Machine Learning-Based Design of Lipid Nanoparticles for mRNA Delivery: A Review

MD IKBAL HUSAIN*, MINAM JONGKEY, NABANITA CHAKMA, PABITRA DEBNATH, SOMENATH ROUTH, SUSMITA SARKAR
Regional Institute of Pharmaceutical Science and Technology
Abhoynagar, Agartala, West Tripura – 799005, India
doi.org/10.64643/IJIRTV12I8-191174-459

***Abstract*—Lipid nanoparticles (LNPs) are widely used as delivery systems for messenger RNA (mRNA) therapeutics due to their ability to protect mRNA from degradation and improve cellular uptake. Their effectiveness was clearly demonstrated during the COVID-19 pandemic, where LNP-based vaccines played a major role in rapid vaccine development. However, traditional LNP formulation methods largely depend on trial-and-error approaches, which are time-consuming and limited in their ability to explore a wide range of lipid compositions. In recent years, machine learning (ML) has emerged as a useful tool to support the rational design of LNPs by analyzing large datasets and predicting formulation performance. This review discusses the basic principles of LNP design, limitations of conventional development strategies and the application of ML techniques in optimizing LNP formulations for mRNA delivery. Recent studies highlighting ML-assisted LNP optimization are also summarized. Overall, the integration of machine learning with experimental research offers a promising approach for improving the efficiency and reliability of mRNA delivery systems.**

***Key Words*— Lipid nanoparticles (LNPs), mRNA, Machine Learning.**

I. INTRODUCTION

Messenger RNA-based therapeutics have gained significant attention in recent years due to their applications in vaccine development, gene therapy and protein replacement therapy. Unlike traditional drugs, mRNA therapies require effective delivery systems to protect the nucleic acid from degradation and ensure efficient entry into target cells. Lipid nanoparticles have become the most widely used carriers for this purpose due to their biocompatibility and ability to encapsulate mRNA molecules [1]. The successful use of LNP-based mRNA vaccines during the COVID-19 pandemic highlighted their clinical importance and potential for rapid therapeutic development. Despite this success, the design and optimization of LNP formulations remain challenging. Several formulation variables, including lipid composition, particle size, surface charge, and encapsulation efficiency, influence delivery performance and biological response [2]. Conventional LNP development relies heavily on empirical screening methods, which are slow, costly and limited in exploring the vast lipid design space. To overcome these limitations, machine learning (ML) techniques have been increasingly applied to analyze experimental data, predict formulation outcomes and guide rational LNP design [3].

II. COMPONENTS OF LIPID NANOPARTICLES

Lipid nanoparticles used for mRNA delivery are generally composed of four main components: ionizable lipids, phospholipids, cholesterol and polyethylene glycol (PEG)-lipids. Each of these components plays a specific and important role in ensuring effective mRNA encapsulation, stability of the nanoparticle and successful delivery into target cells.

Ionizable Lipids:
Ionizable lipids are the most critical component of LNPs for mRNA delivery. These lipids help in binding and encapsulating the negatively charged mRNA molecules during nanoparticle formation. At physiological pH, ionizable lipids remain mostly neutral, which helps reduce toxicity and improves biocompatibility. However, once the LNP enters the acidic environment of the endosome, these lipids become positively charged. This change in charge assists in disrupting the endosomal membrane, allowing the mRNA to escape into the cytoplasm, where it can be translated into protein. The chemical structure and pKa of ionizable lipids strongly influence encapsulation efficiency, transfection efficiency and overall delivery performance. Small

modifications in their structure can lead to significant changes in biological activity [4].

Phospholipids:
Phospholipids act as structural components of LNPs and help maintain the integrity of the lipid bilayer. They support nanoparticle assembly and contribute to membrane stability. Phospholipids also assist in mimicking natural cell membrane components, which can improve compatibility with biological systems. Commonly used phospholipids help stabilize the nanoparticle structure during circulation and support efficient fusion with cellular membranes. Their presence ensures that the LNP maintains its shape and functionality under physiological conditions [5].

Cholesterol:
Cholesterol is included in LNP formulations to enhance membrane rigidity and structural strength. It helps fill gaps between lipid molecules, resulting in a more compact and stable nanoparticle. Cholesterol also improves resistance to mechanical stress and contributes to better stability during storage and circulation. In addition, cholesterol plays a role in improving cellular uptake and endosomal escape by enhancing membrane fusion properties. The amount of cholesterol used can influence particle size, stability, and delivery efficiency [6].

Polyethylene Glycol (PEG)-Lipids:
PEG-lipids are incorporated into LNPs to improve their stability and circulation behavior. The PEG chains form a protective layer on the surface of nanoparticles, reducing aggregation and preventing rapid clearance by the immune system. This effect helps extend the circulation time of LNPs in the bloodstream. While PEG-lipids are beneficial for stability, excessive PEG content may reduce cellular uptake. Therefore, an appropriate balance is required to achieve optimal performance. The length of the PEG chain and the density of PEG-lipids on the nanoparticle surface are important factors influencing LNP behavior [7].

Challenges in Traditional LNP Development:
Traditional LNP development methods are mainly based on repeated experimental screening, making the process slow and resource-intensive. The large number of possible lipid combinations creates a high-dimensional design space that is difficult to explore using conventional approaches [8]. Additionally, results obtained from in vitro studies do not always translate directly to in vivo performance due to complex biological and immunological interactions. These limitations highlight the need for alternative strategies that can predict formulation behavior more efficiently and reduce experimental workload [9].

Role of Machine Learning in Drug Delivery:
Machine learning refers to computational methods that identify patterns in data and generate predictions based on learned relationships [10]. In the context of LNP formulation, ML can be used to link lipid composition and physicochemical properties with biological outcomes such as encapsulation efficiency, biodistribution, transfection efficiency and immune response [11]. Supervised ML models use labeled datasets to predict specific outcomes, while unsupervised learning helps identify hidden patterns within complex datasets. Reinforcement learning approaches can further optimize formulations by iteratively improving performance based on feedback [12]. The application of ML reduces dependence on trial-and-error methods and supports more rational formulation design.

Machine Learning Methods Used in LNP Design:
In recent years, various machine learning methods have been applied to improve the design and optimization of lipid nanoparticles for mRNA delivery. These approaches help analyze complex datasets and identify relationships between lipid composition, physicochemical properties and biological performance. The main ML methods used in LNP design are described below.

i. Predictive Modeling:
Predictive modeling is one of the most commonly used ML approaches in LNP research. Supervised machine learning algorithms are trained using experimental datasets that include lipid composition, particle size, surface charge, encapsulation efficiency and transfection performance. Once trained, these models can predict key LNP performance parameters, such as mRNA delivery efficiency, biodistribution, and stability, for new formulations. Predictive models reduce the need for extensive experimental screening by identifying promising formulations at an early stage. This approach improves development efficiency and supports rational decision-making during formulation design [13].

ii. Generative Modeling:
Generative modeling uses advanced deep learning techniques to design new lipid structures or propose novel lipid combinations with improved delivery properties. These models learn patterns from existing datasets and generate new candidates that are predicted to perform well for specific therapeutic goals, such as enhanced cellular uptake or reduced toxicity. Generative ML approaches expand the accessible lipid design space beyond traditional trial-and-error methods. By suggesting innovative lipid structures, these models support the discovery of next-generation ionizable lipids for mRNA delivery [14].

iii. Optimization Algorithms:
Optimization algorithms are used to identify the best balance between multiple formulation objectives. In LNP design, important factors such as mRNA delivery efficiency, particle stability, toxicity, and immunogenicity often need to be optimized simultaneously. Multi-objective ML frameworks help evaluate trade-offs between these competing parameters and identify formulations with overall optimal performance. This approach is particularly useful when small changes in lipid composition can significantly influence biological outcomes [15].

iv. Transformer-Based and Advanced ML Models:
Recent studies have applied transformer-based deep learning models to predict organ-specific delivery of LNPs and guide lipid selection. These models are capable of handling complex, high-dimensional datasets and learning long-range relationships between lipid structure and biological response. As a result, they can accurately predict tissue targeting and delivery efficiency. Such computational approaches help narrow down a large number of formulation candidates to a manageable set for experimental validation. This reduces time, cost, and experimental workload while increasing the likelihood of successful outcomes [16].

Case Studies and Recent Advancements:
Several recent studies highlight the practical application of machine learning (ML) in the design and optimization of lipid nanoparticles (LNPs) for mRNA delivery. These studies demonstrate how computational approaches can support faster development, improved targeting and enhanced formulation performance.

COVID-19 Vaccine LNPs:
The rapid development of mRNA vaccines during the COVID-19 pandemic represents one of the most successful applications of LNP technology. Machine learning–assisted optimization of ionizable lipids played an important role in improving mRNA encapsulation, stability, and delivery efficiency. ML models helped identify lipid structures with favorable physicochemical properties, contributing to the development of safe and effective vaccine formulations in a short time frame. This example highlights the value of ML in accelerating formulation development under urgent clinical needs [17].

Organ-Targeted Lipid Nanoparticles:
Recent advances in deep learning have enabled the prediction of tissue-specific delivery of LNPs. By analyzing large datasets that link lipid composition with biodistribution patterns, ML models can identify lipid combinations that preferentially target specific organs. This approach has improved the design of organ-targeted LNPs for applications such as liver-specific and lung-specific mRNA delivery. Such targeted delivery strategies help enhance therapeutic efficacy while reducing off-target effects [18].

Integration with High-Throughput Screening:
The integration of machine learning with high-throughput experimental techniques represents another major advancement in LNP development. When combined with microfluidic LNP synthesis platforms, ML algorithms can rapidly analyze data from hundreds of formulations and identify optimal candidates. This approach significantly reduces experimental workload, time, and material consumption while improving formulation selection efficiency. High-throughput ML-guided screening supports systematic and data-driven LNP optimization [19].

Future Scope and Applications:
The continued integration of machine learning with experimental workflows is expected to further improve lipid nanoparticle design for mRNA delivery. ML models may support the development of personalized mRNA therapies by guiding LNP design based on individual biological and immunological characteristics [20]. In addition, automated platforms that combine ML with high-throughput formulation techniques could enable real-time optimization of lipid composition to achieve

improved delivery performance. ML-based prediction tools may also assist in the early identification of potential toxicity and immunogenicity concerns, supporting safer formulation development and regulatory decision-making [21]. Overall, these advancements are likely to accelerate the development of effective and reliable mRNA therapeutics.

III. CONCLUSION

Lipid nanoparticles are essential carriers for mRNA therapeutics, but traditional formulation approaches are limited by their reliance on empirical methods. Machine learning provides a valuable tool for predicting formulation performance, supporting rational design, and improving development efficiency. The integration of ML with experimental research enhances LNP optimization and opens new opportunities for advanced mRNA delivery systems. Continued progress in this field is expected to strengthen the future of nucleic acid-based therapies.

REFERENCE

[1] Liu, Y.; Huang, Y.; He, G.; Guo, C.; Dong, J.; Wu, L. (2024). Development of mRNA Lipid Nanoparticles: Targeting and Therapeutic Aspects. *Int. J. Mol. Sci.* , *25*, 10166.

[2] John, R., Monpara, J., Swaminathan, S., & Kalhapure, R. (2024). Chemistry and Art of Developing Lipid Nanoparticles for Biologics Delivery: Focus on Development and Scale-Up. *Pharmaceutics*, *16*(1), 131.

[3] Abughalia, A., Flynn, M., Clarke, P. F. A., Fayne, D., & Gobbo, O. L. (2025). The Use of Computational Approaches to Design Nanodelivery Systems. *Nanomaterials*, *15*(17), 1354.

[4] Alshehry, Y., Liu, X., Li, W. *et al.* (2025). Lipid Nanoparticles for mRNA Delivery in Cancer Immunotherapy. *AAPS J* 27, 66.

[5] Hald Albertsen C, Kulkarni JA, Witzigmann D, Lind M, Petersson K, Simonsen JB (2022). The role of lipid components in lipid nanoparticles for vaccines and gene therapy. Adv Drug Deliv Rev.188:114416.

[6] Sago CD, Lokugamage MP, Paunovska K, Vanover DA, Monaco CM, Shah NN, et al. (2018). High-throughput in vivo screen of functional mRNA delivery identifies nanoparticles for endothelial cell gene editing. Proc Natl Acad Sci.115(42):E99 44-52.

[7] Pozzi D, Colapicchioni V, Caracciolo G, Piovesana S, Capriotti AL, Palchetti S, et al. (2014). Effect of polyethyleneglycol (PEG) chain length on the bio-nano-interactions between PEGylated lipid nanoparticles and biological fluids: from nanostructure to uptake in cancer cells. Nanoscale. 6(5):2782–92.

[8] Jung HN, Lee SY, Lee S, et al.(2022). Lipid nanoparticles for delivery of RNA therapeutics: current status and the role of in vivo imaging. Theranostics.12(17):7509–7531.

[9] Cheng X, Lee RJ.(2016). The role of helper lipids in lipid nanoparticles (LNPs) designed for oligonucleotide delivery. Adv Drug Deliv. 99(Pt A):129–137. doi: 10.1016/j.addr.2016.01.022

[10] Khera, P., & Kumar, N. (2020). Role of machine learning in gait analysis: a review. *Journal of Medical Engineering & Technology*, *44*(8), 441–467.

[11] Sayadi, H., Fromage, Y., Labriffe, M., Codde, C., Monchaud, C., Marquet, P., Woillard, J. B. (2025). Optimized therapeutic drug monitoring: the role of machine learning models. *Expert Review of Clinical Pharmacology*, 1–10.

[12] Al Meslamani, A. Z., Jarab, A. S., & Ghattas, M. A. (2023). The role of machine learning in healthcare responses to pandemics: maximizing benefits and filling gaps. *Journal of Medical Economics*, *26*(1), 777–780.

[13] Y. Dong et al. (2014)., Lipopeptide nanoparticles for potent and selective siRNA delivery in rodents and nonhuman primates. Proc. Natl. Acad. Sci. U.S.A. 111, 3955.

[14] B. Li, I.O. Raji, A.G.R. Gordon, L. Sun, T.M. Raimondo, F.A. Oladimeji, A.Y. Jiang, A. Varley, R.S. Langer, D.G. Anderson, (2024). Accelerating ionizable lipid discovery for mRNA delivery using machine learning and combinatorial chemistry. Nat. Mater. 23, 1002.

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

[16] K. Yang et al. (2019)., Analyzing learned molecular representations for property prediction. J. Chem. Inf. Model. 59, 3370.

[17] Mulet i Piera, X., Del Campo-Montoya, R., Cuadrado-Tejedor, M., Garcia-Osta, A., Garbayo, E., & Blanco-Prieto, M. J. (2025). Intranasal delivery of lipid-based nanoparticles

for the treatment of neurodegenerative diseases: advances, challenges and future perspectives. *Expert Opinion on Drug Delivery*, 1–19.

[18] K. K., S. P., & Narayansamy, D. (2025). Advancements in nanotechnology for targeted drug delivery in idiopathic pulmonary fibrosis: a focus on solid lipid nanoparticles and nanostructured lipid carriers. *Drug Development and Industrial Pharmacy*, *51*(4), 285–294.

[19] Mittal, R., Patel, A. P., Jhaveri, V. M., Kay, S. I. S., Debs, L. H., Parrish, J. M., … Jayant, R. D. (2018). Recent advancements in nanoparticle based drug delivery for gastrointestinal disorders. *Expert Opinion on Drug Delivery*, *15*(3), 301–318.

[20] Smiatek J, Jung A, Bluhmki E (2020). Towards a digital bioprocess replica: computational approaches in biopharmaceutical development and manufacturing. Trends Biotechnol 38(10):1141–1153.

[21] Xu P (2019). Analytical solution for a hybrid logistic-monod cell growth model in batch and continuous stirred tank reactor culture. Biotechnol Bioeng 117(3):873–878.