RESEARCH ARTICLE

NANO · MICRO small
www.small-journal.com

# Rational Design of Lipid Nanoparticles for Enhanced mRNA Vaccine Delivery via Machine Learning

*Seo-Hyeon Bae, Hosam Choi, Jisun Lee, Min-Ho Kang, Seong-Ho Ahn, Yu-Sun Lee, Huijeong Choi, Sohee Jo, Yeeun Lee, Hyo-Jung Park, Seonghyun Lee, Subin Yoon, Gahyun Roh, Seongje Cho, Youngran Cho, Dahyeon Ha, Soo-Yeon Lee, Eun-Jin Choi, Ayoung Oh, Jungmin Kim, Sowon Lee, Jungmin Hong, Nakyung Lee, Minyoung Lee, Jungwon Park, Dong-Hwa Jeong,* Kiyoun Lee,* and Jae-Hwan Nam**

**Since the coronavirus pandemic, mRNA vaccines have revolutionized the field of vaccinology. Lipid nanoparticles (LNPs) are proposed to enhance mRNA delivery efficiency; however, their design is suboptimal. Here, a rational method for designing LNPs is explored, focusing on the ionizable lipid composition and structural optimization using machine learning (ML) techniques. A total of 213 LNPs are analyzed using random forest regression models trained with 314 features to predict the mRNA expression efficiency. The models, which predict mRNA expression levels post-administration of intradermal injection in mice, identify phenol as the dominant substructure affecting mRNA encapsulation and expression. The specific phospholipids used as components of the LNPs, as well as the N/P ratio and mass ratio, are found to affect the efficacy of mRNA delivery. Structural analysis highlights the impact of the carbon chain length on the encapsulation efficiency and LNP stability. This integrated approach offers a framework for designing advanced LNPs and has the potential to unlock the full potential of mRNA therapeutics.**

## 1. Introduction

The emergence of mRNA vaccines since the 2019 coronavirus pandemic has led to significant advances in vaccinology by utilizing mRNA molecules to encode antigens, eliciting potent immune responses against pathogens.[1] However, mRNA molecules are inherently unstable and have limited immunogenicity, necessitating in-depth advancements in delivery systems for enhanced efficiency.[2] To address these challenges, lipid nanoparticles (LNPs) have emerged as a promising solution to facilitate efficient mRNA delivery, and this area has been the focus of extensive research due to the potential of LNPs across various biomedical applications.[3] Rational designing for LNPs involves optimizing their composition and structure for better mRNA

S.-H. Bae, J. Lee, Y.-S. Lee, H. Choi, S. Jo, Y. Lee, H.-J. Park, S. Lee, S. Yoon, G. Roh, S. Cho, Y. Cho, D. Ha, S.-Y. Lee, E.-J. Choi, A. Oh, J. Kim, S. Lee, J.-H. Nam
Department of Medical and Biological Sciences
The Catholic University of Korea
Bucheon 14662, Republic of Korea
E-mail: jhnam@catholic.ac.kr

S.-H. Bae, H. Choi, M.-H. Kang, S. Jo, Y. Lee, H.-J. Park, S. Lee, S. Yoon, G. Roh, S. Cho, Y. Cho, D. Ha, S.-Y. Lee, E.-J. Choi, A. Oh, J. Kim, S. Lee, J.-H. Nam
Department of Biotechnology
The Catholic University of Korea
Bucheon 14662, Republic of Korea

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smll.202405618



**DOI: 10.1002/smll.202405618**

H. Choi, J. Hong, N. Lee, K. Lee
Department of Chemistry
The Catholic University of Korea
Bucheon 14662, Republic of Korea
E-mail: kiyoun@catholic.ac.kr

M.-H. Kang
Department of Biomedical-Chemical Engineering
The Catholic University of Korea
Bucheon 14662, Republic of Korea

S.-H. Ahn, D.-H. Jeong
Department of Artificial Intelligence
The Catholic University of Korea
Bucheon 14662, Republic of Korea
E-mail: donghwa@catholic.ac.kr

M. Lee, J. Park
School of Chemical and Biological Engineering
Institute of Chemical Processes
Seoul National University
Seoul 08826, Republic of Korea

M. Lee, J. Park
Center for Nanoparticle Research
Institute of Basic Science (IBS)
Seoul 08826, Republic of Korea

delivery and expression efficiency as LNPs typically include four key components: ionizable lipids, helper lipids, cholesterol, and polyethylene glycol (PEG)–lipid conjugates. Each of these components is crucial for mRNA encapsulation, cellular uptake, and protection against degradation.[4]

Notably, ionizable lipids, forming the LNP core, typically consist of a head, linker, and tail. The head contains amino groups and other compounds that determine the charge and solubility, influencing interactions with membranes. The linker comprises esters, amides, and disulfide bonds, while the tail includes chains of 8 to 18 carbon atoms, which can vary in structure and biodegradability.[5] The pH sensitivity of ionizable lipids, influenced by the head group, leads to changes in charge in response to pH fluctuations, affecting release efficiency in the body.[6] Under neutral pH, ionizable lipids in mRNA/LNP complexes remain stable, but in acidic endosomal environments, they protonate, bind to endosomal lipids, destabilize membranes, and release mRNA.[7] The tail of the ionizable lipid also plays a critical role in the release of nucleic acids into the cytoplasm and the overall transfection efficiency. For instance, branching aliphatic chains in the tail can significantly enhance the nucleic acid transfection efficiency by facilitating the formation of crown-like structures, which contribute to rapid nucleic acid release.[5,8] Furthermore, increasing the level of unsaturation in the tail can greatly improve mRNA delivery efficiency.[9]

Despite these mechanisms, current ionizable lipids have limitations such as low delivery efficiency and potential toxicity. Although several injected LNPs are internalized by cells, only a small fraction (<2%) of the nucleic acid cargo escapes the endosome to reach the cytoplasm,[10] limiting clinical utility. The need for higher doses to overcome low endosomal escape efficiency can induce cytotoxic effects, posing a significant challenge in optimizing LNPs for mRNA delivery. Therefore, significant research is focused on developing new ionizable lipids to improve biocompatibility and reduce toxicity.[3]

Piperazine, a heterocyclic compound with two nitrogen atoms and two amine functional groups, has been widely studied as a component of ionizable lipids. Piperazine-mediated ionizable lipids have higher nucleic acid transfection efficiencies and lower cytotoxicity than lipids with other tertiary amine head groups. These characteristics make them attractive for the development of LNPs for mRNA delivery.[11] In addition, the use of piperazine in LNPs has been reported to enhance the permeation of various substances across cell membranes and to improve the efficiency of mRNA delivery.

J. Park
Institute of Engineering Research
College of Engineering
Seoul National University
Seoul 08826, Republic of Korea
J. Park
Advanced Institutes of Convergence Technology
Seoul National University
Suwon-si 16229, Republic of Korea
J.-H. Nam
SML Biopharm
Gwangmyeong 14353, Republic of Korea

Although the design of these ionizable lipids is the most important factor in the efficacy of LNPs, to date, they have been haphazardly designed and tested individually without any theoretical basis. Moreover, designing effective LNPs involves optimizing interactions between ionizable lipids and other components, a process hampered by traditional design of experiment (DOE) approaches that require significant time and resources. Recently, new approaches have been explored that integrate machine learning (ML) and DOE methodologies to accelerate LNP development. Artificial intelligence (AI), especially ML and deep learning, has become a powerful tool for analyzing complex structure–function relationships in LNPs, such as predicting drug–target interactions,[12] drug delivery mechanisms,[13] toxicity effects,[14] and molecular properties.[15]

In this study, a random forest (RF) ensemble ML technique was employed to analyze the design of ionizable lipids and LNPs to predict mRNA expression efficiency. ML approaches provide deeper insight into the effects of subtle structural variations in lipid chemistry on critical LNP properties, including physicochemical characteristics, nucleic acid encapsulation, in vivo mRNA expression, immune responses, and toxicity. The integration of ML-driven screening with traditional design approaches offers a monitoring system for designing smarter LNPs, which is poised to unlock the full potential of mRNA therapeutics.

## 2. Results and Discussion

### 2.1. ML Performance and Feature Importance

A comprehensive investigation was conducted that encompassed 213 LNPs with diverse compositions, including ionizable lipids with various functional heads (FHs), ester bridge (EB) lengths, and tail (T) types (Figure S1, Supporting Information). The novel ionizable lipids featured 14 types of FHs, including aliphatic chains, phenol, and phenyl groups. Phenols are notable for their hydroxyl groups, which facilitate interactions with mRNA nucleobases, improving the stability and expression of mRNA in the body. The numbers of carbon atoms in the EB can range from 5 to 12, whereas the T carbon length ranges from 1 to 2. To determine the influence of the EB and T carbon lengths, we synthesized 14 types of EB–T chains containing carboxylic acid moieties. The delivery efficiencies of these LNPs were assessed using a *Renilla* luciferase (RLuc) assay in mice, demonstrating their potential for enhanced mRNA delivery (Figure S2, Supporting Information). To screen the LNPs and further explore the significance of their substructures and compositions, we employed a RF algorithm.[16] RF is an ML technique known for its interpretability and its robustness against overfitting (**Figure 1a**). We posited that the chemical substructures within the FH, EB, and T are fundamental in influencing efficacy. The specific chemical substructures were counted in the ionizable lipids and were represented by applying the simplified molecular input line entry system (SMILES) and the SMILES Arbitrary Target Specification (SMARTS)[17] using PaDEL-Descriptor software.[18] The occurrence counts of each of the 307 substructure fingerprints (FP1 to FP307) were used as input features for the RF model (Table S1, Supporting Information). Moreover, to investigate how the LNP compositions affect the efficacy, the input

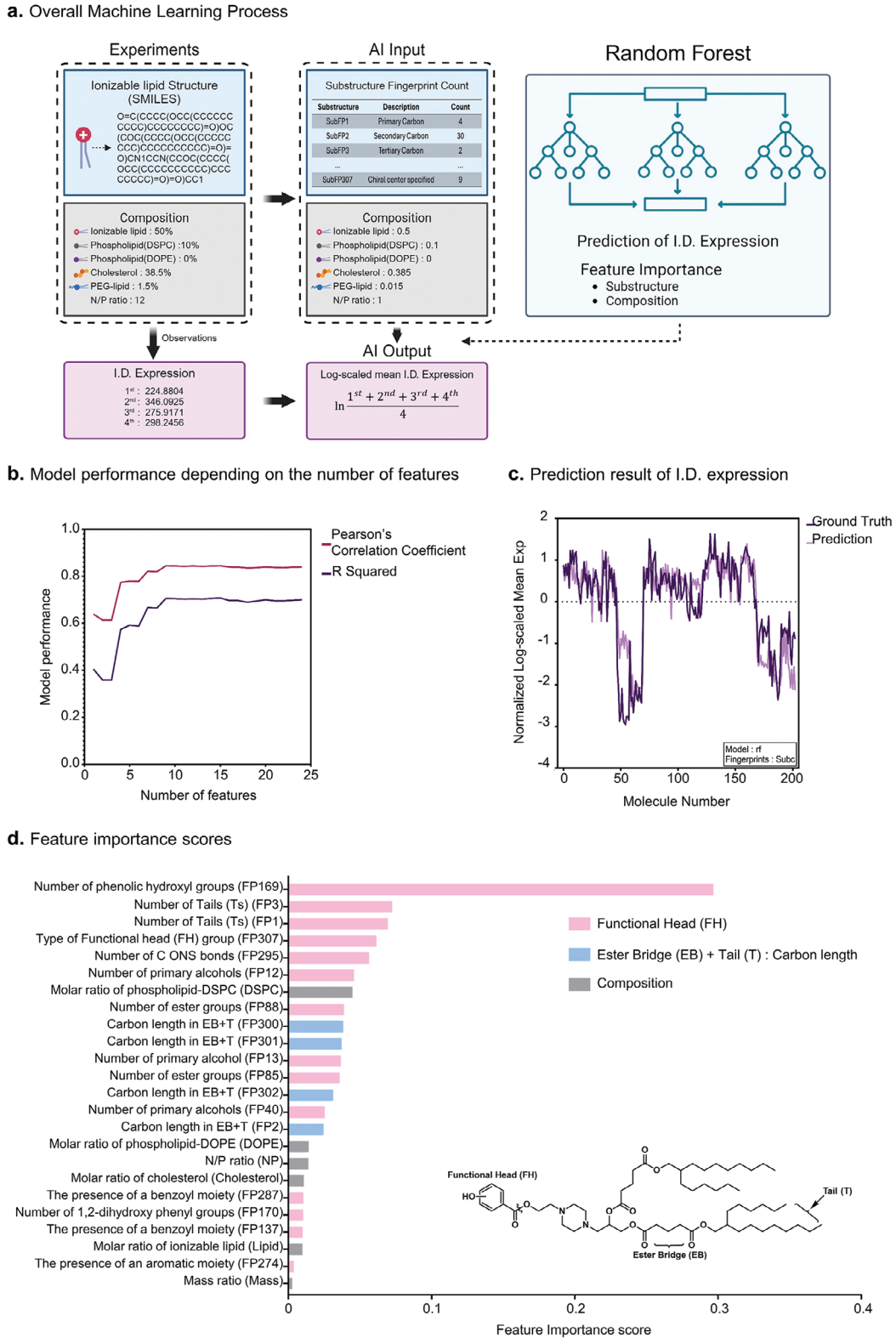


**Figure 1.** Machine learning (ML) analysis for predicting the mRNA expression efficiency and assessing feature importance of the ionizable lipid substructure and LNP composition. a) Overall ML process. Graphical images were created with BioRender.com. b) Model performance depending on the number of features. c) Prediction result of intradermal (I.D.) mRNA expression. d) Feature importance scores.

features for the ML included the ratio of amine groups (N) of the ionizable lipid to phosphate groups (P) of the encapsulated unit (N/P ratio), mass, and molar ratios of ionizable lipids, distearoylphosphatidylcholine (DSPC; saturated phospholipid), dioleoylphosphatidylethanolamine (DOPE; unsaturated phospholipid), cholesterol, and PEG–lipid conjugates. The variations in composition are detailed in Table S2 (Supporting Information). A total of 314 features were collected and used to train the RF regression model to predict the average mRNA expression level of an intradermal (I.D.) injection. To evaluate the improvement in mRNA expression level relative to the baseline, the natural logarithm of the mRNA expression level compared with the baseline was used as the prediction output.

During the evaluation of the RF regression model using tenfold cross-validation, optimal model performance was achieved with the utilization of 15 features, which demonstrated the highest feature importance (Figure 1b). The RF model with 15 features resulted in a coefficient of determination ($R^2$) of 0.708 and a Pearson's correlation coefficient (PCC) of 0.845, indicating successful prediction of the mRNA expression efficiency of the I.D. injection (Figure 1c). To assess the contribution of each feature on the mRNA expression level, the feature importance was examined using the branching criteria in the context of RF regression (Figure 1d). The count of FP169 (phenol) was identified as the most important feature, with a feature importance score of 0.297. Moreover, FP features related to the type of FH significantly influenced the prediction of the I.D. expression, accounting for 10 of the 15 features. While the molar ratio of DSPC ranked seventh, with a feature importance score of 0.045, the molar ratios of DOPE, cholesterol, and ionizable lipids and the N/P ratio adversely affected the model performance. Substructure FPs associated with both EB and Ts, such as the count of FP300 or FP2 (carbon length in EB+T), improved model performance.

### 2.2. Feature Analysis of mRNA Expression Efficiency

The prediction of mRNA expression involved rigorous statistical testing to identify significant features. Phenols (FP169) within the ionizable lipid structure had the highest feature importance score in the RF analysis, making it the primary substructure for FH examination. Lipids with two phenolic hydroxyl groups showed significantly higher I.D. expression levels (**Figure 2a**), while those with three tertiary carbons (FP3) in aliphatic chains showed decreased expression (Figure 2b). Primary alcohol (FP12) molecules as FHs had a significant p-value for I.D. expression, but the difference in levels was not substantial (Figure 2c). Variations in EB and T carbon lengths (FP300, FP302, and FP2) showed a decreasing trend in I.D. expression with increasing carbon length (Figure 2d–f). The composition ratios also affected the LNP efficiency, with a higher N/P ratio exhibiting better mRNA expression (Figure 2g). A mass ratio of 0.5 resulted in significantly higher mRNA expression, but from just 25 cases (Figure 2h). Increasing the DSPC molar ratio reduced efficiency, while replacing DSPC with DOPE improved it, but no statistical difference was observed (Figure 2i,j). Comparisons of the molar ratios of ionizable lipids and cholesterol led to no discernible effect, aligning with AI predictions (Figure 2k,l).

### 2.3. LNPs Formulated with Ionizable Lipids Designed Based on ML

The ML models, combined with the substructure fingerprint (FP)f, primarily recognized the carbon chain length but were unable to selectively distinguish individual substructures such as the EB and T. To investigate the functionality of LNPs in relation to the structural variations of ionizable lipids, we systematically designed combinations by separating the ionizable lipids into three key substructures: EB, which dictates carbon chain length; T, which determines the chain type; and FH, which influences the chemical reactivity and interaction with biological targets. This approach allowed us to independently assess the impact of each substructure on LNP functionality.

To confirm the correlation between the substructural features and I.D. mRNA expression efficiency, we synthesized 36 variations of ionizable lipids using six types of FHs (FH-1 to FH-6), three types of EBs (EB-5 to EB-7), and two types of Ts (T-1 and T-2) (**Figure 3a**). The FHs were aliphatic branched chains (FH-1), aromatic rings without hydroxyl functional groups (FH-2), primary alcohols without FH (FH-3), and three types of phenols with two hydroxyl groups each (FH-4 to FH-6). Based on their importance, phenols were chosen to explore the influence of positional isomers. Particularly, FH-4 and FH-5 were selected to evaluate the effects of the positions of the two phenolic alcohols. Moreover, FH-6 was selected to investigate the effect of the carbon length beyond the position of the phenolic alcohols.

Based on AI predictions and previous statistical analyses (Figure 2), a novel LNP composition was introduced, consisting of ionizable lipid:DOPE:cholesterol:DMG-PEG2000 in a ratio of 25:25:48.5:1.5. These optimized LNP conditions were consistently applied in all subsequent experiments presented to ensure the reliability and comparability of the results. The LNPs were synthesized using a microfluidic device, where lipids from the ethanol phase were combined with RNA dissolved in a pH 4 water phase (Figure 3b). Subsequently, we analyzed the physicochemical properties of 36 selected LNPs using dynamic light scattering (DLS). Most LNPs showed homogeneous size distributions and had polydispersity index (PDI) of below 0.2, indicating highly uniform formulations, and exhibited zeta potentials ranging from −30 to +30 mV, indicating predominantly neutral or weakly charged surfaces near the physiological pH (Figure S3 and Table S3, Supporting Information).

The encapsulation efficiency (EE), which was assessed by quantifying the RNA content inside and outside the LNPs, showed significant variations based on the presence of hydroxyl groups in the FH (Figure 3c). FH-1 and FH-2 exhibited decreased EE due to interference with binding via electrostatic attraction of the tertiary amines. Consequently, ionizable lipids containing FH-1 or FH-2 showed a compromised ability to shield RNA from ribonuclease activity, suggesting diminished protection of mRNA, which can impact the mRNA expression and immune response.[19] Particularly, FH-2 groups, characterized by their aromatic ring structure, are rigid and promote interactions with nucleobases through $\pi$–$\pi$ stacking and other hydrophobic interactions.[20] However, this structure impedes electrostatic attraction with tertiary amines, leading to reduced EE. Hydrogen bonding with RNA was expected to occur with hydroxyl groups in FH-3, FH-4, FH-5, and FH-6. Overall, the quantification of RNA

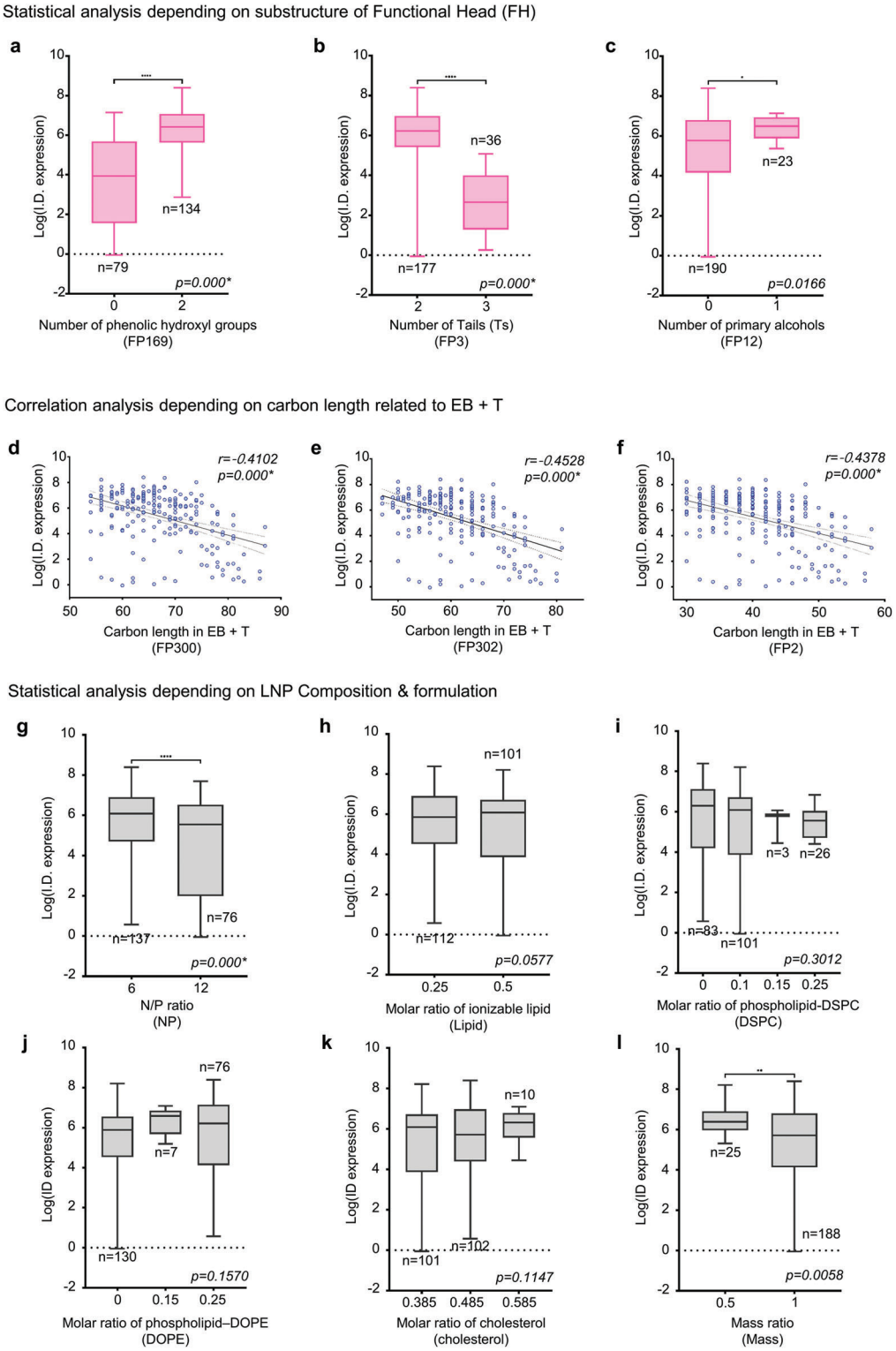


**Figure 2.** Statistical analysis between features and I.D. mRNA expression efficiency. The I.D. mRNA expression levels utilized for statistical analysis were replaced by a logarithmic transformation. a–c) Statistical analysis depending on substructure of FH, d–f) Correlation analysis depending on carbon length in EB and T. g–l) Statistical analysis depending on LNP composition and formulation. In the box plots graphs (a–c and g–l), the central line of each box denotes the median, the box edges indicate the interquartile range (IQR), and the whiskers represent the range of the data (1.5 times the IQR from the first and third quartile). The number of samples (n) is shown for each group. Statistical significance was defined as $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. In the scatter plots (d–f), the solid lines represent the regression line, while the dotted lines indicate the standard error. The correlation coefficient ($r$) is shown for each regression line.

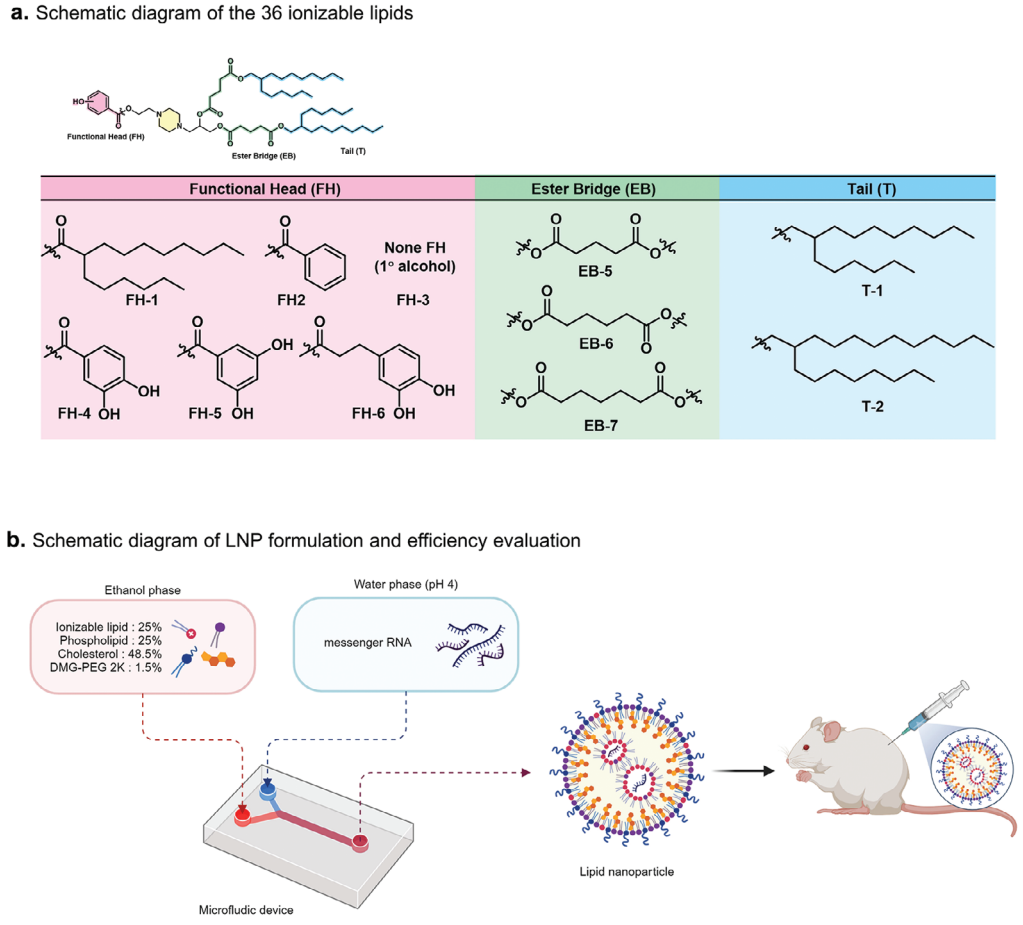


**Figure 3.** Formulation of LNPs using selected ionizable lipids and physicochemical properties of mRNA-loaded LNPs. a) Schematic diagram of 36 ionizable lipids selected for systematic analysis. b) Schematic illustration of RLuc-loaded mRNA-LNPs formulated with ionizable lipids selected using ML. Graphical images were created with BioRender.com. c) Encapsulation efficiency (EE) of mRNA-loaded LNPs. Data are represented as the mean ± standard deviation. Statistical significance was analyzed using two-way ANOVA and defined as $^{*}p < 0.05$, $^{***}p < 0.001$, and $^{****}p < 0.0001$. (P.C, positive control).

content confirmed minimal or negligible RNA loss during LNP production. Accordingly, the presence of a hydroxyl group in the FH substantially affected nucleic acid interactions. Notably, we observed that the EE tended to increase with increasing EB length in the FH-5 and FH-6 groups with T-2, although this enhancement was not statistically significant.

### 2.4. Effects of LNPs on mRNA Expression and Immune Response

Human erythropoietin (hEPO) expression was evaluated using intramuscular (I.M.) injections to examine its correlation with I.D. injection and immunological factors.[21] The delivery efficiency of mRNA encoding the hEPO protein to mice was assessed using 36 LNPs selected from ML analysis. Mice received I.M. administration of hEPO mRNA (10 μg) via the LNPs, and after 6 h, the hEPO expression levels in serum were quantified (**Figure 4a**). The hEPO expression patterns varied with FH types: FH-1 and FH-2 had lower expression, whereas FHs containing one (FH-3) or two (FH-4, FH-5, or FH-6) hydroxyl groups had higher expression (Figure 4b). FH-5 and FH-6 consistently showed high hEPO expression, indicating the significant influence of the FH on expression levels, with no clear patterns for EB or T. A strong correlation ($r > 0.71$) was observed between I.D. and I.M. hEPO expression results used for AI prediction (Figure 4c). To compare the immune responses, cytokines such as monocyte chemoattractant protein 1 (MCP-1) and interleukin 6 (IL-6) were measured in serum. MCP-1, a key inflammation initiator,[22] was detected in FH-5 and FH-6 regardless of mRNA expression levels. MCP-1 patterns varied with FH type: in FH-5, T-1 showed a decreasing trend with increasing EB length, whereas T-2 exhibited an increasing trend; in FH-6, MCP-1 increased with EB length in T-1 but decreased in T-2 (Figure 4d). This indicates that the FH structure influences MCP-1 expression levels, modified by EB and T types. IL-6, inducing local and systemic inflammation,[23] was elevated only in FH-1 and FH-2, independent of T variations, contrasting with MCP-1 findings.

To assess the immune response elicited by the LNPs, we employed an mRNA platform encoding the E6/E7 sequence of human papillomavirus (HPV) types 16 and 18.[24] Subsequently, the HPV mRNA was encapsulated within the selected LNPs. Mice received two immunizations with 10 μg/40 μL of HPV mRNA-LNP at one-week intervals, followed by analysis one week after the final immunization (Figure 4e). Serum levels of IgG1, IgG2a, and total immunoglobulin G (IgG) against the HPV protein were measured using enzyme-linked immunosorbent assay (ELISA). The HPV-loaded mRNA LNP induced high levels of total IgG in LNPs containing FH-5 and FH-6, whereas it did not change for the LNPs with FH-1, FH-2, FH-3, and FH-4 (Figure 4f). A balanced IgG1/IgG2a response was observed in all groups (Figure S4a, Supporting Information).

Numerous studies have demonstrated that T cells play a key role in immunity against cancer vaccines.[25] Thus, to ascertain the efficacy of HPV-loaded mRNA encapsulation on different selected LNPs, we investigated the T-cell responses using enzyme-linked immunospot (ELISpot) for interferon-$\gamma$ (IFN-$\gamma$) and ELISA for tumor necrosis factor-$\alpha$ (TNF-$\alpha$) and interleukin 2 (IL-2). IFN-$\gamma$-producing cells showed significantly higher activity in FH-5 and FH-6 groups compared to FH-1 to FH-4 groups (Figure 4g). TNF-$\alpha$ and IL-2 results showed similar trends (Figure S4b, Supporting Information). Overall, AI-synthesized ionizable lipids with FH-5 and FH-6, EB-5, EB-6, and EB-7, and T-1 and T-2 elicited strong immune responses. While the MCP-1 levels (Figure 4d) aligned with the total IgG levels and IFN-$\gamma$ cell numbers (Figure 4f,g), they did not accurately reflect differences between groups, suggesting MCP-1 levels provide a general trend but not detailed immune response predictions.

### 2.5. Correlation between Molecular Substructures and Immune Response

To further understand the dynamics of molecular structure for functional outcomes in the context of mRNA vaccine design, we investigated the correlation between the AI-predicted features and their implications for both I.M. expression and subsequent immune responses. Guided by AI predictions, our investigation focused on two pivotal structural characteristics–the FP169 ("number of phenolic hydroxyl group") and FP2 ("secondary carbon (carbon length of EB+T)")–as fundamental determinants of LNP performance for mRNA vaccine design.

Throughout our analysis, the significance of phenolic hydroxyl groups consistently emerged, exerting an apparent influence on all facets of LNP behavior. Variations in the "phenolic number" closely aligned with changes in both I.M. mRNA expression and immune response parameters, such as the total IgG levels and expression of inflammatory cytokines, underscoring the pivotal role of these structural properties in shaping the vaccine effectiveness (**Figure 5a–e**). However, when we shifted the focus to "secondary carbon" features that reflect the length (EB+T) of the encapsulated mRNA (Figure 5f–j), the crucial patterns across the diverse range of LNPs under investigation were elusive, despite a significant negative correlation with IFN-$\gamma$ (Figure 5h). Particularly, LNPs containing phenolic hydroxyl groups exhibited no negative trend between the number of secondary carbons and either I.M. mRNA expression or immune responses (Figure 5f–j). Consequently, to clarify why the carbon length variation of EB+T affects the immune response, we purposely selected LNPs that demonstrated distinct results in the immune response and precisely analyzed their morphology.

### 2.6. Correlation between Morphology of mRNA-LNPs and EB+T

To explore the association between the structures identified via AI prediction and the outcomes of the animal experiments, which included mRNA expressions, a scoring system ranging from 0 to 35 was implemented after ranking each analytical parameter. We selected the top three final LNPs based on the sum of the item-specific scores that were obtained previously (**Figure 6a**). Notably, all selected LNPs contained FH-5, and the selected LNPs only differed based on the carbon lengths of EB+T. Particularly, the LNPs of RANK 1 and RANK 2 only differed structurally by one carbon atom in the EB. Therefore, we verified the LNP morphology through cryogenic transmission electron microscopy (cryo-TEM). We also included EB-7 with the same FH group to analyze the effect of carbon length on the structural properties of the LNPs (Figure 6b). The LNPs were prepared for cryo-TEM by

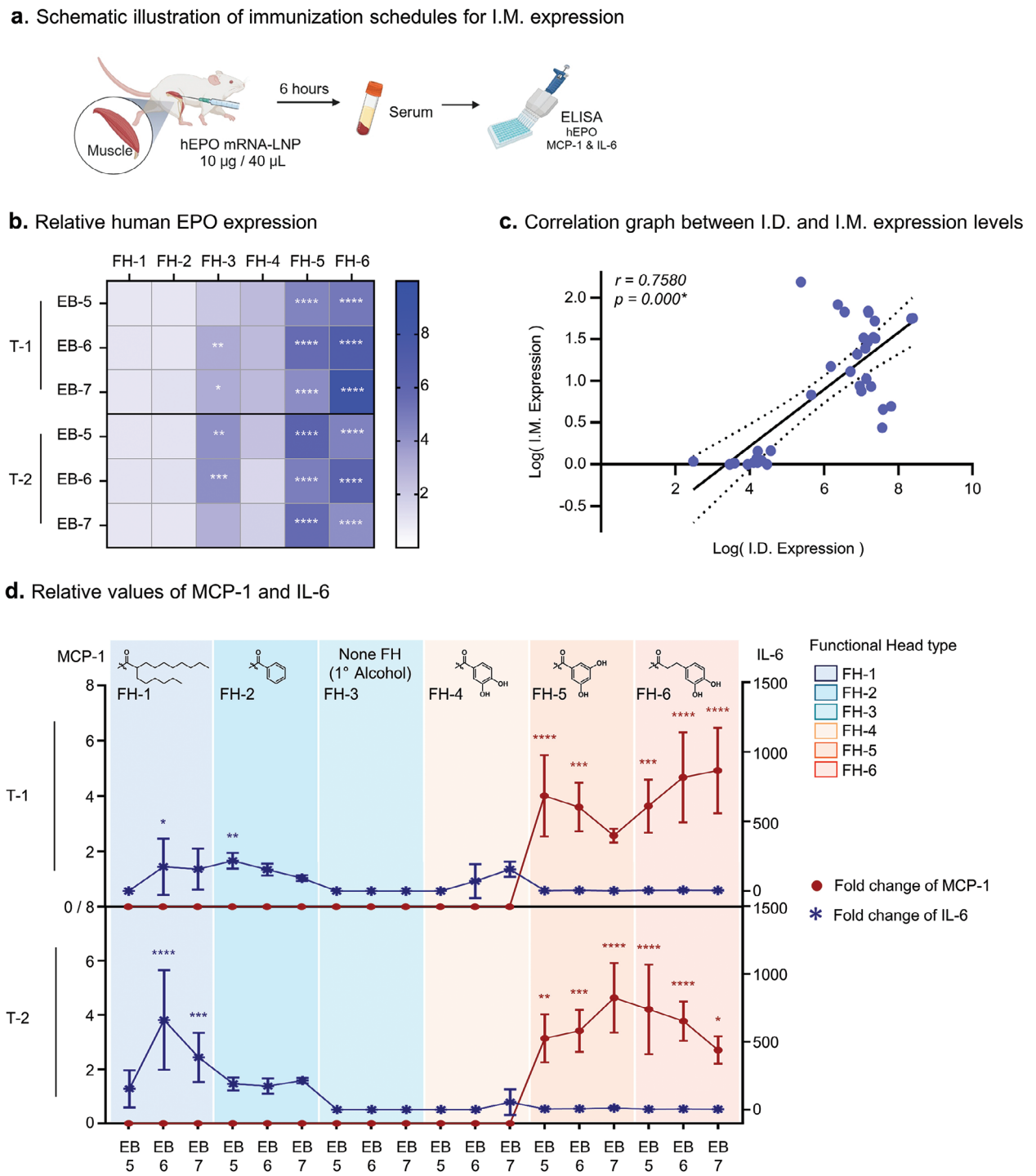


**Figure 4.** In vivo analysis of mRNA expression and immune response depending on LNP types. The results of all experiments were compared by substituting fold change based on the values of the Nil group included in the experiment in question. a) Schematic of immunization schedules for hEPO-LNP. Seven-week-old ICR (Institute of Cancer Research) mice were immunized with 10 μg/40 μL of hEPO-LNPs. Blood was collected 6 h later, and serum was separated. Subsequently, analysis was performed by ELISA. b) Relative hEPO expression. c) Correlation analysis to confirm the results of I.M. and I.D. expression. All results are logarithmically scaled. d) Relative values of MCP-1 and IL-6 in serum. e) Schematic illustration of immunization schedules for HPV-LNP. Seven-week-old C57BL/6 mice prime/boost immunized with LNPs encapsulating HPV mRNA (10 μg) or Dulbecco's phosphate-buffered saline through I.M. injection. f–g) Experimental results were compared by replacing the fold change with the values of the Nil group included in the experiment. f) Serum samples were collected from mice 1 week after the second immunization with mRNA-LNPs, and the levels of total IgG were measured using ELISA. g) Number of HPV peptide-specific IFN-γ cells in splenocytes was measured using ELISpot. Data are represented as the mean ± standard deviation. Statistical significance was analyzed using two-way ANOVA and defined as $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$. Graphical images were created with BioRender.com.

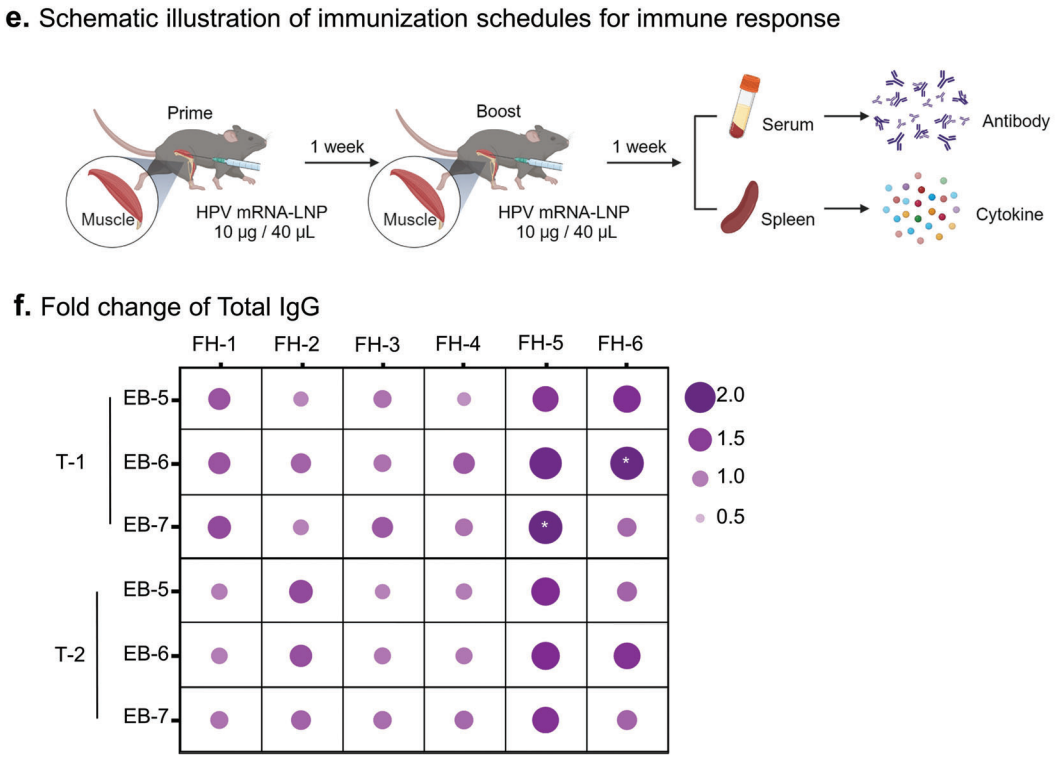


**Figure 4.** *Continued*

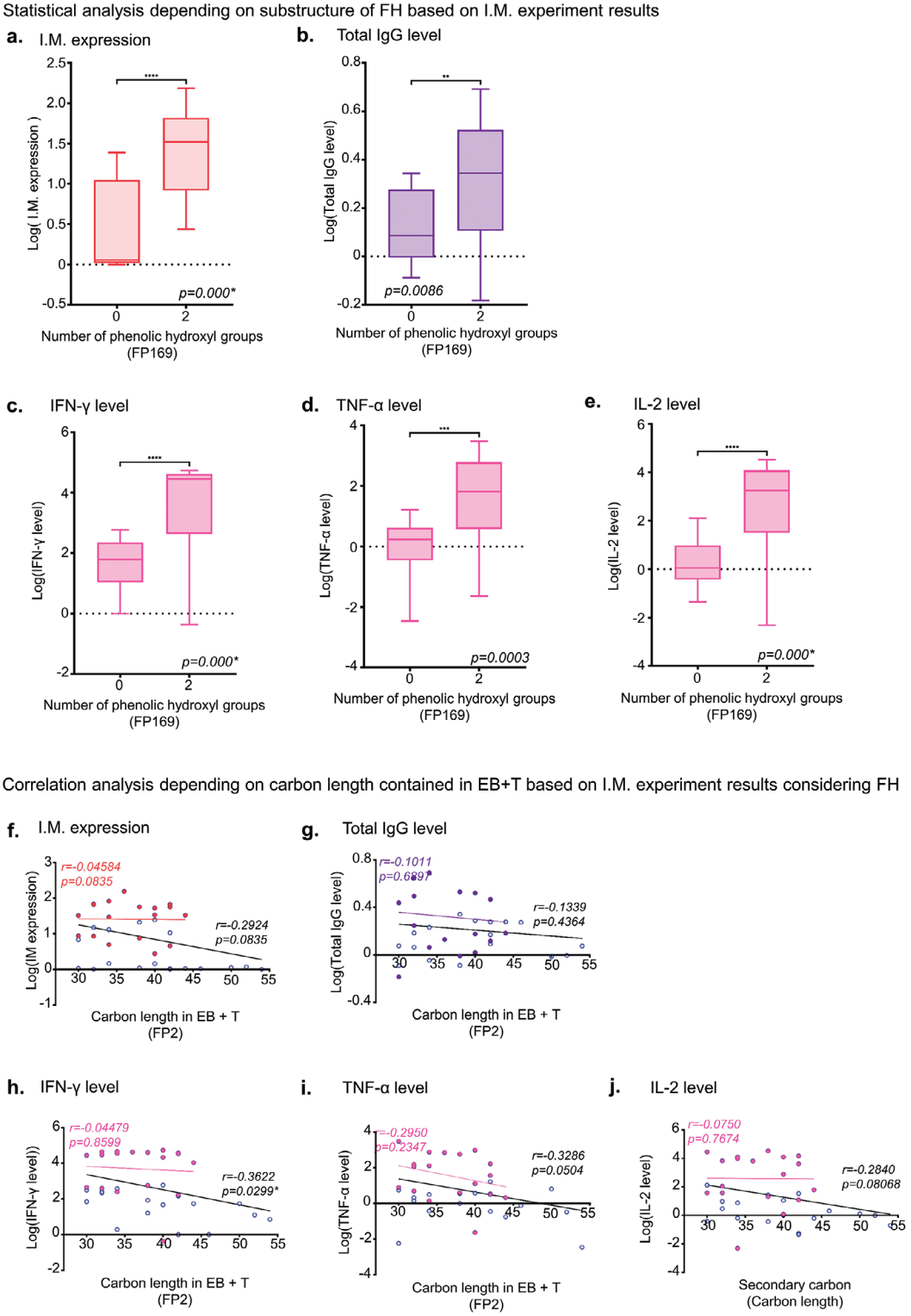


**Figure 5.** Analysis of the main substructural features in ionizable lipids. All results are logarithmically scaled. a–e) Statistical analysis depending on substructure of FH based on I.M. experiment results. f–j) Correlation analysis depending on carbon length of EB+T based on I.M. experiment results considering FH. FHs with phenolic hydroxyl groups (FH-4, FH-5, and FH-6) are indicated by light-colored dots. The black lines denote the correlation analysis for all 36 LNPs. The colored lines show the correlation for FHs with phenolic hydroxyl groups. In the box plots (a–e), the central line of each box denotes the median, the box edges indicate the interquartile range (IQR), and the whiskers represent the range of the data (1.5 times the IQR from the first and third quartile). Statistical significance was defined as $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. In the scatter plots (f–j), the solid lines represent the regression line. The correlation coefficient ($r$) is shown for each regression line.

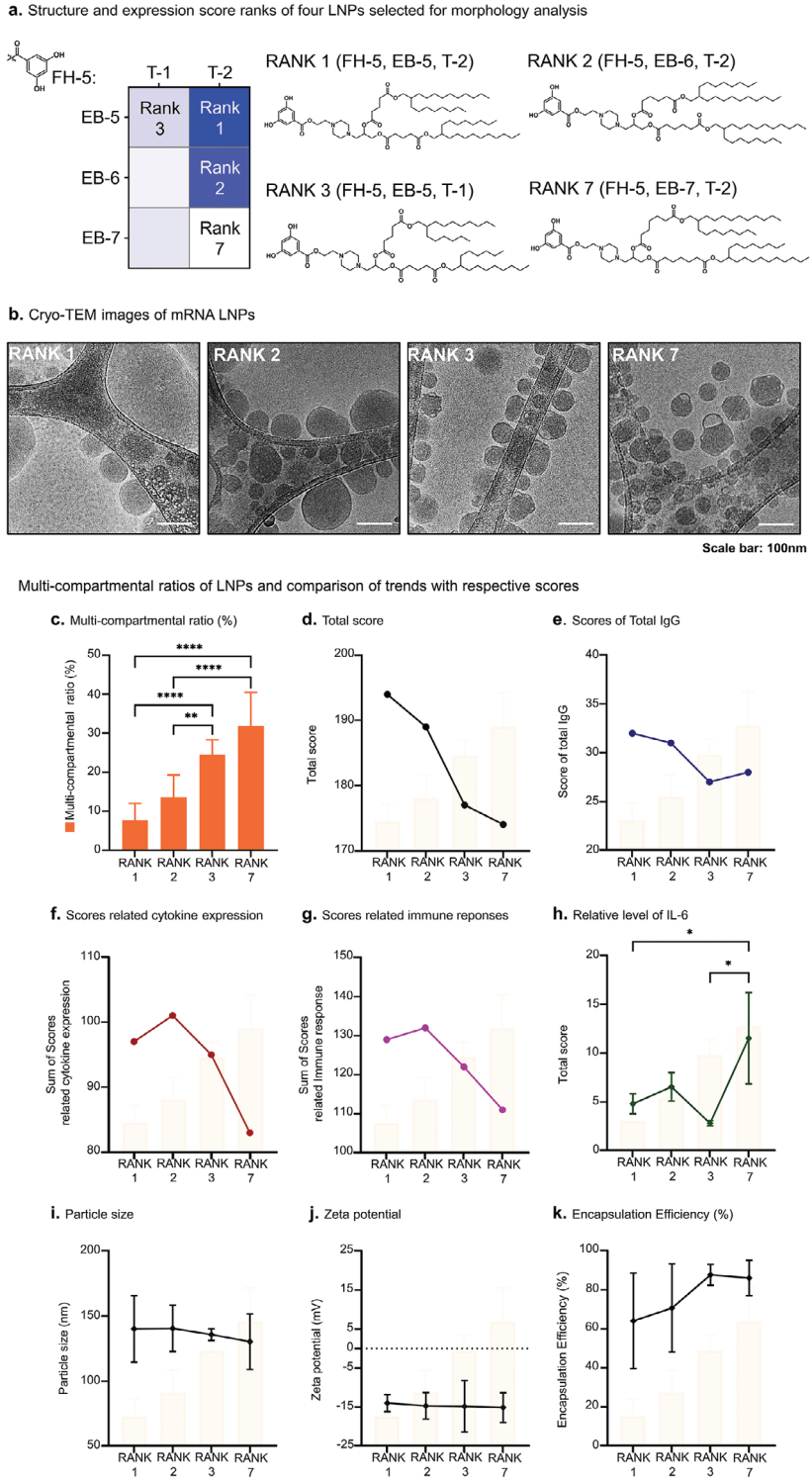


**Figure 6.** Structural relationship of LNPs with expression performance. a) Expression score ranks and the chemical structure of ranked LNPs. b, Cryo-transmission electron microscopy images of ranked LNPs (scale bar: 100 nm). c–k) Relationship between multi-compartmental ratio and expression performance of LNPs by expressing multi-compartmental ratio as a column in each graph. c, Multi-compartmental ratio of ranked LNPs. d) Total score (sum of expression level and immune response level) of ranked LNPs. e) Scores of total IgG of ranked LNPs. f) Scores of related cytokine expression (IFN-$\gamma$, IL-2, and TNF-$\alpha$) of ranked LNPs. g) Scores of related immune responses (total IgG and cytokine expression) of ranked LNPs. h) Relative level of IL-6 of ranked LNPs. The results were compared by substituting the fold change based on the values of the Nil group included in the experiment in question. This part is related to the IL-6 levels shown in Figure 4. i) Particle size of ranked LNPs. j) Zeta potential of ranked LNPs. k) Encapsulation efficiency (EE, %) of ranked LNPs. Data are represented as the mean ± standard deviation. Statistical significance was analyzed using two-way ANOVA and defined as $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$.

mounting them on a lacey carbon grid and freezing in liquid ethane. LNPs adhere to the surface of lacey carbon membranes because of their affinity for carbon-based materials.[26]

Most LNPs exhibit an electron-dense internal amorphous structure, thus achieving superior encapsulation of mRNA.[27] However, LNPs with a multi-compartmental structure have been observed, which result from the incomplete fusion of the initial LNPs during dialysis, which is a crucial step in LNP assembly.[28] Moreover, it has been reported that mRNA may reside in the solvent filled multi-compartment of LNPs,[29] which may affect the EE. Therefore, we evaluated the ratio of the population of multi-compartmental LNPs based on cryo-TEM images and attempted to analyze its impact based on key item-specific scores (Figure 6c–k). We confirmed that the multi-compartmental ratio gradually increased as the ranking of the ionizable lipids decreased (Figure 6c). Notably, the multi-compartmental ratio for RANK 3, with the shortest EB+T carbon length, was over 20%, indicating that ionizable lipids cannot be infinitely short–a suitable carbon length is required to maintain lipid membrane stability.

The multi-compartmental ratio was also negatively correlated with various functional parameters of the mRNA-LNPs, including the total score (Figure 6d), which is the sum of all analytes, that is, the combined scores for total IgG (Figure 6e), IFN-$\gamma$, IL-2, and TNF-$\alpha$ (Figure 6f), and the sum of the previous two scores (cytokines and total IgG) (Figure 6g). The levels of inflammatory cytokine IL-6 were relatively higher for RANK 7, which had the highest multi-compartmental ratio (Figure 6h). Combined, these findings suggest that the immune efficacy of the LNPs decreases as the EB+T chain length increases, and that the multi-compartmental ratio plays a significant role in the LNP performance.

As the multi-compartmental ratio increased, the size and zeta potential of the LNPs (as measured by DLS) gradually decreased, although the change was not statistically significant (Figure 6i,j). However, the size and zeta potential of RANK 1, RANK 2, RANK 3, and RANK 7 all fell within ranges favorable for mRNA delivery and expression, and the differences in these characteristics did not appear to result in significant variations in mRNA delivery efficiency. Notably, the EE increased slightly from RANK 1 to RANK 7. Slight increase of EE may have affected by the increase of the proportion of solvent-filled multi-compartment in which mRNA may reside,[29] Furthermore, RANK 1, with its low multi-compartmental ratio, had a relatively simple internal structure, which may limit its efficiency for mRNA encapsulation. By contrast, RANK 2, RANK 3, and RANK 7 have more complex internal structures, leading to more effective encapsulation. However, it is difficult to suggest that multi-compartmental ratio individually attributed to the EE it should be interpreted in a more complex context. Considering these findings, the EB+T chain length, derived through AI predictions, significantly impacts the functionality of LNPs. Thus, when designing ionizable lipids, it is crucial to select a carbon chain length that optimizes both membrane stability and performance.

## 3. Conclusion

This study highlighted the pivotal role of rational LNP design in the optimization of mRNA delivery for therapeutic applications. By leveraging ML techniques, we identified key structural features within ionizable lipids that significantly influenced mRNA encapsulation and expression efficiency. The findings emphasize the importance of phenolic hydroxyl groups and specific lipid composition ratios in enhancing the performance of LNPs. Experimental validation demonstrated the predictive power of the ML model, providing a robust framework for designing LNPs with tailored properties. Cryo-TEM analysis revealed the structural impact of carbon length on the LNP morphology and stability, further corroborating the ML predictions. By elucidating the intricate relationship between LNP composition, structure, and therapeutic efficacy, this study offers valuable insights for the rational design of LNPs for mRNA delivery, paving the way for the development of next-generation mRNA therapeutics with enhanced potency and safety profiles.

## 4. Experimental Section

*Feature Extraction (Molecular Description)*: Chemical substructures were extracted from ionizable lipid structures to examine their influence on I.D. mRNA expression using substructure fingerprints.[15a] Substructure fingerprints directly encode the molecular structures into binary bits, indicating the presence of specific substructures within a molecule. The SMARTS list of substructure patterns is predefined as a substructure dictionary, comprising 881 molecular substructures. For instance, the first substructure fingerprint (FP1) corresponds to primary carbon, whereas FP12 and FP169 represent alcohol and phenol, respectively, as listed in Table S2 (Supporting Information). Thus, the substructure fingerprint offers the advantage of establishing a one-to-one correspondence between each molecular structure and its fingerprint vector. PaDEL-Descriptor was utilized to convert the SMILES patterns of ionizable lipids into substructure fingerprint vectors.[18] Then, the occurrence count of each substructure in every ionizable lipid was extracted and utilized as a feature in the ML.

The compositions of the LNPs, including factors such as the N/P ratio, mass, and molar ratios of ionizable lipids, DSPC, DOPE, cholesterol, and PEG–lipid, were employed as features. In total, 314 features, consisting of 307 substructure fingerprints and 7 composition values, were utilized to train ML models for predicting I.D. expression. Each feature was standardized to have a mean of 0 and a standard deviation of 1 to avoid variability and bias in the ML models. The I.D. expression was observed four times as an output of the ML models. The four observations were averaged and converted to a logarithmic scale using a natural logarithm. Finally, the ML models were trained using 314 input features to forecast the average logarithmic-scaled I.D. expression.

*ML Models (RF)*: To predict the I.D. expression using the extracted features, RF, an ensemble learning method that utilizes multiple decision trees, was employed.[16] RF addresses both classification and regression problems using a bootstrap aggregation (bagging) process. In the bootstrap process, the observed data were randomly sampled with replacement to create multiple subsets. Subsequently, each decision tree was trained on a different subset to ensure input variation for each learner. The decision tree algorithm repeatedly conducts a greedy search to identify the optimal splitting criteria within the input data with the aim of minimizing variance. During the aggregation process, the final prediction was determined by combining the predicted results from multiple decision trees trained on different subsets. The bagging process helps to reduce the prediction variance, particularly that arising from outliers. As RF demonstrates superior performance compared to other ML techniques, particularly owing to its robustness to overfitting caused by outliers, it was adequate for predicting the I.D. expression, given the significant variation in the experimental data. Furthermore, because the splitting criterion is created by evaluating all possible features individually, the inclusion of multiple decision trees can provide the feature importance that will aid in

examining the impact of the chemical substructures and molecular ratios on the I.D. expression.

*Evaluation Metrics*: To evaluate the prediction performance of the RFs, tenfold cross-validation was employed. In this method, 213 LNPs were randomly divided into 10 equal-sized subgroups. During each iteration, one subgroup was used to test the model, and the remaining nine subgroups were used to train the model. This process was repeated ten times, and the results, including the model performance metrics and feature importance scores, were averaged across the folds.

To evaluate the RF regression model, two metrics, namely $R^2$ and PCC, were used to assess the model performance. $R^2$ was used to assess the goodness of fit of a regression model by comparing the variance of the observed data with the variance of the predicted values obtained from the model. It is calculated as the ratio of the regression sum of squares ($SS_{residual}$) to the total sum of squares ($SS_{total}$), which is subtracted from 1. The formula for $R^2$ is as follows:

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)}{\sum_i (y_i - \mu_y)} \quad (1)$$

where $y_i$ represents the observed data, $\hat{y}_i$ is the predicted value, and $\mu_y$ is the mean of the observed data. The $R^2$ value ranges from 0 to 1, with higher values indicating a better fit of the model to the data. The PCC is a statistical measure used to assess the linear relationship between two continuous variables. This is calculated by dividing the covariance of two variables by the product of their standard deviations. The formula for PCC is as follows:

$$PCC = \frac{cov_{y,\hat{y}}}{\sigma_y \sigma_{\hat{y}}} = \frac{\sum_i (y_i - \mu_y)(\hat{y}_i - \mu_{\hat{y}})}{\sqrt{\sum_i (y_i - \mu_y)^2} \sqrt{\sum_i (\hat{y}_i - \mu_{\hat{y}})^2}} \quad (2)$$

where $\mu_{\hat{y}}$ is the mean of the predicted values. The PCC ranges from −1 to 1, with −1 indicating a perfect negative correlation, 0 indicating no linear relationship, and 1 indicating a perfect positive correlation.

*Statistical Analysis*: The influence of the substructural patterns and compositions on the expressions was evaluated by statistically comparing the experimental results of the I.D. and I.M. expressions. All I.D. and I.M. expression results were transformed using a natural logarithm, consistent with the approach used in the ML process. For screening using the I.D. expression of 213 LNPs, features that exhibited high importance scores, including the count of substructure patterns and compositions, were analyzed using the following steps. First, the Kolmogorov–Smirnov test was conducted for each feature to examine whether the distribution of each group was normal. If all groups exhibited a normal distribution, a comparison between the two groups was performed using Welch's t-test, considering unequal sample sizes. For comparisons involving three or four groups, analysis of variance (ANOVA) was utilized. The PCC was employed to analyze the linear relationship between the features and experimental results when the feature range varied. Nonparametric tests were performed if at least one group exhibited a non-Gaussian distribution. Specifically, the Mann–Whitney U test, Kruskal–Wallis H-test, and Spearman rank-order correlation were substituted for Welch's t-test, ANOVA, and PCC, respectively. Similarly, for the analysis of I.M. expression, the results based on the phenol count were statistically tested using Welch's t-test, and the results based on the count of secondary carbons were analyzed using PCC, as all groups exhibited a normal distribution. Statistical tests were performed in Prism9 (GraphPad Software Inc.) and MATLAB 2021b (The MathWorks Inc.).

*Synthesis of Ionizable Lipids*: EB–T chain synthesis and a representative synthesis method for cyclic diamine ionizable lipids are described in the Supplementary Methods and Figures S5–S9 (Supporting Information).

*mRNA Synthesis*: Two types of RNA platforms were used: an internal ribosome entry site (IRES)-based cap-independent RNA platform and a cap-dependent RNA platform with capping.

The DNA template for the IRES-based cap-independent RNA platform was designed using the ribosome entry site of the encephalomyocarditis virus.[30] Moreover, the DNA template had a multicloning site with four types of restriction enzymes, which were utilized by inserting genes of interest, such as RLuc, whose in vivo expression had been previously evaluated. This platform is referred to as CUK2.

The DNA template for the cap-dependent RNA platform consisted of a 5′ untranslated region (UTR), 3′ UTR, multicloning site, and poly-A tail, as reported in a previous study.[31] Similar to the cap-independent RNA platform, it can be used to encode antigens such as hEPO to assess the in vivo expression and E-6/E-7 of HPV 16 and 18 to assess the immune response.[24] This platform is referred to as CUK3-1. Plasmids containing the desired sequences were linearized using the Not1 restriction enzyme. The DNA was then purified and quantified using spectrophotometry. mRNA was generated from the DNA templates by in vitro transcription using an EZ T7 high yield in vitro transcription kit (Enzynomics, Daejeon, South Korea). Moreover, the capping of the CUK3-1 mRNA platform was performed using SC101 (ST Pharm Co., Ltd. Seoul, South Korea), and uridine-5-triphosphate (UTP) was replaced with N1-methylpseudouridine (TriLink BioTechnologies, California, USA). The mRNA platform reactions were efficiently transcribed by incubating overnight at 37 °C. Any remaining DNA on the template was removed via deoxyribonuclease 1 treatment at the same temperature for 30 min. The resulting transcripts were precipitated by adding lithium chloride and incubating at −20 °C for 30 min. After centrifugation at 13 000 rpm for 15 min, the pellets were washed with 70% ethanol (700 µL). The mRNA pellets were resuspended in sterile distilled water and subjected to a second purification step using cellulose to minimize double-stranded RNA contamination. The concentration of synthesized mRNA was determined using a NanoDrop 2000 spectrophotometer. The resuspended RNA was then aliquoted and stored at −80 °C for subsequent experimental procedures. mRNAs containing RLuc, hEPO, and HPV sequences in open reading frames were named CUK2 RLuc, CUK3-1 hEPO, and CUK3-1 HPV mRNA, respectively.

*Formulation of mRNA-Loaded LNPs*: Diamine-based ionizable lipids were synthesized as described in the Supplementary Methods. DSPC, DOPE, cholesterol, and DMG-PEG2000 were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA). SM-102 was purchased from Hanmi Fine Chemical Co. Ltd. (Siheung, South Korea). Other reagents were purchased from Sigma–Aldrich. The LNP components listed in Table S2 (Supporting Information), including phospholipids, cholesterol, and DMG-PEG2000, were dissolved in a mixed chloroform/methanol solvent (1:1, v/v) at a concentration of 50 µg µL$^{-1}$. The lipid components were then mixed in the molar ratios shown in Figure 1 and concentrated under reduced pressure to form a lipid film. The lipid film, which was dissolved in ethanol, and mRNA, which was dissolved in citric acid buffer (pH 4.0, 50 mM), were mixed in a 1:3 volume ratio to prepare mRNA-loaded LNPs using enCell (enParticle, Busan, South Korea). The mRNA-loaded LNPs were washed twice using 1X Dulbecco's phosphate-buffered saline (DPBS), centrifuged (LABOGENE Co, South Korea), and concentrated using Amicon Ultra-15 centrifugal filters (Merck Millipore, Germany).

*Z-Average, Đ, and Zeta Potential Measurements of mRNA-LNPs*: Measurements and analyses were conducted after 100-fold dilution of the LNPs in 1X DPBS for Z-average and Đ measurements and in deionized water for zeta potential measurements. The Z-average, Đ, and zeta potential of the mRNA-loaded LNPs were measured using a Zetasizer Nano ZS (Malvern, USA) with 12 mm square polystyrene cuvettes (DTS0012, for Z-average and Đ) and folded capillary zeta cells (DTS1070, for zeta potential).

*Measurement of EE for mRNA-Loaded LNPs*: The efficacy of the mRNA-loaded LNPs was evaluated by Fluorescent Quanti-it RiboGreen assay.[32] Two types of LNPs were prepared to ensure the identification of all the RNA inside and outside the LNPs: a diluted LNP solution lysed with Tris–EDTA (TE) buffer supplemented with 0.5% Triton X, and an LNP solution diluted with only the TE buffer. The mRNA-loaded LNPs were then diluted 200-fold using 1X TE buffer, mixed with the Fluorescent Quant-it RiboGreen dye (Invitrogen, USA), and loaded into 96-well plates. The fluorescence intensity of the 96-well plate was measured at 520 nm using the Synergy H1 microplate reader (BioTek, USA) using an excitation wavelength of 485 nm. The amount of mRNA in the sample was determined by

analyzing the mRNA EE of the LNPs using a standard curve formed after serial dilutions of the loaded mRNA in 1X TE buffer.

*Animal Experiments*: Six-week-old female ICR mice ($n = 3$) were obtained from Daehan Biolink and used for luciferase activity and hEPO expression measurements. BALB/c mice were used for immunization experiments. All mice were acclimated for 1 week and housed under specific pathogen-free conditions at a controlled temperature of 23 ± 2 °C with a 12-h light/dark cycle. The animal experiments were approved by the Institutional Animal Care and Use Committee of the Catholic University of Korea (approval number: CUK-IACUC-2023-023) in strict accordance with ethical guidelines.

*RLuc Assay to Optimize LNP Formulations for In Vivo Expression*: To compare LNP-specific RLuc expression, the study immunized mice via I.D. injection into the ear with CUK RLuc-loaded LNPs (5 µg/20 µL) using a 30G insulin syringe (BD, NJ, USA). The mice were anesthetized using 5% isoflurane prior to injection. After 6 h, the mice were euthanized, and their ears were harvested. The harvested ears were placed in *Renilla* lysis buffer (300 µL) and finely homogenized using a homogenizer. After a brief 1-min spin-down, the RLuc activity was measured using the RLuc assay system and GloMax instrument following the manufacturer's instructions.

*Serum Analysis*: CUK3-1 hEPO-loaded LNPs (10 µg/40 µL) were injected intramuscularly into ICR mice. After 6 h, the mice were euthanized and serum was collected for analysis, as following the protocol previously established.[33] The hEPO levels were measured using a hEPO DuoSet ELISA kit (DY286, R&D Systems, USA), and the cytokine levels were quantified using an uncoated cytokine ELISA kit (Invitrogen, USA) according to the manufacturer's specifications.

*Immunization Using CUK3-1 HPV*: To analyze the immune response to the antigen delivered by LNPs, 6-week-old C57BL/6J mice were immunized twice with CUK3-1 HPV-loaded LNPs (10 µg/40 µL) at 1-week intervals via I.M. injection in the upper thigh. The mice received CUK3-1 HPV-loaded LNPs (10 µg/40 µL) for each injection. One week after the final immunization, all animals were euthanized. Blood and spleen samples were collected for further analysis.

*Enzyme-Linked Immunospot Assay*: Splenocytes ($5 \times 10^5$) were stimulated with a mixture of HPV-specific peptides (HPV serotype 16 E6 (YDFAFRDL) 5 µg mL$^{-1}$, HPV serotype 16 E7 (RAHYNIVTF) 2 µg mL$^{-1}$, and HPV 18 E6 (KCIDFYSRI) 2 µg mL$^{-1}$) for 48 h at 37 °C. Peptides were synthesized using Peptron (Daejeon, South Korea). IFN-$\gamma$ secretion by T cells was measured using a mouse IFN-$\gamma$ ELISpotBASIC assay (Mabtech, Stockholm, Sweden) according to the manufacturer's instructions.

*Enzyme-Linked Immunosorbent Assay*: ELISAs were conducted to assess the levels of antigen-specific total IgG, IgG1, and IgG2a in the mouse serum. Ninety-six-well plates (Corning, Corning, NY, USA) were coated with an HPV protein mixture (100 ng well$^{-1}$) and incubated for 2 h at room temperature (20–25 °C). Following the incubation period, the wells were washed three times using phosphate-buffered saline (PBS; 200 µL) containing 0.05% Tween 20 (PBS-T) and subsequently blocked with blocking buffer (PBS containing 1% bovine serum albumin; 100 µL) for 1 h at room temperature. The serum samples were diluted 1/50 in the blocking buffer, added to the wells, and incubated for 2 h at room temperature. After incubation, the wells were washed three times using PBS-T (200 µL). Anti-mouse IgG1 (Invitrogen, Carlsbad, CA, USA), IgG2c (Novus Biologicals, Centennial, CO, USA), and IgG-HRP (Bethyl Laboratories, Inc., Montgomery, TX, USA)-conjugated antibodies were diluted 1/5000 in blocking buffer and incubated for 1 h at room temperature. Following five washes with PBS-T, tetramethylbenzidine substrate (Invitrogen) was added, and the samples were incubated for 10 min. The reaction was terminated by adding 2 N $H_2SO_4$. The optical density (OD) was measured at 450 nm using a GloMax Explorer microplate reader (Promega).

To evaluate the cytokine levels in the supernatants of splenocyte cultures, splenocytes were obtained from immunized mice and seeded at a density of $5 \times 10^5$ cells well$^{-1}$ in a 96-well plate. Subsequently, the cells were stimulated with 500 ng well$^{-1}$ of an HPV-specific peptide mixture and incubated for 72 h at 37 °C. After the incubation period, the concentrations of IL-2 and TNF-$\alpha$ were determined using ELISA kits (Invitrogen and Thermo Fisher Scientific Inc.) following the manufacturer's specifications. OD values were measured at 450 nm using a GloMax Explorer microplate reader (Promega). The concentrations of these cytokines were calculated based on their respective standard curves.

*Cryogenic Transmission Electron Microscopy*: The structure of the mRNA-loaded LNPs was analyzed using cryo-TEM. Cryogenic samples were prepared prior to performing cryo-TEM. A lacey carbon grid (Lacey Carbon, 200 mesh Cu, Ted Pella Inc., USA) was treated with a glow discharger at 15 mA for 60 s and loaded into a Vitrobot Mark IV (Thermo Fisher Scientific, SNU CMCI). The temperature and humidity inside the Vitrobot chamber were maintained at 15 °C and 100%, respectively, during sample preparation. The mRNA LNP solution (3 µL) with a concentration of 0.5 µg µL$^{-1}$ was mounted on the grid, and excess solution was blotted away. The blotted specimens were immediately plunged in liquid ethane for freezing, and stored in liquid nitrogen before cryo-TEM imaging.

The cryogenic sample was transferred to a cryo-TEM holder (626 single tilt cryo-TEM holder, Gatan, USA), which maintained the temperature of the sample at ≈−180 °C. The cryo-TEM holder was loaded into a transmission electron microscope (JEM-2100F, JEOL, Japan), and LNP images were acquired at an accelerating voltage of 200 kV with a total dose of ~15 $e^-$ Å$^{-2}$.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

S.-H.B., H.C., J.L., and M.-H.K. contributed equally to this work. This work was supported by grants from the Ministry of Food and Drug Safety (grant number: 22213MFDS421, J.-H.N.; RS-2023-00217074, J.L.), National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT; No. 2023R1A2C1004182 to K.L.) and partially supported by the Brain Korea 21 Four Program.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Keywords

lipid nanoparticles (LNPs), machine learning (ML) techniques, mRNA expression efficiency, mRNA vaccines, structural optimization

Received: July 8, 2024
Revised: August 17, 2024
Published online:

[1] N. Pardi, M. J. Hogan, F. W. Porter, D. Weissman, *Nat. Rev. Drug Discovery* **2018**, *17*, 261.
[2] X. Hou, T. Zaks, R. Langer, Y. Dong, *Nat. Rev. Mater.* **2021**, *6*, 1078.
[3] X. Han, H. Zhang, K. Butowska, K. L. Swingle, M.-G. Alameh, D. Weissman, M. J. Mitchell, *Nat. Commun.* **2021**, *12*, 7233.
[4] C. Hald Albertsen, J. A. Kulkarni, D. Witzigmann, M. Lind, K. Petersson, J. B. Simonsen, *Adv. Drug Delivery Rev.* **2022**, *188*, 114416.

[5] Y. Xu, A. Golubovic, S. Xu, A. Pan, B. Li, *J. Mater. Chem. B* **2023**, *11*, 6527.
[6] P. Patel, N. M. Ibrahim, K. Cheng, *Trends Pharmacol. Sci.* **2021**, *42*, 448.
[7] M. Schlich, R. Palomba, G. Costabile, S. Mizrahy, M. Pannuzzo, D. Peer, P. Decuzzi, *Bioeng. Transl. Med.* **2021**, *6*, e10213.
[8] K. J. Hassett, K. E. Benenato, E. Jacquinet, A. Lee, A. Woods, O. Yuzhakov, S. Himansu, J. Deterling, B. M. Geilich, T. Ketova, C. Mihai, A. Lynn, I. McFadyen, M. J. Moore, J. J. Senn, M. G. Stanton, Ö. Almarsson, G. Ciaramella, L. A. Brito, *Mol. Ther. Nucleic Acids* **2019**, *15*, 1.
[9] O. S. Fenton, K. J. Kauffman, R. L. McClellan, J. C. Kaczmarek, M. D. Zeng, J. L. Andresen, L. H. Rhym, M. W. Heartlein, F. DeRosa, D. G. Anderson, *Angew. Chem., Int. Ed.* **2018**, *57*, 13582.
[10] J. Gilleron, W. Querbes, A. Zeigerer, A. Borodovsky, G. Marsico, U. Schubert, K. Manygoats, S. Seifert, C. Andree, M. Stöter, H. Epstein-Barash, L. Zhang, V. Koteliansky, K. Fitzgerald, E. Fava, M. Bickle, Y. Kalaidzidis, A. Akinc, M. Maier, M. Zerial, *Nat. Biotechnol.* **2013**, *31*, 638.
[11] a) M. M. Billingsley, N. Singh, P. Ravikumar, R. Zhang, C. H. June, M. J. Mitchell, *Nano Lett.* **2020**, *20*, 1578; b) H. Ni, M. Z. C. Hatit, K. Zhao, D. Loughrey, M. P. Lokugamage, H. E. Peck, A. D. Cid, A. Muralidharan, Y. Kim, P. J. Santangelo, J. E. Dahlman, *Nat. Commun.* **2022**, *13*, 4766; c) K. Xu, Y. Xu, J. Sun, X. Cheng, C. Lu, W. Chen, B. He, T. Jiang, *Nano Res.* **2024**, *17*, 7357.
[12] a) L. Wang, Z. -H. You, X. Chen, X. Yan, G. Liu, W. Zhang, *Curr. Protein Pept. Sci.* **2018**, *19*, 445; b) Z. Li, P. Han, Z. H. You, X. Li, Y. Zhang, H. Yu, R. Nie, X. Chen, *Sci. Rep.* **2017**, *7*, 11174; c) Y. Zhang, Z. Jiang, C. Chen, Q. Wei, H. Gu, B. Yu, *Interdiscip. Sci.* **2022**, *14*, 311; d) Y. Li, X.-Z. Liu, Z.-H. You, L.-P. Li, J.-X. Guo, Z. Wang, *Int. J. Intell. Syst.* **2021**, *36*, 593.
[13] a) W. Wang, S. Feng, Z. Ye, H. Gao, J. Lin, D. Ouyang, *Acta Pharm. Sin. B* **2022**, *12*, 2950; b) P. J. Harrison, H. Wieslander, A. Sabirsh, J. Karlsson, V. Malmsjö, A. Hellander, C. Wählby, O. Spjuth, *Nanomedicine* **2021**, *16*, 1097.
[14] a) A. Tharwat, Y. S. Moemen, A. E. Hassanien, *J. Biomed. Inform.* **2017**, *68*, 132; b) X. Yan, T. Yue, D. A. Winkler, Y. Yin, H. Zhu, G. Jiang, B. Yan, *Chem. Rev.* **2023**, *123*, 8575.
[15] a) J. Shen, F. Cheng, Y. Xu, W. Li, Y. Tang, *J. Chem. Inf. Model.* **2010**, *50*, 1034; b) A. A. Öztürk, A. B. Gündüz, O. Ozisik, *Comb Chem. High Throughput Screen* **2018**, *21*, 693; c) R. Maharjan, S. Hada, J. E. Lee, H. -K. Han, K. H. Kim, H. J. Seo, C. Foged, S. H. Jeong, *Int. J. Pharm.* **2023**, *640*, 123012.
[16] L. Breiman, *Machine Learning* **2001**, *45*, 5.
[17] C. Laggner, *SMARTS patterns for functional group classification*, Inte: Ligand Software-Entwicklungs und Consulting GmbH, **2005**, Available at: https://github.com/openbabel/openbabel/blob/master/data/SMARTS_InteLigand.txt.
[18] C. W. Yap, *J. Comput. Chem.* **2011**, *32*, 1466.
[19] A. K. Blakney, P. F. McKay, B. I. Yus, Y. Aldon, R. J. Shattock, *Gene Ther.* **2019**, *26*, 363.
[20] E. A. Kataev, T. A. Shumilova, B. Fiedler, T. Anacker, J. Friedrich, *J. Org. Chem.* **2016**, *81*, 6505.
[21] S. Sabnis, E. S. Kumarasinghe, T. Salerno, C. Mihai, T. Ketova, J. J. Senn, A. Lynn, A. Bulychev, I. McFadyen, J. Chan, Ö. Almarsson, M. G. Stanton, K. E. Benenato, *Mol. Ther.* **2018**, *26*, 1509.
[22] M. Kim, M. Jeong, G. Lee, Y. Lee, J. Park, H. Jung, S. Im, J. -S. Yang, K. Kim, H. Lee, *Bioeng. Transl. Med.* **2023**, *8*, e10556.
[23] S. Ndeupen, Z. Qin, S. Jacobsen, A. Bouteau, H. Estanbouli, B. Z. Igyártó, *iScience* **2021**, *24*, 103479.
[24] S. Lee, H. Yoon, S. H. Hong, S. P. Kwon, J. J. Hong, H. W. Kwak, H. -J. Park, S. Yoo, S. -H. Bae, H. -J. Park, J. Lee, Y. -J. Bang, Y. -S. Lee, J. -Y. Kim, S. Yoon, G. Roh, Y. Cho, Y. Kim, D. Kim, S. -I. Park, D. -H. Kim, S. Lee, A. Oh, D. Ha, S. -Y. Lee, M. Park, E. -H. Hwang, G. Bae, E. Jeon, S. H. Park, et al., *J. Med. Virol.* **2023**, *95*, e29309.
[25] A. D. Waldman, J. M. Fritz, M. J. Lenardo, *Nat. Rev. Immunol.* **2020**, *20*, 651.
[26] R. M. Glaeser, *Curr. Opin. Colloid Interface Sci.* **2018**, *34*, 1.
[27] a) L. Schoenmaker, D. Witzigmann, J. A. Kulkarni, R. Verbeke, G. Kersten, W. Jiskoot, D. J. A. Crommelin, *Int. J. Pharm.* **2021**, *601*, 120586; b) A. K. Leung, I. M. Hafez, S. Baoukina, N. M. Belliveau, I. V. Zhigaltsev, E. Afshinmanesh, D. P. Tieleman, C. L. Hansen, M. J. Hope, P. R. Cullis, *J. Phys. Chem. C Nanomater. Interfaces* **2012**, *116*, 18440; c) S. Patel, N. Ashwanikumar, E. Robinson, Y. Xia, C. Mihai, J. P. Griffith, 3rd, S. Hou, A. A. Esposito, T. Ketova, K. Welsher, J. L. Joyal, Ö. Almarsson, G. Sahay, *Nat. Commun.* **2020**, *11*, 983.
[28] a) J. A. Kulkarni, M. M. Darjuan, J. E. Mercer, S. Chen, R. van der Meel, J. L. Thewalt, Y. Y. C. Tam, P. R. Cullis, *ACS Nano* **2018**, *12*, 4787; b) M. J. Carrasco, S. Alishetty, M. -G. Alameh, H. Said, L. Wright, M. Paige, O. Soliman, D. Weissman, T. E. t. Cleveland, A. Grishaev, M. D. Buschmann, *Commun. Biol.* **2021**, *4*, 956.
[29] M. L. Brader, S. J. Williams, J. M. Banks, W. H. Hui, Z. H. Zhou, L. Jin, *Biophys. J.* **2021**, *120*, 2766.
[30] H. L. Ko, H. -J. Park, J. Kim, H. Kim, H. Youn, J. -H. Nam, *J. Microbiol. Biotechnol.* **2019**, *29*, 127.
[31] H. J. Park, Y. J. Bang, S. P. Kwon, W. Kwak, S. I. Park, G. Roh, S. H. Bae, J. Y. Kim, H. W. Kwak, Y. Kim, S. Yoo, D. Kim, G. Keum, E. K. Bang, S. H. Hong, J. H. Nam, *NPJ Vaccines* **2023**, *8*, 84.
[32] R. E. McKenzie, J. J. Minnell, M. Ganley, G. F. Painter, S. L. Draper, *Curr. Protoc.* **2023**, *3*, e898.
[33] S. H. Bae, S. Yoo, J. Lee, H. J. Park, S. P. Kwon, H. Jin, S. I. Park, Y. S. Lee, Y. J. Bang, G. Roh, S. Lee, S. B. Youn, I. W. Kim, H. R. Oh, A. K. El-Damasy, G. Keum, H. Kim, H. Youn, J. H. Nam, E. K. Bang, *Bioact. Mater.* **2024**, *38*, 486.