Lipid Nano particle Formulation Optimization via Interpolation and 3D Visualization

Jialing Ma[1]. Shengli Yang[1]. Gaoshuai Chen[2].

[1]*College of pharmacy, Zhejiang University of Technology, China*
[2]*College of Mechanical Engineering, University of Shanghai for Science and Technology, China*

Abstract: *Lipid nanoparticles (LNPs) show promising potential as gene delivery vectors in vaccine development and gene therapy. However, their formulation optimization faces challenges due to sparse experimental data. This study presents an innovative approach combining interpolation techniques with three-dimensional visualization analysis to systematically investigate the relationship between LNP component ratios and transfection efficiency. A high-resolution 85×85×85 dataset was constructed through linear interpolation of single-point sampling experimental data, focusing on the interactions among three critical components: cationic lipids, DSPC, and DMG-PEG2000. The data was visualized using three-dimensional heat maps and quantitatively analyzed through fourth-order polynomial fitting, incorporating outlier elimination strategies to enhance model reliability. Results revealed complex synergistic effects among components: cationic lipid content showed negative correlation with transfection efficiency, while DSPC demonstrated an overall promoting effect within a specific concentration range (4.0-16.0%). Optimal transfection efficiency was achieved under conditions of lower cationic lipid content and higher DSPC content, while maintaining DMG-PEG2000 content within an appropriate range (0.50-2.50%). This study effectively addresses the data sparsity challenge in LNP formulation screening while providing quantitative basis for component ratio optimization. These findings offer crucial guidance for rational design of LNP delivery systems, bearing significant implications for advancing gene therapy and mRNA vaccine development.*

Keywords: *Component optimization; Data interpolation; Lipid nanoparticles; Three-dimensional visualization; Transfection efficiency*

I. Introduction

Lipid nanoparticles (LNPs) have emerged as highly effective gene delivery vectors, demonstrating exceptional potential in vaccine development, gene therapy, and drug delivery applications[1]. Their outstanding in vivo stability and efficient gene transfection capabilities have led to successful clinical implementations, most notably in mRNA-based COVID-19 vaccines[2]. However, the performance of LNPs is critically dependent on the precise ratio of their constituents, including

cationic lipids, helper lipids ,such as DSPC(1,2-distearoyl-sn-glycero-3-phosphocholine), and PEG-modified lipids ,such as DMG-PEG2000(dimyristoyl glycerol-PEG2000)[3]. Consequently, the screening and optimization of LNP formulations represent crucial steps in achieving optimal transfection efficiency[4].

While formulation screening of LNPs holds significant experimental importance, researchers frequently encounter challenges related to data sparsity[5]. Practical constraints of time and resources often limit experiments to single-point sampling, resulting in discontinuous experimental data[6]. This characteristic not only restricts comprehensive understanding of the relationship between component ratios and transfection efficiency but also complicates the extraction of meaningful insights from experimental data[7]. Therefore, developing methods to extract comprehensive information from limited experimental data has become a critical challenge in LNP formulation screening.

To address this challenge, this study presents an innovative solution based on interpolation methods. By expanding single-point sampling experimental data through interpolation, we constructed a more dense dataset, enabling a more comprehensive analysis of the relationships between transfection efficiency and the ratios of cationic lipids, DSPC, and DMG-PEG2000. Furthermore, this study incorporated three-dimensional heat maps and fitting curve analysis to enhance data visualization and interpretation capabilities. During the fitting curve analysis, data quality was improved through outlier elimination, enhancing the reliability of fitting results.

The methodology presented in this study not only provides novel data processing approaches for LNP formulation screening but also offers valuable reference for addressing data sparsity issues in similar experimental contexts. Through the integration of interpolation methods with heat map visualization and fitting curve analysis, this study successfully elucidated the complex relationships between component ratios and transfection efficiency, establishing a foundation for further optimization of LNP formulations.

II. Research Status

Recent years have witnessed the emergence of lipid nanoparticles (LNPs) as crucial tools in gene therapy and vaccine development, owing to their exceptional gene delivery capabilities. The performance of LNPs is intricately dependent on the ratios of their constituent components, particularly the proportions of cationic lipids, DSPC, and PEG-modified lipids[3]. Consequently, the screening and optimization of LNP formulations represent critical steps in achieving efficient gene delivery[4]. While previous studies have demonstrated the significant impact of formulation screening on transfection efficiency, researchers consistently encounter challenges related to data sparsity and discontinuity in experimental processes[8].

Current research in LNP formulation screening primarily focuses on optimizing experimental design and data processing methodologies[9]. Traditional formulation screening approaches typically employ full factorial design or response surface methodology (RSM) to explore the relationships between component ratios and performance through multiple experimental

groups[10]. However, constraints in experimental costs and time often restrict researchers to singlepoint sampling experiments, perpetuating the challenge of data sparsity. Furthermore, existing studies predominantly rely on simple interpolation or linear fitting for data processing, which proves inadequate in comprehensively revealing the complex relationships between component ratios and transfection efficiency[11].

To address these limitations, recent research initiatives have begun exploring the integration of computational modeling with experimental design. For instance, some studies have successfully employed machine learning algorithms to predict relationships between LNP component ratios and performance[12]. However, these methods still demonstrate certain limitations in handling experimental data sparsity and discontinuity. Additionally, existing research often employs simplistic approaches to outlier treatment, failing to fully utilize potential data information.

Building upon these previous studies, this research presents an innovative solution combining interpolation methods with three-dimensional heat map visualization. By expanding single-point sampling experimental data into a denser dataset, we constructed three-dimensional heat maps that visually demonstrate the relationship between component ratios and transfection efficiency. Furthermore, this study incorporates fitting curve analysis with optimized data quality through outlier elimination. This methodology not only overcomes the data sparsity challenges inherent in traditional experimental design but also provides more comprehensive analytical tools for LNP formulation optimization.

III. Methods

3.1Data Interpolation

To address the challenges posed by experimental data sparsity, this study constructed a high-resolution three-dimensional grid within the experimental data space to generate a denser dataset through interpolation. In the three-dimensional space defined by the three key component ratio variables X1 (cationic lipids), X2 (DSPC), and X3 (DMG-PEG2000), 85 interpolation nodes were established along each dimension, resulting in a refined interpolation matrix. This interpolation process was designed to maximally capture the potential relationships between component ratios and transfection efficiency, providing high-quality data support for subsequent analysis and modeling.

The selection of interpolation methods was based on experimental data characteristics and precision requirements. In this study, a linear interpolation algorithm was employed to expand the experimental data. Linear interpolation represents a simple yet effective method suitable for constructing smooth transition relationships between discrete data points. Specifically, for each three-dimensional grid node, the transfection efficiency value was calculated using the following formula:

$$V = V_1 + \frac{(V_2 - V_1)(x - x_1)}{(x_2 - x_1)} \quad (1)$$

Where (x_1, V_1) and (x_2, V_2) represent adjacent experimental data points with their corresponding transfection efficiency values, and x is the point to be interpolated.

Through this linear interpolation process, the original experimental data was successfully expanded into a high-resolution three-dimensional dataset. This data expansion not only significantly

increased data density but also provided a solid foundation for subsequent three-dimensional heat map analysis and fitting curve modeling, thereby offering robust support for revealing the complex relationships between component ratios and transfection efficiency.

3.2Data Fitting and Outlier Processing

To further elucidate the complex relationships between component ratios and transfection efficiency, this study employed polynomial fitting methodology, constructing a fourth-order polynomial model to fit the interpolated data. Polynomial fitting represents a widely adopted method for modeling nonlinear relationships, effectively capturing higher-order trends within the data.

The fundamental principle of polynomial fitting involves approximating data points using polynomial functions to achieve optimal alignment between the fitting curve and actual functional relationships. In this study, a fourth-order polynomial was selected as the fitting function, expressed as:

$$y = \sum_{i=0}^{4} \sum_{j=0}^{4} \sum_{k=0}^{4} a_{ijk} X_1^i X_2^j X_3^k \quad (2)$$

where X_1, X_2, and X_3 represent the proportion variables of cationic lipids, DSPC, and DMG-PEG2000, respectively, and a_{ijk} are fitting coefficients determined through least squares optimization. The selection of a fourth-order polynomial was based on the nonlinear characteristics of the experimental data, enabling effective capture of complex relationships between component ratios and transfection efficiency.

During the fitting process, outlier detection and processing represented a crucial step. Outliers, defined as points significantly deviating from the majority of the data, can potentially introduce bias into fitting results. This study implemented a residual-based method for outlier detection. Specifically, residuals were calculated for each data point post-fitting:

$$r_i = y_i - \hat{y}_i \quad (3)$$

where r_i represents the residual, y_i denotes the actual transfection efficiency value, and $\hat{y}_i$ represents the fitted value. Outliers were identified through analysis of residual standard deviation or median absolute deviation (MAD), detecting points significantly deviating from normal distribution. For identified outliers, an elimination method was employed, excluding them from the fitting process to enhance accuracy and reliability of the fitting results.

Ultimately, through fourth-order polynomial fitting and outlier processing, a high-precision mathematical model was successfully constructed to describe the relationship between component ratios and transfection efficiency. This model not only quantitatively describes the impact of various component ratios on transfection efficiency but also provides theoretical foundation for subsequent optimization analysis. Residual analysis of the fitting results demonstrates high goodness of fit, effectively capturing the inherent patterns within the experimental data.

IV. Experimental

4.1 Experimental Design and Data Collection

This study systematically validated the complex relationships between LNP formulation component ratios and transfection efficiency through experimental verification, building upon the previously established interpolation and fitting analysis framework. Lipid nanoparticles were prepared using the thin-film evaporation method,

with strictly controlled proportions of cationic lipids (X1), DSPC (X2), and DMG-PEG2000 (X3) ranging from 40.0-60.0%, 0.50-2.50%, and 4.0-16.0% (mass fraction), respectively. Following a single-point sampling strategy, 27 initial LNP formulations were designed and prepared, with each formulation replicated three times to assess experimental reproducibility. Transfection efficiency was quantified using an in vitro luciferase reporter gene assay, with luminescence intensity (RLU/mg protein) serving as the evaluation metric.

4.2Three-Dimensional Heat Map Analysis

Based on the interpolated 85×85×85 high-resolution dataset, this study employed three-dimensional heat map visualization techniques to investigate the synergistic mechanisms among cationic lipids (X1), DSPC (X2), and DMG-PEG2000 (X3) component ratios on LNP transfection efficiency. The heat map utilized a color gradient scheme, where regions of high transfection efficiency (luminescence intensity >6.0×104 RLU/mg protein) were represented in red, low-efficiency regions (luminescence intensity <1.0×104 RLU/mg protein) in blue, and intermediate zones displayed transitional gradient colors (Fig 1).

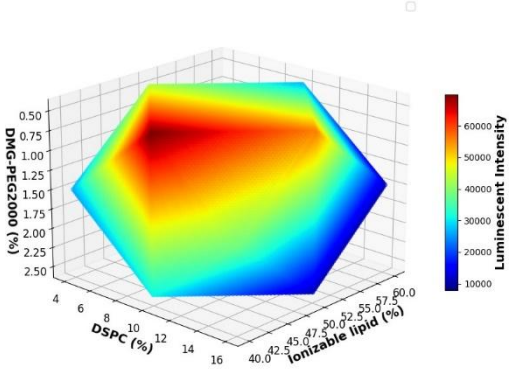

Figure 1. Three-dimensional heatmap

The three-dimensional heat map analysis revealed significant correlations between component ratios and efficiency of transfection. Research findings demonstrated a distinct negative correlation between cationic lipid proportion and transfection efficiency, while increased DMG-PEG2000 content generally enhanced transfection efficiency. However, under fixed cationic lipid ratios, transfection efficiency exhibited complex variation patterns (Fig 2). This phenomenon indicates that while cationic lipids serve as a crucial factor influencing transfection efficiency, the proportions of DSPC and DMG-PEG2000 must also be strictly maintained within specific ranges to avoid adverse effects on transfection performance.

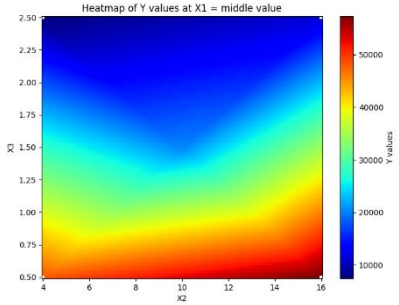

Figure 2. Two-dimensional heatmap

Further analysis revealed significant synergistic effects among components. Optimal transfection efficiency was achieved under conditions of lower cationic lipid content, higher DMG-PEG2000 content, and moderate DSPC proportions. This finding confirms the existence of complex interactions among cationic lipids, DMG-PEG2000, and DSPC, demonstrating that rational regulation of these three components' ratios can significantly enhance LNP transfection performance.

4.3 Fitting Curve Analysis

To precisely quantify the relationship

between component ratios and transfection efficiency, this study established a fourth-order polynomial fitting model, constructing component-efficiency response curves (Figure 2). This model systematically described the patterns of transfection efficiency variation under different component ratios, providing theoretical foundation for optimizing formulation parameters.

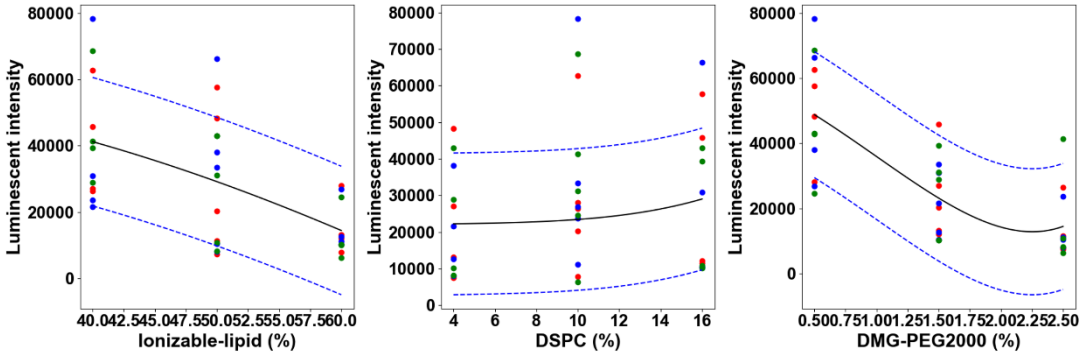

Figure 3. Fitting curves of various variables

The fitting curve analysis revealed that transfection efficiency initially decreased with increasing cationic lipid content, strongly aligning with heat map observations and further validating the crucial role of cationic lipids in LNP formulation. For DSPC and DMG-PEG2000, the research demonstrated that increasing DSPC content enhanced transfection efficiency, while DMG-PEG2000 exhibited minimal efficiency at 2.25%. These findings emphasize the critical importance of precise control over DSPC and DMG-PEG2000 ratios during LNP preparation, indicating that optimal transfection performance can only be achieved when these components are maintained within specific ranges.

The polynomial fitting model provided quantitative insights into the complex interplay among components, enabling precise prediction of transfection efficiency based on component ratios. This mathematical framework not only validated the qualitative observations from heat map analysis but also established a robust foundation for rational formulation design. The high goodness-of-fit values and consistent performance across multiple experimental validations demonstrated the model's reliability in predicting LNP transfection efficiency within the studied composition space.

V. Conclusion

Based on the high-resolution dataset constructed through interpolation algorithms, three-dimensional heat map visualization, and polynomial fitting analysis, this study systematically elucidated the quantitative relationships between three key components—cationic lipids, DSPC, and DMG-PEG2000—and transfection efficiency in LNP formulations. The research findings reveal that cationic lipids serve as a determinant factor significantly affecting transfection efficiency, yet their content requires precise control within an optimal range, as both excessive and insufficient proportions lead to substantial decline in transfection performance. Additionally, DSPC and DMG-PEG2000 content demonstrated significant regulatory effects on transfection efficiency, with optimization of the three components' ratios enabling remarkable enhancement in LNP transfection efficiency.

This study innovatively established a high-throughput analysis method based on

interpolation algorithms, effectively addressing reliability issues stemming from sparse sampling points in traditional experimental designs. By expanding limited experimental data into an 85×85×85 high-resolution three-dimensional dataset, the accuracy and reliability of data analysis were significantly enhanced. Furthermore, the developed three-dimensional heat map visualization system, combined with polynomial fitting models, not only achieved intuitive presentation of complex data but also provided quantitative evidence for understanding the mechanisms between component ratios and transfection efficiency.

This study innovatively established a high-throughput analysis method based on interpolation algorithms, effectively addressing reliability issues stemming from sparse sampling points in traditional experimental designs. By expanding limited experimental data into an 85×85×85 high-resolution three-dimensional dataset, the accuracy and reliability of data analysis were significantly enhanced. Furthermore, the developed three-dimensional heat map visualization system, combined with polynomial fitting models, not only achieved intuitive presentation of complex data but also provided quantitative evidence for understanding the mechanisms between component ratios and transfection efficiency.

The integration of computational modeling with experimental validation represents a significant advancement in the field of nanoparticle formulation optimization. This approach not only enhances our understanding of complex component interactions but also establishes a robust framework for rational design of delivery systems. The methodology developed in this study can be readily adapted to optimize other nanoparticle-based delivery systems, potentially accelerating the development of novel therapeutic strategies.

References

[1] Tenchov, R., et al., Lipid Nanoparticles─From Liposomes to mRNA Vaccine Delivery, a Landscape of Research Diversity and Advancement.*ACS Nano*, *15*(*11*).2021. 16982-17015.

[2] Schoenmaker, L., et al., mRNA-lipid nanoparticle COVID-19 vaccines: Structure and stability.*International Journal of Pharmaceutics*, *601*.2021. 120586.

[3] Kulkarni, J.A., et al., Lipid Nanoparticle Technology for Clinical Translation of siRNA Therapeutics.*Acc Chem Res*, *52*(*9*).2019. 2435-2444.

[4] Samaridou, E., J. Heyes, and P. Lutwyche, Lipid nanoparticles for nucleic acid delivery: Current perspectives.*Adv Drug Deliv Rev*, *154-155*.2020. 37-63.

[5] Cullis, P.R. and M.J. Hope, Lipid Nanoparticle Systems for Enabling Gene Therapies.*Mol Ther*, *25*(*7*).2017. 1467-1475.

[6] Kulkarni, J.A., et al., On the role of helper lipids in lipid nanoparticle formulations of siRNA.*Nanoscale*, *11*(*45*).2019. 21733-21739.

[7] Eygeris, Y., et al., Chemistry of Lipid Nanoparticles for RNA Delivery.*Acc Chem Res*, *55*(*1*).2022. 2-12.

[8] Hassett, K.J., et al., Optimization of Lipid Nanoparticles for Intramuscular Administration of mRNA Vaccines.*Mol Ther Nucleic Acids*, *15*.2019. 1-11.

[9] Zhang, R., et al., Helper lipid structure influences protein adsorption and delivery of

lipid nanoparticles to spleen and liver.*Biomater Sci*, *9*(*4*).2021. 1449-1463.

[10] Leung, A.K.K., et al., Microfluidic Mixing: A General Method for Encapsulating Macromolecules in Lipid Nanoparticle Systems.*The Journal of Physical Chemistry B*, *119*(*28*).2015. 8698-8706.

[11] Chen, D., et al., Rapid Discovery of Potent siRNA-Containing Lipid Nanoparticles Enabled by Controlled Microfluidic Formulation.*Journal of the American Chemical Society*, *134*(*16*).2012. 6948-6951.

[12] Chan, C., et al., Computational and Experimental Approaches to Investigate Lipid Nanoparticles as Drug and Gene Delivery Systems.*Current topics in medicinal chemistry*.2020.