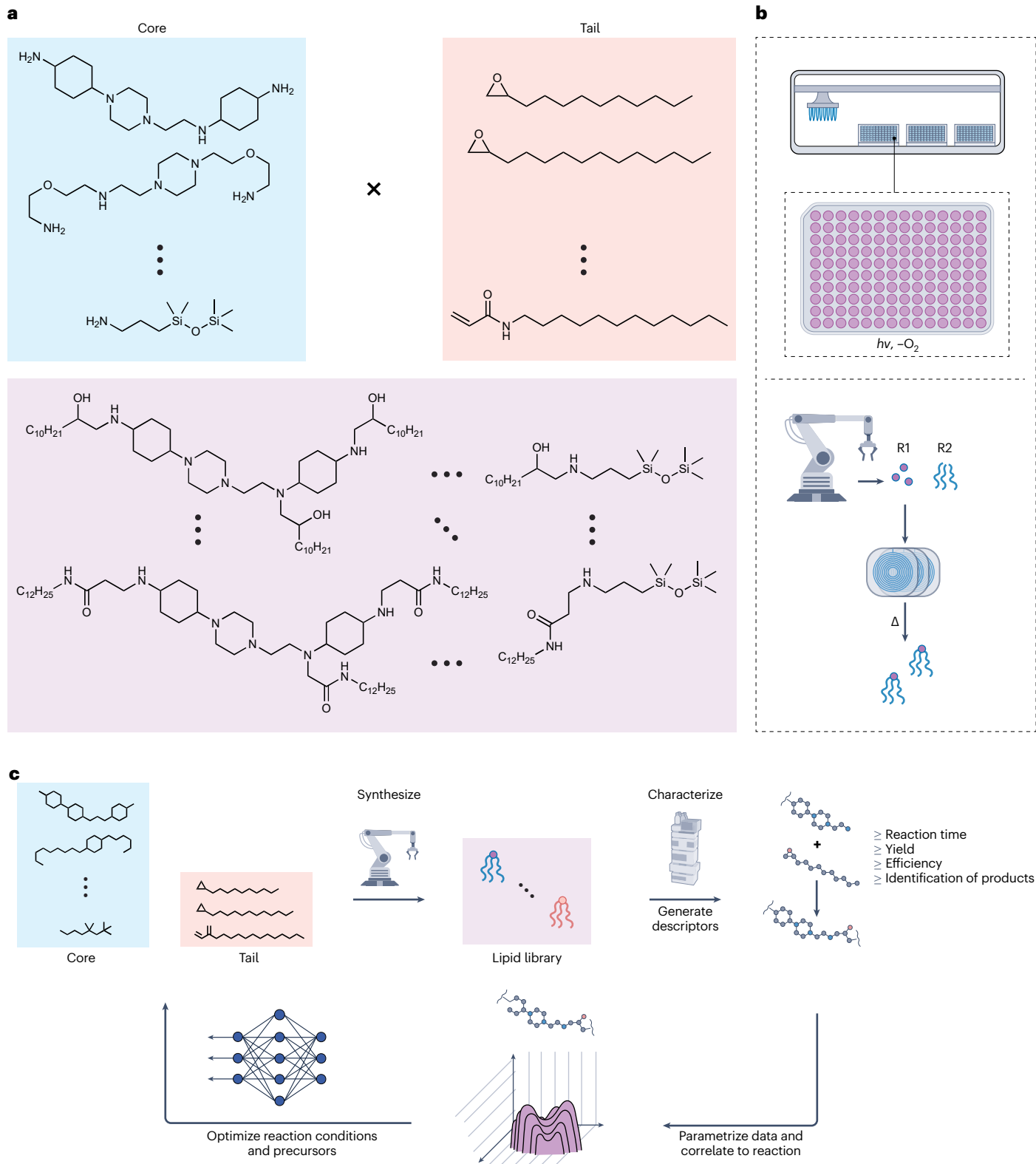
Ionizable lipid synthetic schemes are widely amenable to existing platforms for high-throughput materials generation, such as plate-based liquid handlers, robotic chemists or automated modular flow cell schemes (Fig. 3b). Automated liquid handlers have been leveraged in some synthetic schemes, allowing them to synthesize libraries of hundreds of ionizable lipids^{70,77}. Such liquid handlers can be further modified to process reactions in air-free conditions^{4,78}. A plate automation strategy incorporating independently controllable light sources has also been used to synthesize large libraries of morpholines and other N-heterocycles⁷⁹, broadening the potential structural diversity of ionizable lipid cores. However, these plate-based synthetic strategies often lack the modularity required to incorporate in-line

Table 1 | Examples of variations in lipid structure and overall LNP composition across validated formulations

Formulation name	Ionizable lipid structure	Lipid molar ratio	Application
C14-494 (ref. 37)		35% C14-494 46.5% cholesterol 16% DOPE 2.5% 14:0 PEG2000 PE	Ex vivo T cell transfection
C12-200 (ref. 47)		35% C12-200 46.5% Cholesterol 16% DOPE 2.5% 14:0 PEG2000 PE	RNA delivery to the liver
306-N16B (ref. 51)		50% 306-N16B 38.5% cholesterol 10% DOPC 1.5% DMG-PEG 2000	Cas9 delivery to the lung
SM102 (Moderna)		50% SM-102 38.5% Cholesterol 10% DSPC 1.5% DMG-PEG 2000	mRNA vaccine against SARS-CoV-2

DOPC, 1,2-dioleoyl-*sn*-glycero-3-phosphocholine; DMG-PEG, 1,2-dimyristoyl-*rac*-glycero-3-methoxypolyethylene glycol-2000; DOPE, 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine; DSPC, 1,2-distearoyl-*sn*-glycero-3-phosphocholine; LNP, lipid nanoparticles; PEG, polyethylene glycol; PEG2000 PE, 1,2-dimyristoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[methoxy(polyethylene glycol)-2000].

Perspective



characterization or purification, are poorly suited for handling volatile solvents, and are limited in their scalability (both in reaction volume and in number of parallel syntheses)⁸⁰.

Mobile robotic chemists that handle laboratory equipment could alleviate some of the limitations of synthesis using plate robotics^{81,82}. These robots can perform many of the same synthetic tasks that a

Fig. 3 | Strategies for the automation of lipid synthesis. The combination of robotics and machine learning has facilitated a closed-loop system for chemical synthesis, wherein products can be generated and characterized to determine the optimal reaction conditions for a given reagent and to suggest new reagents for faster or higher-yield syntheses. **a**, Depiction of potential ionizable lipids as the outer product of 'core' structures and 'tail' structures. **b**, Examples of high-throughput lipid synthesis platforms, including automated plate synthesis and robotically enabled synthesis in flow. Certain automated plate interfaces are now capable of performing photocatalysed reactions ($h\nu$) or performing reactions in an air-free environment ($-O_2$). Automation incorporating modular fluidic units can combine different reagents (R1 and R2) in flow under heat (Δ) or

microwave to accelerate the rate of reaction. **c**, Workflow diagram of the closed-loop optimization of ionizable lipid synthesis, wherein lipids are synthesized autonomously, characterized and featured for inclusion in a machine learning model. Automation facilitates the synthesis of different lipids with varying reaction parameters, such as temperature, solvent or catalysts, while optimizing for factors such as reaction yield or efficiency. This cycle continues until some output criterion, such as a certain yield, is sufficiently satisfied or optimized. Closed-loop automation coupled with machine learning enables the rapid optimization of the synthetic strategy to generate new potential lipid nanoparticle candidates.

human can but, like a human, they are limited by large space requirements and throughput. Given these drawbacks, these robots have not been widely applied to the synthesis of new lipids. To reduce the necessary operating footprint and increase synthetic throughput, robotic arms which interchange modular fluidic units have been developed^{66,83}. These fluidic units allow for heating and cooling of reagents, as well as for in-line characterization and work-up, allowing synthesis to scale easily. Fluidic automation has been coupled with ML-based optimization schemes to assist with reagent acquisition, synthesis planning and tuning of conditions such as temperature and reaction time⁶⁶. Data from the downstream analysis of the synthetic products produced by these systems can be fed into ML-based feedforward optimization schemes, wherein output variables such as reaction yield and purity are used to dynamically adjust reaction conditions and to inform future related reactions^{66,82–86}. Using dynamically optimized reaction conditions or synthesis planners, higher yield syntheses can be performed at lower cost⁸⁷ and with lower environmental impact. Open-access resources such as ASKCOS have expanded the use of models already pretrained on these reaction outputs.

Given the relative simplicity of most ionizable lipid syntheses, these automated platforms can massively accelerate the rate at which new compounds can be synthesized. Standard off-the-shelf liquid handlers may be able to perform thousands of two-component, single-step reactions per day. However, to access a broader chemical space, throughput drastically declines and the complexity of automation may increase with each additional reaction step necessary. Although there are still extensive swaths of chemical space in single-step, multicomponent reactions that have yet to be explored¹⁴, an important advantage from the application of ML models to lipid synthesis is the generative prediction of lipids from previously unexplored design spaces, potentially involving more complicated synthetic strategies. More modular synthetic instrumentation, such as modular fluidic reactors⁶⁶, could widen the accessible chemical space, but throughput would decrease drastically owing to a lack of parallelization. Furthermore, downstream screening processes such as particle formulation and screening are still probably rate limiting at their current throughput. Thus, developments in the application of automation to lipid synthesis should carefully consider the current balance between increasing lipid complexity and throughput, as well as downstream bottlenecks in the screening process.

Purification and characterization methods

Despite advances in the parallelization and automation of synthesis, purification and characterization still challenge lipid generation throughput. Although protocols for the purification of some components, such as steroids, are well established, the amphiphilic character

of the ionizable lipid, PEG lipid and phospholipid can render these compounds difficult to purify at high throughput; for example, traditional purification techniques, such as column chromatography, are often inadequate, resulting in broad elution peaks comprising multiple species. Characterization of ionizable lipids in particular often requires careful buffer consideration and the addition of charge-stabilizing agents such as ammonium hydroxide^{69,72}. In the automated modular flow example mentioned above, processes involving a greater number of processing steps may introduce substantially more complicated fluidic distribution systems, thus hindering throughput⁶⁶.

Nonetheless, improvements to in-line characterization techniques have assisted in the feedforward optimization of lipid syntheses (Fig. 3c). High-throughput liquid chromatography–mass spectrometry (LC–MS)⁸⁸ has been used for the characterization of ionizable lipids using charged-aerosol detection (CAD), a high-sensitivity method of detecting compounds that lack a UV-detectable chromophore^{89,90}. For example, CAD-based high-performance liquid chromatography (HPLC) has been applied towards the high-throughput analysis of lipids and lipid-like materials in large-scale lipidomic studies of plasma⁹¹. Emerging developments in high-throughput characterization include methods of incorporating HPLC data into feedforward decision-making schemes⁹². These systems have been coupled with automated nuclear magnetic resonance (NMR) set-ups^{93,94} for more thorough characterization. ML has been increasingly useful for the optimization both of reaction schemes to yield simpler purification⁶⁶ and for automating structure identification from MS and NMR spectra⁹⁵.

LNP formulation and characterization

Although techniques for the automated and parallelized synthesis of many lipids have existed since the early 2000s, a current limitation in the throughput of LNP screening is the controlled mixing of the organic phase solubilizing the lipids with an acidic, aqueous buffer bearing the nucleic acid cargo. Current efforts to improve the throughput of this mixing process to generate distinct LNP formulations focus on adapting existing automation, such as liquid handlers, to mix together organic and aqueous phases, developing new interfaces between mixing devices and many distinct fluidic inputs, or implementing parallelized architectures for the distribution of fewer fluidic inputs to generate many formulation outputs. New technologies must deliver not only improved throughput but also high batch-to-batch consistency and the ability to use the same mixing technique across all scales of clinical development.

High-throughput mixing methods

Even though individual lipid components can be generated at high throughput, these components must be combined in a nanoprecipitation reaction with a nucleic acid cargo to generate an LNP.

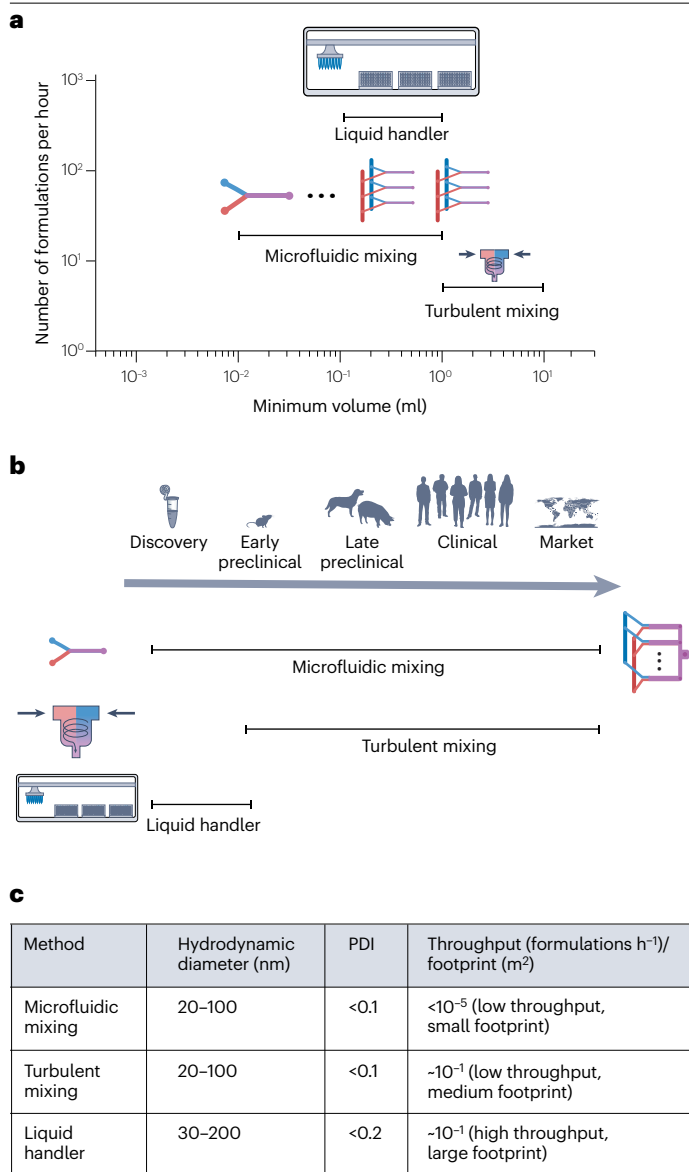


Fig. 4 | Throughput and scalability of mixing methods for LNPs. Different mixing methods result in lipid nanoparticles (LNPs) of different structures and characteristics. **a**, Plot of the minimum volume possible for each LNP formulation method compared with potential throughput. **b**, Evaluation of scale of production for each mixing method. Formulation reoptimization may be required if the formulation method used at discovery scale differs from those used at clinical or market scale. **c**, Summary of the differences in physicochemical outputs as they relate to the relative throughput of each mixing method. The footprint of each mixing method dictates how well a given mixing method could be parallelized given the spatial constraints of good laboratory practice or good manufacturing practice facilities. Although many mixing methods for LNPs exist at different scales of development, the ideal approach would combine rapid throughput in mixing distinct formulations with scalable production capabilities for clinical trial or market scale, all under similar local operating conditions. PDI, polydispersity index.

LNP nucleation typically involves nanoprecipitation of lipids in ethanol into a buffered aqueous countersolvent, wherein the ionizable lipid undergoes electrocomplexation with the cargo solubilized in the

aqueous phase⁹⁶. Nucleation that occurs at different Reynolds numbers (correlating to the laminarity or turbulence of a flow regime) results in LNPs with different internal structures^{97,98} and ultimately different downstream biological effects^{99,100}.

The most common mixing methods used for the high-throughput formulation of LNPs include bulk precipitation¹⁰¹, turbulent mixing^{102,103} and chaotic advective microfluidic mixing^{104,105} (Fig. 4a). Bulk precipitation, typically seen in pipette mixing, can generate LNPs at high throughput using automated liquid handlers^{101,106,107}, but mixing is intrinsically heterogeneous and inefficient, resulting in batch-to-batch variations in nanoparticle physicochemical properties^{99,101,108}. Turbulent mixers such as confined impinging jet mixers have been used far more widely at high volumes in polymeric and lipid nanoparticle applications, and are often capable of achieving mixing times below one millisecond¹⁰⁹. However, turbulent mixers are generally difficult to use at low volumes¹¹⁰, and thus suboptimal for generating large libraries of particles owing to material limitations (Fig. 4a). Furthermore, intrinsic nonlinearities and flow rate variance in scaling turbulent flow result in the unpredictability of local flow conditions, meaning that though mixing is occurring rapidly across the entire solution, mixing is not necessarily even throughout¹¹¹.

As the mixing environment has an important effect on the physicochemical and ultimately biological properties of LNPs, an ideal LNP formulation system would allow for identical local mixing conditions at every step of the drug development process, from discovery to batch scale¹⁰⁷ (Fig. 4b). Particles across the device should ‘experience’ the same mixing conditions at all the flow rates necessary for development, as a particle produced differently at the discovery and scale up phases may behave differently in a biological setting. For scale-up of LNP production in industrial settings, flash nanoprecipitation¹¹², a formulation method involving highly turbulent mixing, is widely used for its design simplicity and robustness to continuous processing. However, turbulent mixing techniques often change total flow rates through the mixer dependent on the amount of formulation being made, thus resulting in potentially different local mixing conditions for different particle batch sizes. Liquid handlers could theoretically scale to higher volumes by parallelizing the mixing across multiple wells, but they suffer from the same variabilities experienced in turbulent regimes. Additionally, scaling to higher volumes using liquid handlers requires the use of different pipettes, therefore different geometries and thus mixing environments, to operate effectively across all scales of development. Microfluidic mixing techniques are capable of generating LNPs in quantities relevant for small animal testing^{18,68}. When this mixing architecture is parallelized on the same device, the production volumes required for larger animal studies and human-scale applications can be met^{104,105}. Thus, microfluidic mixing is the only scale-invariant technique of those mentioned here.

Microfluidic mixing techniques offer high batch-to-batch consistency, but they are often fairly low throughput owing to inherent flow rate limitations through micrometre-scale devices and challenges in rapidly interfacing many distinct solutions through the same device¹¹³ (Fig. 4c). A number of microfluidic mixing devices have been used to generate LNPs^{114–118}. These devices often result in thorough mixing, but their performance tends to decline at higher flow rates required for clinical-scale particle production¹¹⁹; typical operating flow rates for bifurcating mixers range from 0.6 ml min⁻¹ (ref. 97) to 20 ml min⁻¹ (NxGen technology at Cytiva), which are flow rates sufficient for generating enough material at discovery to early preclinical scales. However, to generate diverse LNP formulations, fluidic inputs bearing different

lipids must serially interface with the same mixing device to generate new products. To improve throughput, scaffolds have been developed for connecting multiple fluidic inputs to the same chip¹¹³ and input delivery and output collection have been automated¹²⁰. Some schemes for microfluidic automation exist, producing LNP formulations on the order of ten formulations per hour (**Unchained Labs Sunscreen**), yet this technology does not currently allow for generation and screening of multiple nanoparticle formulations at the rate of an automated liquid handler. An emerging automated, microfluidic platform using parallelized delivery of single fluidic inputs but at different stoichiometric ratios to many independent mixing units has scaled in throughput beyond that of automated liquid handlers, to a throughput of around one distinct LNP formulation every 3 s, while delivering the precision mixing capabilities of microfluidics¹²¹.

Emerging and future work to accelerate formulation rates should include additional means of monitoring the mixing process, enabling the engineering of highly consistent, scalable mixing devices wherein distinct lipid and nucleic acid inputs can be automatically interfaced with a mixing device and many different stoichiometric ratios can be formulated simultaneously.

High-throughput characterization

High-throughput LNP generation demands matching throughput in characterization to determine whether each batch of formulation falls within a set of standards of quality. Metrics to evaluate formulated products include particle size and polydispersity¹²². Most commonly, dynamic light scattering or optical tracking methods are used for determining LNP size. For example, a technology similar to nanoparticle tracking analysis but performed in-line on a microfluidic chip has been developed¹²³. It is also important to determine the relative encapsulation of nucleic acid in the nanoparticle, most commonly through fluorescence assays such as the RiboGreen. Parts of the assay have been automated (**Unchained Labs Sunscreen** and **Unchained Labs Stunner**), yet separate robotic units are required for formulation and dialysis, requiring transfer between each.

Although size and encapsulation efficiency are useful metrics, advanced characterization techniques are also needed to further probe core structure, some of which have been adapted for high-throughput characterization. By coupling liquid handlers with a beamline, high-throughput small angle X-ray scattering (SAXS)¹⁰⁶ has been performed on a library of LNPs, providing data about lipid packing, core density and core phase in an automated manner at a throughput of around one formulation per minute. To enable in-line formulation, processing and characterization, additional workflows have integrated SAXS directly after asymmetric field flow fractionation (AF4) – a commonly used downstream processing step for LNPs to remove excess organic solvent¹²⁴. However, the generally slow throughput of AF4 combined with the complexity of SAXS data means that the latency between LNP formulation and analysis can be up to 40 min.

Although characterization techniques to confirm the structure and determine the purity of individual lipid components have been largely adapted from existing workflows, analysing the compositional ratios of LNP at the single-particle level remains a challenge. Single LNPs provide insufficient signal for standard chemical characterization techniques such as HPLC or NMR. To address this, microfluidic platforms have been used for in-line optical detection, allowing high-throughput assessment of size, relative composition and number of RNA per LNP^{125,126}. In these workflows, fluorescently labelled lipid and cargo components are separated in a micrometre scale capillary

and analysed with fluorescent coincidence detection at a throughput of 3,000–5,000 events per minute. Additionally, solution-phase Raman analysis has allowed for high-throughput quantification of relative lipid amounts¹²⁷ and quantification of particle surface ligand in antibody-coated LNPs (**SPARTA analysis of LNPs**). To move beyond labelled approaches, mass spectrometry has emerged as an alternative method to determine the number of mRNA per particle^{128,129}. At the bulk scale, compositional distributions across entire LNP formulations have been assessed using HPLC coupled with charged aerosol detection¹³⁰.

Although few feedforward ML-driven optimization schemes exist¹³¹, data-driven analyses of the relation between lipid ratios and physicochemical properties have been performed^{61,62,132}. Autonomous systems wherein the output of a system continuously informs input conditions have been widely adopted in applications from chemical synthesis⁸⁴ to inorganic materials generation¹³³ and droplet microfluidics¹³⁴. Rapid in-line characterization techniques may be used in the future to optimize physical conditions, such as total flow rate, or chemical conditions, such as the selection and combination of lipid components, for the production of LNPs with a desired set of properties. Currently, no physicochemical characterization method provides the same depth of insight into LNP performance as those available in analogous fields such as inorganic materials synthesis, wherein diffraction-based methods have enabled closed-loop discovery of new materials^{135–137}. However, larger studies involving advanced solution-based characterization methods combined with ML-driven data analysis may elucidate design rules relating LNP performance in biological systems to formulation input parameters, presenting a path forward for closed-loop LNP optimization.

In vitro and in vivo characterization

In vitro and in vivo screening data correlate most strongly with the therapeutic success of an LNP and, thus, provide the most relevant data for training ML models. In vitro metrics include toxicity against a particular cell type¹³⁸, transfection or expression efficiency^{139–142} (**Unchained Labs Sunscreen**), rate of transcytosis¹⁴³, identity of protein corona^{54,144}, and rate of cell association¹²² (NxGen technology). Although there is only a minimal correlation between in vitro and in vivo performance, in vitro metrics can help to narrow the potential design space to be tested in animal models, by identifying which particles have little to no capacity for transfection²⁴. Methods for screening formulated particles in a pooled fashion, such as LNP barcoding, have drastically accelerated the throughput of in vivo assays (Fig. 5a).

High-throughput in vitro assays

The most general paradigm for scaling the throughput of in vitro assays involves the automation of liquid handling using plate robotics, flow-based assays, or optical techniques. For example, one common early-stage screen involves the quantification of a luminescence reporter – a proxy for relative LNP transfection – using automated reagent and plate handling¹⁴⁵. Plate automation can also be used in flow-based assays to interchange samples for high-resolution in vitro analysis of LNP effects. Highly multiplexed flow cytometry has been leveraged to obtain time-resolved single-cell measurements of LNP transfection¹⁴¹. Mass cytometry can multiplex detection of mRNA transcripts and proteins within single cells to identify transcriptomic alterations brought about by LNP transfection^{139,146}. Optical methods have also been used for single-cell measurements¹⁴⁷. For example, confocal microscopy has been used to quantify fluorescent protein

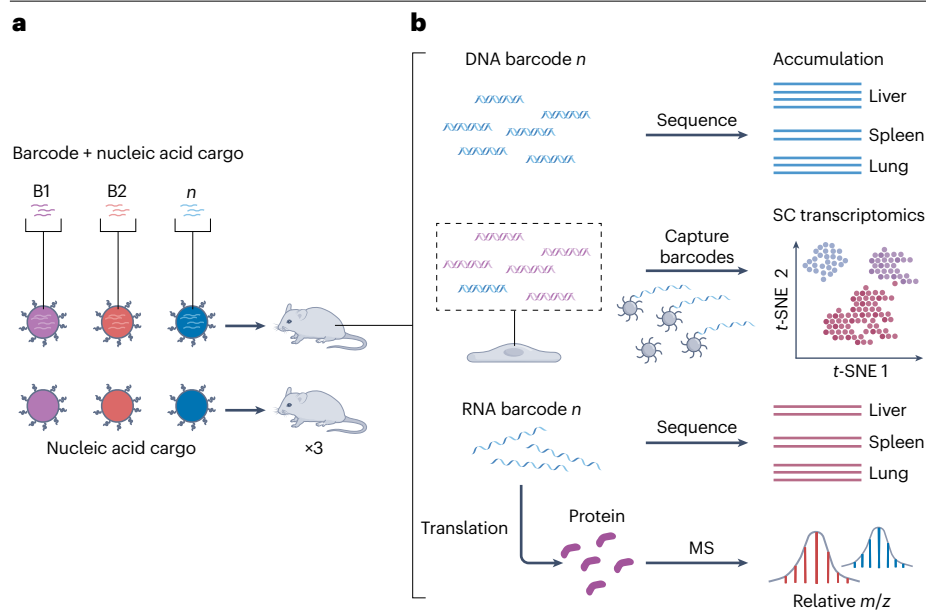


Fig. 5 | Options for parallelizing the in vivo screening of LNPs. **a**, Experimental workflow using barcoded samples and non-barcoded samples, wherein barcoded samples can be pooled for delivery to a single animal. **b**, Potential types of, and applications for, lipid nanoparticle (LNP) barcodes. The most common types of barcodes include nucleic acid barcodes, such as DNA or RNA barcodes. Their accumulation in different cell or tissue types is often quantified using deep sequencing, or in the case of RNA barcodes, their translated product can be identified and quantified using mass spectrometry (MS). **c**, Barcode types and relevant characteristics. Determining the optimal parallelized in vivo barcoding strategy, thus, relies on the practical degree of parallelization of each barcoding technique within a single animal, the ease and scalability of readout method, and the resolution of the readout required for each technique. *m/z*, mass-to-charge ratio; MS/MS, tandem mass spectrometry; NGS, next-generation sequencing; SC, single-cell; $t_{1/2}$, half-life; t-SNE, *t*-distributed Stochastic Neighbour Embedding.

c

Type of barcode	Method of analysis	Median barcode stability	Multiplexing capability (barcodes per animal)	Resolution
DNA	NGS	$t_{1/2}$ = -days	>100	Single cell
RNA	NGS	$t_{1/2}$ = -hours	>100	Single cell
Peptide	MS/MS	$t_{1/2}$ = -days to >1 wk	>100	Tissue or organ
Chemical	MS; imaging	$t_{1/2}$ = hours	~10	Single cell

products expressed from nucleic acid cargo (or fluorescent dyes), assess the relative transfection efficiency of single cells¹⁴⁸, evaluate cytotoxicity¹⁴⁹, or image potential membrane damage^{138,140}. For the use of ex vivo cell engineering, as in the case of CAR T cells or natural killer cells¹⁵⁰, these higher-throughput readouts are immediately applicable for identifying top-performing formulations. In the screening process for the discovery of in vivo therapeutics, in vitro assays are used largely to eliminate potential LNP formulations on the grounds of poor transfection efficiency. However, improving the throughput of higher resolution in vitro assays may provide new or supplementary information which can help to prevent against the inclusion of false positives or negatives in later animal model screens.

LNP barcoding for high-throughput in vivo analysis

Since the early 2020s, LNP barcoding has been increasingly leveraged to enable the simultaneous testing of multiple formulations in the same animal^{151,152}. LNP barcodes primarily consist of DNA¹⁵¹, mRNA¹⁵³, protein products from RNA transcripts¹⁵⁴ or distinct molecules^{155,156} (Fig. 5b,c). By leveraging next-generation sequencing and bioinformatic tools, nucleic acid barcoding of LNPs enables the assessment of biodistribution and translational efficiency of individual LNPs within a pool (Fig. 5c). Each 'barcode' is a nucleic acid sequence that can be isolated

from genomic DNA¹⁵¹. Together with a unique molecular identifier, these sequences can be demultiplexed, and their relative abundance in different cell and tissue types can be quantitated^{151,153}. Barcode analysis has been used to study LNP tropism^{151,157} – the relative accumulation of LNPs in different organ types – along with cellular association and uptake within different organs^{158,159}, cellular internalization pathways, and has even been combined with single-cell RNA sequencing to obtain multi-omics readouts from individual cells^{158,160}. Among nucleic acid barcodes, DNA barcodes were the first to be used to screen pooled LNPs¹⁵¹. In this approach, part of the therapeutic cargo is replaced with a DNA barcode, allowing many different barcodes to be used across LNPs carrying identical therapeutic cargo. RNA barcodes offer comparable multiplexing capability and readout resolution, but they generally have shorter half-lives than DNA barcodes (Fig. 5c). Structural and chemical modifications can be made to mRNA to prolong its half-life, but the biological implications of these modifications have not been widely assayed across cell types¹⁶¹. Although DNA barcodes are more stable and can be more easily interchanged than RNA barcodes¹⁵¹, the inclusion of a DNA barcode within LNPs already carrying RNA as the therapeutic cargo may result in physicochemical differences to the LNP, potentially altering biological data for a given formulation¹⁵³. Peptide barcodes provide an additional level of detail from RNA barcodes in that they

identify not only relative accumulation of a nucleic acid but its relative activity within a given population¹⁵⁴. However, because the primary readout of peptide barcodes involves tandem mass spectrometry, achieving resolution beyond the tissue scale remains low through-put owing to the time-consuming steps of cell lysis and transfer into the MS inlet. Finally, small molecules have been increasingly used as barcodes for nanoparticles^{155,156}. Although the theoretical structural landscape and, thus, potential number of small-molecule barcodes are massive, these species must often be detectable either optically – thus suffering from spectral overlap¹⁶² – or by using analytical techniques such as tandem mass spectrometry or gas chromatography–mass spectrometry, wherein they are often difficult to distinguish from the molecular background signal of the cell and have a high limit of detection relative to nucleic acids¹⁵⁵. Strategies incorporating chemically distinct aryl-halide-based compounds called halocodes¹⁵⁵ show low to mid-nanomolar limits of detection; however, the ability to multiplex beyond around ten samples has yet to be shown.

Thus far, the number of barcoded LNPs tested per animal experimentally falls short of the theoretical limit dictated by the information density of the barcode. Although an eight-nucleotide sequence can generate $4^8 = 65,536$ unique barcodes, LNP barcoding experiments have only screened on the order of 10^2 barcodes per animal^{163,164}. Although error prevention techniques, such as differentiating barcode sequences at multiple positions, reduce the number of usable barcodes¹⁶⁵, the practical number of barcoded particles used at any one time is probably further limited by physicochemical disparities between particles, which may cause interparticle interactions or ‘drug–drug’ interactions in vivo. Additionally, limitations in the throughput of other parts of the LNP formulation process are prohibitive to the screening of greater numbers of barcoded LNPs at a single time.

Largely as a result of this scalability and the ubiquity of next-generation sequencing, nucleic acid barcoding has been the most widely used form of parallelized in vivo screening for LNPs. Nonetheless, as mass spectrometry or imaging-based readouts improve in throughput, the scalability of techniques such as peptide barcoding, which would provide more resolution in terms not only of LNP localization but of cargo translation, or chemical barcoding, wherein intracellular localization could be visualized, may improve.

Ultimately, because in vivo and in vitro environments fundamentally differ, in vitro data – such as transfection or cytotoxicity – are limited in their capacity to serve as training data for ML models used to predict in vivo LNP behaviour. Although this lack of correlation is well profiled, ML models capable of integrating highly multimodal data sets¹⁶⁶ could undergo multiple rounds of fine tuning, first on large ($>10^3$ – 10^4 samples) in vitro data sets, upstream of smaller in vivo data sets. As seen in an active learning approach¹³¹, multiple rounds of in vitro screening could be used to iteratively improve the predictive capacity of the model before moving to in vivo testing. As the automation and parallelization of LNP formulation improve, these approaches could be more increasingly leveraged to minimize resource and time costs.

Outlook

High-throughput screening has substantially accelerated the development of LNPs for a broadening set of therapeutic indications. These advancements have been enabled by recent advances increasing screening throughput, despite the many complex processes involved from lipid synthesis to in vivo testing. As more advanced characterization techniques are developed, the data sets generated by LNP screens are

becoming increasingly large and multifactorial, rendering computational tools more useful in understanding how the inputs for a given formulation – that is, the different lipid structures, the ratio of lipids, and the mixing method – alter its biological effects.

ML models are now being leveraged in the discovery of new LNP formulations. Model architectures such as support vector machines, random forests or gradient boosting, trained on relatively small training data sets (tens to hundreds of samples) of largely physicochemical and in vitro data, have been used to predict the transfection efficiency of new formulations, as shown across the polymeric nanoparticle^{167–169} and increasingly the LNP literature^{14,61,62}. These applications demonstrate the success of these models in predicting some output variable such as transfection efficiency across a small number of input features – for example, compositional variables such as phospholipid-to-PEG ratio, and the type of phospholipid/sterol used. Explanatory frameworks such as Shapley additive explanations are further able to deconvolute the relevance of distinct features as they relate to biological performance such as transfection efficiency^{61,62}.

Increasingly, far more data-intensive neural network-based models^{61,131,166} have been used to predict entirely new ionizable lipid structures. These models reference embed lipid structure as some high-dimensional representation, commonly a graph, which is coupled to some experimental output such as transfection efficiency, to fine-tune an existing neural network which has been trained on millions of virtual chemical compounds. Thus far, in the LNP space, ML models pretrained on physicochemical and in vitro data from hundreds to thousands of samples have been used to fine-tune these models to ultimately generate new ionizable lipid structures^{61,131,169}. These models have either leveraged large existing data sets or been used to develop new automation to generate and screen sample sizes large enough to fine-tune neural networks largely pretrained on existing graph-based molecular descriptors. In future iterations, these models might be able to suggest new lipid structures and their relative ratios might be able to generate an LNP with a particular function in vivo^{37,154}.

Ultimately, developing models for the function-led generative design of LNPs using neural networks would require the generation of large libraries of sufficient size, at least thousands of samples each, to be tested across multiple relevant biological models^{166,170,171}. Given the inherent inconsistency of most biological data in comparison to physicochemical data, future predictive models will probably require larger data sets than are currently generated in single studies. Although discrepancies in the methods used to generate a data set will always exist, the number of different factors – lipid purity, mixing method and post-formulation processing, to name a few – that can confound biological performance is drastically higher in LNPs than in small-molecule development, wherein ML models have successfully improved drug performance^{172,173}. Moving towards the generation of open-source predictive frameworks that, for a given biological function, can suggest lipids and their relative ratios for an LNP formulation would require both that data be generated in a more standardized fashion and that source data be made more widely available. Examples of such standardized databases in formats ready for use in learning applications are present in the polymer literature^{174,175}. Such collaborative initiatives would incentivize the development of better standardized methods of synthesis, characterization, formulation and screening to generate publicly available data sets for future LNP discovery. Resources for ML implementation in nanoparticle and biomaterials applications^{13,176,177} can assist in the cleaning and encoding of these data. Finally, as automation for LNP synthesis and formulation improves and becomes more

widely available, the use of such methods may further streamline the process of experimental standardization.

As of now, the LNP design space must be screened for each distinct biological application. For example, the optimal LNP for lung gene editing^{44,50,178} is entirely distinct from the optimal LNP for in vivo delivery to macrophages or monocytes^{68,179}. Nonetheless, advancements in the automation of lipid synthesis and LNP formulation will decrease the timescale necessary to perform these screens. Ultimately, consistently generated, publicly available data may be used to train models through multi-objective learning, which might be trained on transfection efficiency, tissue tropism or even more detailed multi-omics readouts to generate a ‘foundation model’ that can best predict LNPs for a particular desired biological response. These generative models, which might suggest new lipid structures or stoichiometries, could be heavily weighted to suggest lipids that are chemically distinct from previous lipids, while simultaneously assigning high weight to their ‘synthesizability’, potentially by integrating existing software to suggest reagents and possible synthetic strategies.

Although each therapeutic or prophylactic indication requires different biological properties from a delivery vehicle, the use of ML models for the prediction of new LNP formulations substantially narrows the potential design space for a given application, drastically shortening the discovery and early development process for each drug. Combining automation with these computational tools will enable the development of LNPs for a broad set of unmet clinical needs and enable a much more rapid response to emerging infectious diseases in the future.

Published online: 08 September 2025

References

- Anderson, D. G., Lynn, D. M. & Langer, R. Semi-automated synthesis and screening of a large library of degradable cationic polymers for gene delivery. *Angew. Chem. Int. Ed.* **42**, 3153–3158 (2003).
- Breunig, M., Lungwitz, U., Liebl, R. & Goepferich, A. Breaking up the correlation between efficacy and toxicity for nonviral gene delivery. *Proc. Natl Acad. Sci. USA* **104**, 14454–14459 (2007).
- Lynn, D. M. & Langer, R. Degradable poly(β -amino esters): synthesis, characterization, and self-assembly with plasmid DNA. *J. Am. Chem. Soc.* **122**, 10761–10768 (2000).
- Sieglwart, D. J. et al. Combinatorial synthesis of chemically diverse core-shell nanoparticles for intracellular delivery. *Proc. Natl Acad. Sci. USA* **108**, 12996–13001 (2011).
- Wightman, L. et al. Different behavior of branched and linear polyethylenimine for gene delivery in vitro and in vivo. *J. Gene Med.* **3**, 362–372 (2001).
- Dahlman, J. E. et al. In vivo endothelial siRNA delivery using polymeric nanoparticles with low molecular weight. *Nat. Nanotechnol.* **9**, 648–655 (2014).
- Patra, J. K. et al. Nano based drug delivery systems: recent developments and future prospects. *J. Nanobiotechnol.* **16**, 71 (2018).
- Anselmo, A. C. & Mitragotri, S. Nanoparticles in the clinic: an update. *Bioeng. Transl. Med.* **4**, e10143 (2019).
- Maurer, M. S. et al. Patisiran treatment in patients with transthyretin cardiac amyloidosis. *N. Engl. J. Med.* **389**, 1553–1565 (2023).
- Hassett, K. J. et al. Optimization of lipid nanoparticles for intramuscular administration of mRNA vaccines. *Mol. Ther. Nucleic Acids* **15**, 1–11 (2019).
- Verma, M. et al. The landscape for lipid-nanoparticle-based genomic medicines. *Nat. Rev. Drug Discov.* **22**, 349–350 (2023).
- Dehghani-Ghahnaviyeh, S. et al. Ionizable amino lipids distribution and effects on D5PC/cholesterol membranes: implications for lipid nanoparticle structure. *J. Phys. Chem. B* **127**, 6928–6939 (2023).
- McDonald, S. M. et al. Applied machine learning as a driver for polymeric biomaterials design. *Nat. Commun.* **14**, 4838 (2023).
- Li, B. et al. Accelerating ionizable lipid discovery for mRNA delivery using machine learning and combinatorial chemistry. *Nat. Mater.* **23**, 1002–1008 (2024).
- Han, X. et al. Fast and facile synthesis of amidine-incorporated degradable lipids for versatile mRNA delivery in vivo. *Nat. Chem.* **16**, 1687–1697 (2024).
- Han, X. et al. Optimization of the activity and biodegradability of ionizable lipids for mRNA delivery via directed chemical evolution. *Nat. Biomed. Eng.* **8**, 1412–1424 (2024).
- Xue, L. et al. Combinatorial design of siloxane-incorporated lipid nanoparticles augments intracellular processing for tissue-specific mRNA therapeutic delivery. *Nat. Nanotechnol.* <https://doi.org/10.1038/s41565-024-01747-6> (2024).
- Han, X. et al. Adjuvant lipidoid-substituted lipid nanoparticles augment the immunogenicity of SARS-CoV-2 mRNA vaccines. *Nat. Nanotechnol.* **18**, 1105–1114 (2023).
- Han, X. et al. An ionizable lipid toolbox for RNA delivery. *Nat. Commun.* **12**, 7233 (2021).
- Cheng, Q. et al. Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing. *Nat. Nanotechnol.* **15**, 313–320 (2020).
- Dilliard, S. A. & Sieglwart, D. J. Passive, active and endogenous organ-targeted lipid and polymer nanoparticles for delivery of genetic drugs. *Nat. Rev. Mater.* **8**, 282–300 (2023).
- Mitchell, M. J. et al. Engineering precision nanoparticles for drug delivery. *Nat. Rev. Drug Discov.* **20**, 101–124 (2021).
- Billingsley, M. M. et al. Orthogonal design of experiments for optimization of lipid nanoparticles for mRNA engineering of CAR T cells. *Nano Lett.* **22**, 533–542 (2022).
- Lokugamage, M. P. et al. Optimization of lipid nanoparticles for the delivery of nebulized therapeutic mRNA to the lungs. *Nat. Biomed. Eng.* **5**, 1059–1068 (2021).
- Hou, X., Zaks, T., Langer, R. & Dong, Y. Lipid nanoparticles for mRNA delivery. *Nat. Rev. Mater.* **6**, 1078–1094 (2021).
- Patel, S. et al. Naturally-occurring cholesterol analogues in lipid nanoparticles induce polymorphic shape and enhance intracellular delivery of mRNA. *Nat. Commun.* **11**, 983 (2020).
- Lokugamage, M. P., Sago, C. D. & Dahlman, J. E. Testing thousands of nanoparticles in vivo using DNA barcodes. *Curr. Opin. Biomed. Eng.* **7**, 1–8 (2018).
- Zheng, L., Bandara, S. R., Tan, Z. & Leal, C. Lipid nanoparticle topology regulates endosomal escape and delivery of RNA to the cytoplasm. *Proc. Natl Acad. Sci. USA* **120**, e2301067120 (2023).
- Koltover, I., Salditt, T., Rädler, J. O. & Safinya, C. R. An inverted hexagonal phase of cationic liposome-DNA complexes related to DNA release and delivery. *Science* **281**, 78–81 (1998).
- Álvarez-Benedicto, E. et al. Optimization of phospholipid chemistry for improved lipid nanoparticle (LNP) delivery of messenger RNA (mRNA). *Biomater. Sci.* **10**, 549–559 (2022).
- Herrera, M., Kim, J., Eygeris, Y., Jozic, A. & Sahay, G. Illuminating endosomal escape of polymorphic lipid nanoparticles that boost mRNA delivery. *Biomater. Sci.* **9**, 4289–4300 (2021).
- Radmand, A. et al. Cationic cholesterol-dependent LNP delivery to lung stem cells, the liver, and heart. *Proc. Natl Acad. Sci. USA* **121**, e2307801120 (2024).
- Mui, B. L. et al. Influence of polyethylene glycol lipid desorption rates on pharmacokinetics and pharmacodynamics of siRNA lipid nanoparticles. *Mol. Ther. Nucleic Acids* **2**, e139 (2013).
- Gindy, M. E. et al. Mechanism of macromolecular structure evolution in self-assembled lipid nanoparticles for siRNA delivery. *Langmuir* **30**, 4613–4622 (2014).
- Kauffman, K. J. et al. Optimization of lipid nanoparticle formulations for mRNA delivery in vivo with fractional factorial and definitive screening designs. *Nano Lett.* **15**, 7300–7306 (2015).
- Akinc, A. et al. A combinatorial library of lipid-like materials for delivery of RNAi therapeutics. *Nat. Biotechnol.* **26**, 561–569 (2008).
- Billingsley, M. M. et al. Ionizable lipid nanoparticle-mediated mRNA delivery for human CAR T cell engineering. *Nano Lett.* **20**, 1578–1589 (2020).
- Whitehead, K. A. et al. Degradable lipid nanoparticles with predictable in vivo siRNA delivery activity. *Nat. Commun.* **5**, 4277 (2014).
- Haji, K. A. et al. Branched-tail lipid nanoparticles potently deliver mRNA in vivo due to enhanced ionization at endosomal pH. *Small* **15**, 1805097 (2019).
- Yonezawa, S., Koide, H. & Asai, T. Recent advances in siRNA delivery mediated by lipid-based nanoparticles. *Adv. Drug Deliv. Rev.* **154**, 64–78 (2020).
- Li, H. et al. Circular RNA cancer vaccines drive immunity in hard-to-treat malignancies. *Theranostics* **12**, 6422–6436 (2022).
- McKay, P. F. et al. Self-amplifying RNA SARS-CoV-2 lipid nanoparticle vaccine candidate induces high neutralizing antibody titers in mice. *Nat. Commun.* **11**, 3523 (2020).
- Palanki, R. et al. Ionizable lipid nanoparticles for therapeutic base editing of congenital brain disease. *ACS Nano* **17**, 13594–13610 (2023).
- Wei, T. et al. Lung SORT LNPs enable precise homology-directed repair mediated CRISPR/Cas genome correction in cystic fibrosis models. *Nat. Commun.* **14**, 7322 (2023).
- Thatte, A. S. et al. mRNA lipid nanoparticles for ex vivo engineering of immunosuppressive T cells for autoimmunity therapies. *Nano Lett.* **23**, 10179–10188 (2023).
- Billingsley, M. M. et al. In vivo mRNA CAR T cell engineering via targeted ionizable lipid nanoparticles with extrahepatic tropism. *Small* **20**, 2304378 (2024).
- Love, K. T. et al. Lipid-like materials for low-dose, in vivo gene silencing. *Proc. Natl Acad. Sci. USA* **107**, 1864–1869 (2010).
- Forsyth, V., Rzeczycki, P., Taylor, D. & Ariosa, A. Enhancing transfection efficiency of primary immune cells through lipid nanoparticle mediated delivery. *J. Immunol.* **212**, 0857_6526 (2024).
- Dilliard, S. A., Cheng, Q. & Sieglwart, D. J. On the mechanism of tissue-specific mRNA delivery by selective organ targeting nanoparticles. *Proc. Natl Acad. Sci. USA* **118**, e2109256118 (2021).
- Sun, Y. et al. In vivo editing of lung stem cells for durable gene correction in mice. *Science* **384**, 1196–1202 (2024).
- Qiu, M. et al. Lung-selective mRNA delivery of synthetic lipid nanoparticles for the treatment of pulmonary lymphangioliomyomatosis. *Proc. Natl Acad. Sci. USA* **119**, e2116271119 (2022).
- LoPresti, S. T., Arral, M. L., Chaudhary, N. & Whitehead, K. A. The replacement of helper lipids with charged alternatives in lipid nanoparticles facilitates targeted mRNA delivery to the spleen and lungs. *J. Control. Release* **345**, 819–831 (2022).

53. Omo-Lamai, S. et al. Physicochemical targeting of lipid nanoparticles to the lungs induces clotting: mechanisms and solutions. *Adv. Mater.* **36**, 2312026 (2024).
54. Walkey, C. D. et al. Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles. *ACS Nano* **8**, 2439–2455 (2014).
55. Walkey, C. D., Olsen, J. B., Guo, H., Emili, A. & Chan, W. C. W. Nanoparticle size and surface chemistry determine serum protein adsorption and macrophage uptake. *J. Am. Chem. Soc.* **134**, 2139–2147 (2012).
56. Cai, R. et al. Dynamic intracellular exchange of nanomaterials' protein corona perturbs proteostasis and remodels cell metabolism. *Proc. Natl Acad. Sci. USA* **119**, e2200363119 (2022).
57. Abdelkhalig, A. et al. Impact of nanoparticle surface functionalization on the protein corona and cellular adhesion, uptake and transport. *J. Nanobiotechnol.* **16**, 70 (2018).
58. Baimanov, D. et al. In situ analysis of nanoparticle soft corona and dynamic evolution. *Nat. Commun.* **13**, 5389 (2022).
59. Lee, Y. K., Choi, E.-J., Webster, T. J., Kim, S.-H. & Khang, D. Effect of the protein corona on nanoparticles for modulating cytotoxicity and immunotoxicity. *Int. J. Nanomed.* **10**, 97–113 (2015).
60. Mahmoudi, M., Landry, M. P., Moore, A. & Coreas, R. The protein corona from nanomedicine to environmental science. *Nat. Rev. Mater.* **8**, 422–438 (2023).
61. Xu, Y. et al. AGILE platform: a deep learning powered approach to accelerate LNP development for mRNA delivery. *Nat. Commun.* **15**, 6305 (2024).
62. Cheng, L. et al. Machine learning elucidates design features of plasmid deoxyribonucleic acid lipid nanoparticles for cell type-preferential transfection. *ACS Nano* **18**, 28735–28747 (2024).
63. Dormán, G. The rise, fall and revival of combinatorial chemistry. *Nachr. Chem.* **70**, 70–72 (2022).
64. Liu, Y. et al. Ultrastiff metamaterials generated through a multilayer strategy and topology optimization. *Nat. Commun.* **15**, 2984 (2024).
65. Ha, C. S. et al. Rapid inverse design of metamaterials based on prescribed mechanical behavior through machine learning. *Nat. Commun.* **14**, 5765 (2023).
66. Coley, C. W. et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).
67. Hashiba, K. et al. Branching ionizable lipids can enhance the stability, fusogenicity, and functional delivery of mRNA. *Small Sci.* **3**, 2200071 (2023).
68. Mukalel, A. J. et al. Oxidized mRNA lipid nanoparticles for in situ chimeric antigen receptor monocyte engineering. *Adv. Funct. Mater.* **34**, 2312038 (2024).
69. Riley, R. S. et al. Ionizable lipid nanoparticles for in utero mRNA delivery. *Sci. Adv.* **7**, eaba1028 (2021).
70. Chen, J. et al. Combinatorial design of ionizable lipid nanoparticles for muscle-selective mRNA delivery with minimized off-target effects. *Proc. Natl Acad. Sci. USA* **120**, e2309472120 (2023).
71. Fenton, O. S. et al. Synthesis and biological evaluation of ionizable lipid materials for the in vivo delivery of messenger RNA to B lymphocytes. *Adv. Mater.* **29**, 1606944 (2017).
72. Fenton, O. S. et al. Bioinspired alkenyl amino alcohol ionizable lipid materials for highly potent in vivo mRNA delivery. *Adv. Mater.* **28**, 2939–2943 (2016).
73. Hatit, M. Z. C. et al. Nanoparticle stereochemistry-dependent endocytic processing improves in vivo mRNA delivery. *Nat. Chem.* **15**, 508–515 (2023).
74. Da Silva Sanchez, A. J. et al. Substituting racemic ionizable lipids with stereopure ionizable lipids can increase mRNA delivery. *J. Control. Release* **353**, 270–277 (2023).
75. Li, Z. et al. Enzyme-catalyzed one-step synthesis of ionizable cationic lipids for lipid nanoparticle-based mRNA COVID-19 vaccines. *ACS Nano* **16**, 18936–18950 (2022).
76. Jacobsen, E. N. Asymmetric catalysis of epoxide ring-opening reactions. *Acc. Chem. Res.* **33**, 421–431 (2000).
77. Li, B. et al. Combinatorial design of nanoparticles for pulmonary mRNA delivery and genome editing. *Nat. Biotechnol.* **41**, 1410–1415 (2023).
78. Lee, J. et al. A fully automated platform for photoinitiated RAFT polymerization. *Digit. Discov.* **2**, 219–233 (2023).
79. Götz, J. et al. High-throughput synthesis provides data for predicting molecular properties and reaction success. *Sci. Adv.* **9**, ead3214 (2023).
80. Kong, F., Yuan, L., Zheng, Y. F. & Chen, W. Automatic liquid handling for life science: a critical review of the current state of the art. *SLAS Technol.* **17**, 169–185 (2012).
81. Burger, B. et al. A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
82. Dai, T. et al. Autonomous mobile robots for exploratory synthetic chemistry. *Nature* **635**, 890–897 (2024).
83. Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
84. Nambiar, A. M. K. et al. Bayesian optimization of computer-proposed multistep synthetic routes on an automated robotic flow platform. *ACS Cent. Sci.* **8**, 825–836 (2022).
85. Shields, B. J. et al. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).
86. Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **360**, 186–190 (2018).
87. Fromer, J. C. & Coley, C. W. An algorithmic framework for synthetic cost-aware decision making in molecular design. *Nat. Comput. Sci.* **4**, 440–450 (2024).
88. Kyranos, J. N., Cai, H., Wei, D. & Goetzinger, W. K. High-throughput high-performance liquid chromatography/mass spectrometry for modern drug discovery. *Curr. Opin. Biotechnol.* **12**, 105–111 (2001).
89. Ginsburg-Moraff, C. et al. Integrated and automated high-throughput purification of libraries on microscale. *SLAS Technol.* **27**, 350–360 (2022).
90. Su, W.-C. et al. A platform method for simultaneous quantification of lipid and nucleic acid components in lipid nanoparticles. *J. Chromatogr. A* **1746**, 465788 (2025).
91. Munjoma, N. et al. High throughput LC-MS platform for large scale screening of bioactive polar lipids in human plasma and serum. *J. Proteome Res.* **21**, 2596–2608 (2022).
92. Haas, C. P. et al. Open-source chromatographic data analysis for reaction optimization and screening. *ACS Cent. Sci.* **9**, 307–317 (2023).
93. Herck, J. V. et al. Operator-independent high-throughput polymerization screening based on automated inline NMR and online SEC. *Digit. Discov.* **1**, 519–526 (2022).
94. Ammini, G. D., Hooker, J. P., Herck, J. V., Kumar, A. & Junkers, T. Comprehensive high-throughput screening of photopolymerization under light intensity variation using inline NMR monitoring. *Polym. Chem.* **14**, 2708–2716 (2023).
95. Hu, G. & Qiu, M. Machine learning-assisted structure annotation of natural products based on MS and NMR data. *Nat. Prod. Rep.* **40**, 1735–1753 (2023).
96. Kulkarni, J. A. et al. On the formation and morphology of lipid nanoparticles containing ionizable cationic lipids and siRNA. *ACS Nano* **12**, 4787–4795 (2018).
97. Maeki, M. et al. Understanding the formation mechanism of lipid nanoparticles in microfluidic devices with chaotic micromixers. *PLoS ONE* **12**, e0187962 (2017).
98. Jung, H. N., Lee, S.-Y., Lee, S., Youn, H. & Im, H.-J. Lipid nanoparticles for delivery of RNA therapeutics: current status and the role of in vivo imaging. *Theranostics* **12**, 7509–7531 (2022).
99. Strelkova Petersen, D. M., Chaudhary, N., Arral, M. L., Weiss, R. M. & Whitehead, K. A. The mixing method used to formulate lipid nanoparticles affects mRNA delivery efficacy and organ tropism. *Eur. J. Pharm. Biopharm.* **192**, 126–135 (2023).
100. Padilla, M. et al. Solution biophysics identifies lipid nanoparticle non-sphericity, polydispersity, and dependence on internal ordering for efficacious mRNA delivery. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.12.19.629496> (2025).
101. Fan, Y. et al. Automated high-throughput preparation and characterization of oligonucleotide-loaded lipid nanoparticles. *Int. J. Pharm.* **599**, 120392 (2021).
102. Valente, I., Celasco, E., Marchisio, D. L. & Barresi, A. A. Nanoprecipitation in confined impinging jets mixers: production, characterization and scale-up of pegylated nanospheres and nanocapsules for pharmaceutical use. *Chem. Eng. Sci.* **77**, 217–227 (2012).
103. He, Z. et al. Size-controlled lipid nanoparticle production using turbulent mixing to enhance oral DNA delivery. *Acta Biomater.* **81**, 195–207 (2018).
104. Shepherd, S. J. et al. Throughput-scalable manufacturing of SARS-CoV-2 mRNA lipid nanoparticle vaccines. *Proc. Natl Acad. Sci. USA* **120**, e2303567120 (2023).
105. Shepherd, S. J. et al. Scalable mRNA and siRNA lipid nanoparticle production using a parallelized microfluidic device. *Nano Lett.* **21**, 5671–5680 (2021).
106. Hammel, M. et al. Correlating the structure and gene silencing activity of oligonucleotide-loaded lipid nanoparticles using small-angle X-ray scattering. *ACS Nano* **17**, 11454–11465 (2023).
107. Valencia, P. M., Farokhzad, O. C., Karnik, R. & Langer, R. Microfluidic technologies for accelerating the clinical translation of nanoparticles. *Nat. Nanotechnol.* **7**, 623–629 (2012).
108. Kroth, I., Karimov, M. & Karongo, R. Automated preparation of oligonucleotide-loaded lipid nanoparticles using Andrew™ pipetting robot for high-throughput in-vitro screening. *Waters* <https://www.waters.com/content/dam/waters/en/app-notes/2023/720008090/720008090-en.pdf> (2024).
109. Metzger, L. & Kind, M. On the mixing in confined impinging jet mixers — time scale analysis and scale-up using CFD coarse-graining methods. *Chem. Eng. Res. Des.* **109**, 464–476 (2016).
110. Shepherd, S. J., Issadore, D. & Mitchell, M. J. Microfluidic formulation of nanoparticles for biomedical applications. *Biomaterials* **274**, 120826 (2021).
111. Sreenivasan, K. R. & Antonia, R. A. The phenomenology of small-scale turbulence. *Annu. Rev. Fluid Mech.* **29**, 435–472 (1997).
112. Feng, J., Markwalter, C. E., Tian, C., Armstrong, M. & Prud'homme, R. K. Translational formulation of nanoparticle therapeutics from laboratory discovery to clinical scale. *J. Transl. Med.* **17**, 200 (2019).
113. Wilhelm, E., Neumann, C., Duttonhofer, T., Pires, L. & Rapp, B. E. Connecting microfluidic chips using a chemically inert, reversible, multichannel chip-to-world-interface. *Lab Chip* **13**, 4343–4351 (2013).
114. Ripoll, M. et al. Optimal self-assembly of lipid nanoparticles (LNP) in a ring micromixer. *Sci. Rep.* **12**, 9483 (2022).
115. Kimura, N. et al. Development of a microfluidic-based post-treatment process for size-controlled lipid nanoparticles and application to siRNA delivery. *ACS Appl. Mater. Interfaces* **12**, 34011–34020 (2020).
116. Chen, D. et al. Rapid discovery of potent siRNA-containing lipid nanoparticles enabled by controlled microfluidic formulation. *J. Am. Chem. Soc.* **134**, 6948–6951 (2012).
117. Bellevue, 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).
118. Leung, A. K. K., Tam, Y. Y. C., Chen, S., Hafez, I. M. & Cullis, P. R. Microfluidic mixing: a general method for encapsulating macromolecules in lipid nanoparticle systems. *J. Phys. Chem. B* **119**, 8698–8706 (2015).
119. Williams, M. S., Longmuir, K. J. & Yager, P. A practical guide to the staggered herringbone mixer. *Lab Chip* **8**, 1121–1129 (2008).
120. Longwell, S. A. & Fordyce, P. M. micrIO: an open-source autosampler and fraction collector for automated microfluidic input-output. *Lab Chip* **20**, 93–106 (2019).

121. Hanna, A. R. et al. Automated and parallelized microfluidic generation of large and precisely-defined lipid nanoparticle libraries. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.05.26.656157> (2025).
122. Nag, K. et al. DoE-derived continuous and robust process for manufacturing of pharmaceutical-grade wide-range LNPs for RNA-vaccine/drug delivery. *Sci. Rep.* **12**, 9394 (2022).
123. McLeod, E. et al. High-throughput and label-free single nanoparticle sizing based on time-resolved on-chip microscopy. *ACS Nano* **9**, 3265–3273 (2015).
124. Graewert, M. A. et al. Quantitative size-resolved characterization of mRNA nanoparticles by in-line coupling of asymmetrical-flow field-flow fractionation with small angle X-ray scattering. *Sci. Rep.* **13**, 15764 (2023).
125. Li, S. et al. Payload distribution and capacity of mRNA lipid nanoparticles. *Nat. Commun.* **13**, 5561 (2022).
126. Li, S. et al. Single-particle spectroscopic chromatography reveals heterogeneous RNA loading and size correlations in lipid nanoparticles. *ACS Nano* **18**, 15729–15743 (2024).
127. Penders, J. et al. Single particle automated Raman trapping analysis. *Nat. Commun.* **9**, 4256 (2018).
128. Lowenthal, M. S., Antonishek, A. S. & Phinney, K. W. Quantification of mRNA in lipid nanoparticles using mass spectrometry. *Anal. Chem.* **96**, 1214–1222 (2024).
129. Medina, J. et al. Omic-scale high-throughput quantitative LC–MS/MS approach for circulatory lipid phenotyping in clinical research. *Anal. Chem.* **95**, 3168–3179 (2023).
130. Kinsey, C. et al. Determination of lipid content and stability in lipid nanoparticles using ultra high-performance liquid chromatography in combination with a corona charged aerosol detector. *Electrophoresis* **43**, 1091–1100 (2022).
131. Cui, H. et al. LUMI-lab: a foundation model-driven autonomous platform enabling discovery of new ionizable lipid designs for mRNA delivery. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.02.14.638383> (2025).
132. Maharjan, R., Kim, K. H., Lee, K., Han, H.-K. & 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).
133. Stach, E. et al. Autonomous experimentation systems for materials development: a community perspective. *Matter* **4**, 2702–2726 (2021).
134. Land, O., Seider, W. D. & Lee, D. Convolutional neural network augmented soft-sensor for autonomous microfluidic production of uniform bubbles. *Chem. Eng. J.* **499**, 156494 (2024).
135. Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86–91 (2023).
136. Chen, J. et al. Navigating phase diagram complexity to guide robotic inorganic materials synthesis. *Nat. Synth.* **3**, 606–614 (2024).
137. Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).
138. Jan, E. et al. High-content screening as a universal tool for fingerprinting of cytotoxicity of nanoparticles. *ACS Nano* **2**, 928–938 (2008).
139. Frei, A. P. et al. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* **13**, 269–275 (2016).
140. Krohn-Grimberghe, M. et al. Nanoparticle-encapsulated siRNAs for gene silencing in the haematopoietic stem-cell niche. *Nat. Biomed. Eng.* **4**, 1076–1089 (2020).
141. Connors, J. et al. Lipid nanoparticles (LNP) induce activation and maturation of antigen presenting cells in young and aged individuals. *Commun. Biol.* **6**, 188 (2023).
142. Haley, R. M. et al. Lipid nanoparticle delivery of small proteins for potent in vivo RAS inhibition. *ACS Appl. Mater. Interfaces* **15**, 21877–21892 (2023).
143. Han, E. L. et al. Predictive high-throughput platform for dual screening of mRNA lipid nanoparticle blood-brain barrier transfection and crossing. *Nano Lett.* **24**, 1477–1486 (2024).
144. Liu, R., Jiang, W., Walkley, C. D., Chan, W. C. W. & Cohen, Y. Prediction of nanoparticles-cell association based on corona proteins and physicochemical properties. *Nanoscale* **7**, 9664–9675 (2015).
145. Alabi, C. A. et al. Multiparametric approach for the evaluation of lipid nanoparticles for siRNA delivery. *Proc. Natl Acad. Sci. USA* **110**, 12881–12886 (2013).
146. Roosa, C. A. et al. Conjugation of IL-33 to microporous annealed particle scaffolds enhances type 2-like immune responses in vitro and in vivo. *Adv. Healthc. Mater.* **13**, 2400249 (2024).
147. Sahay, G. et al. Efficiency of siRNA delivery by lipid nanoparticles is limited by endocytic recycling. *Nat. Biotechnol.* **31**, 653–658 (2013).
148. Rui, Y. et al. High-throughput and high-content bioassay enables tuning of polyester nanoparticles for cellular uptake, endosomal escape, and systemic in vivo delivery of mRNA. *Sci. Adv.* **8**, eabk2855 (2022).
149. Ponsoda, X., Jover, R., Castell, J. V. & Gómez-Lechón, M. J. Measurement of intracellular LDH activity in 96-well cultures: a rapid and automated assay for cytotoxicity studies. *J. Tissue Cult. Methods* **13**, 21–24 (1991).
150. Marin, D. et al. Safety, efficacy and determinants of response of allogeneic CD19-specific CAR-NK cells in CD19⁺ B cell tumors: a phase 1/2 trial. *Nat. Med.* **30**, 772–784 (2024).
151. 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).
152. Kim, J., Vaughan, H. J., Zamboni, C. G., Sunshine, J. C. & Green, J. J. High-throughput evaluation of polymeric nanoparticles for tissue-targeted gene expression using barcoded plasmid DNA. *J. Control. Release* **337**, 105–116 (2021).
153. Guimaraes, P. P. G. et al. Ionizable lipid nanoparticles encapsulating barcoded mRNA for accelerated in vivo delivery screening. *J. Control. Release* **316**, 404–417 (2019).
154. Rhyim, L. H., Manan, R. S., Koller, A., Stephanie, G. & Anderson, D. G. Peptide-encoding mRNA barcodes for the high-throughput in vivo screening of libraries of lipid nanoparticles for mRNA delivery. *Nat. Biomed. Eng.* **7**, 901–910 (2023).
155. Vaidya, K. et al. Pooled nanoparticle screening using a chemical barcoding approach. *Angew. Chem. Int. Ed.* **137**, e202420052 (2025).
156. Boehnke, N. et al. Massively parallel pooled screening reveals genomic determinants of nanoparticle delivery. *Science* **377**, eabm5551 (2022).
157. El-Mayta, R. et al. A nanoparticle platform for accelerated in vivo oral delivery screening of nucleic acids. *Adv. Ther.* **4**, 2000111 (2021).
158. Hatit, M. Z. C. et al. Species-dependent in vivo mRNA delivery and cellular responses to nanoparticles. *Nat. Nanotechnol.* **17**, 310–318 (2022).
159. Hamilton, A. G. et al. High-throughput in vivo screening identifies differential influences on mRNA lipid nanoparticle immune cell delivery by administration route. *ACS Nano* **18**, 16151–16165 (2024).
160. Dobrowolski, C. et al. Nanoparticle single-cell multiomic readouts reveal that cell heterogeneity influences lipid nanoparticle-mediated messenger RNA delivery. *Nat. Nanotechnol.* **17**, 871–879 (2022).
161. Zhang, Q. et al. Predictable control of RNA lifetime using engineered degradation-tuning RNAs. *Nat. Chem. Biol.* **17**, 828–836 (2021).
162. Bruchez, M., Moronne, M., Gin, P., Weiss, S. & Alivisatos, A. P. Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**, 2013–2016 (1998).
163. Paunovska, K. et al. A direct comparison of in vitro and in vivo nucleic acid delivery mediated by hundreds of nanoparticles reveals a weak correlation. *Nano Lett.* **18**, 2148–2157 (2018).
164. Sago, C. D. et al. High-throughput in vivo screen of functional mRNA delivery identifies nanoparticles for endothelial cell gene editing. *Proc. Natl Acad. Sci. USA* **115**, E9944–E9952 (2018).
165. Buschmann, T. & Bystrykh, L. V. Levenshtein error-correcting barcodes for multiplexed DNA sequencing. *BMC Bioinform.* **14**, 272 (2013).
166. Witten, J. et al. Artificial intelligence-guided design of lipid nanoparticles for pulmonary gene therapy. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-024-02490-y> (2024).
167. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. & Lopez, A. A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing* **408**, 189–215 (2020).
168. Yamankurt, G. et al. Exploration of the nanomedicine-design space with high-throughput screening and machine learning. *Nat. Biomed. Eng.* **3**, 318–327 (2019).
169. Kumar, R. et al. Efficient polymer-mediated delivery of gene-editing ribonucleoprotein payloads through combinatorial design, parallelized experimentation, and machine learning. *ACS Nano* **14**, 17626–17639 (2020).
170. Pattipeiluhu, R. et al. Anionic lipid nanoparticles preferentially deliver mRNA to the hepatic reticuloendothelial system. *Adv. Mater.* **34**, 2201095 (2022).
171. Mandl, H. K. et al. Optimizing biodegradable nanoparticle size for tissue-specific delivery. *J. Control. Release* **314**, 92–101 (2019).
172. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, eaap7885 (2018).
173. Tropsha, A., Isayev, O., Varnek, A., Schneider, G. & Cherkasov, A. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. *Nat. Rev. Drug Discov.* **23**, 141–155 (2024).
174. Walsh, D. J. et al. Community Resource for Innovation in Polymer Technology (CRIPT): a scalable polymer material data structure. *ACS Cent. Sci.* **9**, 330–338 (2023).
175. Kuenneth, C. & Ramprasad, R. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nat. Commun.* **14**, 4099 (2023).
176. Meyer, T. A., Ramirez, C., Tamasi, M. J. & Gormley, A. J. A user's guide to machine learning for polymeric biomaterials. *ACS Polym. Au* **3**, 141–157 (2023).
177. Xue, K. et al. Biomaterials by design: harnessing data for future development. *Mater. Today Bio* **12**, 100165 (2021).
178. Haley, R. M. et al. Lipid nanoparticles for in vivo lung delivery of CRISPR-Cas9 ribonucleoproteins allow gene editing of clinical targets. *ACS Nano* **19**, 13790–13804 (2025).
179. Tang, C. et al. mRNA-laden lipid-nanoparticle-enabled in situ CAR-macrophage engineering for the eradication of multidrug-resistant bacteria in a sepsis mouse model. *ACS Nano* **18**, 2261–2278 (2024).

Acknowledgements

M.J.M. acknowledges support from the US National Institutes of Health (NICHD R01 HD115877), the Burroughs Wellcome Fund Career Award at the Scientific Interface (CASI), a US National Science Foundation (NSF) CAREER award (CBET-2145491), the American Cancer Society Research Scholar Grant (RSG-22-122-01-ET) and the Cystic Fibrosis Foundation (MITCHE2410). D.A.I. acknowledges support from the Wellcome Leap R3 programme, a NSF Materials Research Science and Engineering Centers (MRSECs) grant (DMR-2309034), NSF Biofoundry and the Center for Precision Engineering for Health at University of Pennsylvania. A.R.H. is supported by the US National Science Foundation Graduate Research Fellowship.

Author contributions

The authors contributed equally to all aspects of the article.

Competing interests

M.J.M. is named on patents describing the use of lipid nanoparticles and lipid compositions for nucleic acid delivery that are discussed in this article. D.A.I. is a founder and holds equity in InfiniFluidics. A.R.H. declares no competing interests.

Additional information

Peer review information *Nature Reviews Materials* thanks Xing-Jie Liang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2025