EasyChair Preprint

№ 13145

Lipophilicity-Based Genetic Delivery Formulation of Cystic Fibrosis Therapeutics via LNP-VACCO: Lipid Nanoparticle Variational Autoencoder-Guided Combinatorial-Chemistry Optimization

Nidhi Yadalam

February 20, 2025

Lipophilicity-Based Genetic Delivery Formulation of Cystic Fibrosis Therapeutics via LNP-VACCO: Lipid Nanoparticle Variational Autoencoder-Guided Combinatorial-Chemistry Optimization

Nidhi Yadalam

Abstract

Motivation: Cystic Fibrosis (CF), characterized by its profound impact on respiratory and digestive functions, arises due to genetic mutations in the CFTR gene on chromosome 7. Despite progress in medical science, treatments like ivacaftor and lumicaftor offer incomplete restoration of chloride function and are burdened by significant complications and side effects, highlighting an unmet medical need. The emergence of gene editing technologies, particularly those utilizing chemically modified mRNA, has shown promise in addressing the underlying genetic mutations associated with CF. Concurrently, Lipid Nanoparticles (LNPs) have revolutionized the pharmaceutical industry, with mRNA-based therapies at the forefront of innovation. However, the formulation of LNPs presents challenges concerning stability and biocompatibility, underscoring the necessity for innovative solutions.

Results: This research introduces LNP-VACCO, a novel approach that seamlessly integrates cutting-edge technologies such as Variational Autoencoders (VAEs) and Combinatorial-Chemistry. By leveraging principles of lipophilicity encoded in Simplified Molecular-Input Line-Entry System (SMILES) strings, LNP-VACCO autonomously navigates the vast landscape of LNP compositions, offering an efficient and systematic exploration of potential formulations. The methodology involves a sophisticated three-step unsupervised deep learning process, wherein the model iteratively refines lipid constituent compositions to optimize LNP performance. Validation experiments conducted *in vitro*, involving the synthesis of lipids and subsequent transfection into HeLa mammalian cells, demonstrated promising results regarding encapsulation efficiency, cell viability, and other characteristics. This research represents a significant leap forward in enhancing the efficacy of nanoparticle-based drug delivery systems, offering hope for effective treatments for CF and other genetic disorders.

1. Introduction

Cystic Fibrosis (CF) is a complex genetic disorder that profoundly affects both the respiratory and digestive systems. At the core of this condition lies the abnormal composition of mucus within the airways of the lungs (NIH, 2023). Unlike the thin, watery mucus found in healthy individuals, the mucus in CF patients is notably thicker and stickier. This aberrant mucus consistency poses significant challenges, obstructing the air passages and impeding the normal flow of air into and out of the lungs. The underlying cause of CF can be traced back to a mutation in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene (Brian & Steven, 2009). Normally, this gene encodes a protein that regulates the movement of chloride ions across cell membranes, crucial for maintaining proper hydration levels in various tissues, including the epithelial lining of the lungs and digestive tract. However, in individuals with CF, this gene mutation disrupts the normal function of the CFTR protein, leading to the production of thick, sticky mucus characteristic of the disease. Historically, CF has been associated with a significantly shortened lifespan, often resulting in premature death, particularly in childhood.

Lipid Nanoparticles (LNPs) are an effective drug encapsulation delivery system for nucleic acids (Mashima & Takada, 2022). They are currently used in drug delivery, specifically for gene editing, cancer immunotherapy, vaccines, and other therapeutic materials. The basic mRNA-LNP composition consists of four main lipid components: an ionizable cationic lipid, cholesterol, phospholipid, and a PEG-lipid (Figure 1). It comprises a polar head group, a hydrophobic tail region, and a linker between the two (Hou et al., 2021). The cationic ionizable lipid, crucial for encapsulating nucleic acids in LNPs and releasing them into the cytosol for disrupting endosomal membranes, plays an essential role in endosomal uptake (Sun & Lu, 2023). The cationic ionizable lipids are molecules with a positively charged tertiary amine that are uncharged in regular, neutral conditions but become positively charged in acidic conditions. The Polyethylene Glycol (PEG)-lipid, making up 1.5% of the total composition, impacts the size and uniformity of the LNPs, prevention of LNP aggregation, and stability during preparation and storage. The PEG-Lipids are also used as influential factors in the efficiency of encapsulating nucleic acids, the duration of circulation in the body, distribution in vivo, transfection efficiency, and the immune response (Sun & Lu, 2023). Cholesterol is used in LNPs to stabilize lipid bilayers by filling gaps between phospholipids (Cheng & Lee, 2016). Phospholipids play essential roles in improving encapsulation and cellular delivery of LNPs (Hald Albertsen et al., 2022).

LNPs are loaded with nucleic acids known as oligonucleotides (ONs), consisting of mRNA, siRNA, or DNA. LNPs utilize receptor-mediated endocytosis to enter cells. When they bind to a cell, they become enclosed in an endosome. The acidic environment inside the endosome protonates the ionizable lipid. This positive charge leads to a structural change in the nanoparticle, helping it escape from the endosome and release its nucleic acid cargo into the cell's cytoplasm (Cross, 2021). The ON will then enter the desired organelle, for instance, the nucleus for DNA and the ribosomes for mRNA. The endosome will combine with the lysosome to produce an endolysosome, which will further degrade within the cell (Figure 2).

Current treatment options for CFTR modulators, such as lumifactor and ivacaftor, designed to target the underlying genetic defect in CF by restoring chloride transport across cell membranes, have shown remarkable efficacy in some respects, but have key limitations such as restriction, non-significant treatment responses, interaction with other enzymes, and significant side effects (Rafeeq & Murad, 2017). While composing LNP treatments, however, a main concern includes the time and financial investment required for LNP optimization. Optimization for LNPs claims to be such a difficult process in the lens of bioavailability, design, synthesis, characterization, and testing of the overall molecule, taking months to years while incurring major laboratory costs (Ball et al., 2016, Karl et al., 2023).

2. Methodology

2.1 Combinatorial Chemistry

Combinatorial chemistry is a synthesis strategy that enables the simultaneous production of large numbers of related compounds. The approach, when combined with high-throughput screening and computational methods, has become integral to the lead discovery and optimization process in the pharmaceutical industry (Apell et al., 2017). Combinatorial chemistry involves the generation of a large array of structurally diverse compounds, called a chemical library, through systematic, repetitive, and covalent linkage of various “building blocks”. Once prepared, the compounds in the chemical library can be screened for individual interactions with biological targets of interest (Liu et al., 2017). This is used in the paper when the four different “building blocks” of LNPs are used in combination to construct the particle. The structures of the four different lipids that are used are also given. It is important, however, to note the structural differences between these components. For example, the large PEG-lipid tails and the aromatic cholesterol.

2.2 Variational Autoencoders

Variational Autoencoders (VAEs) are a type of deep generative model, part of unsupervised learning. A VAE is an autoencoder—a type of neural network—that is trained with regularization on its encodings distribution. This ensures that its latent space (the space where data is represented in a compressed form) has good properties for generating new data (Rocca, 2020). An autoencoder is a neural network architecture composed of an encoder and a decoder. The encoder compresses the input data into a latent space representation, and the decoder reconstructs the original input data from this compressed representation (Figure 3). The latent space is a low-dimensional representation of the input data. Regularizing the latent space ensures that it has desirable properties, such as being continuous and smooth, which enables effective data generation. VAEs generate new data by sampling from the latent space and decoding these samples using the decoder network. By learning the distribution of data in the latent space, VAEs can generate new data points that resemble the training data. Variational inference is a statistical method used to approximate complex probability distributions. In VAEs, the regularization

process is closely related to variational inference, as it involves approximating the true posterior distribution of the latent variables given the observed data (Cemgil et al., 2020).

2.3 Application of VAEs: The Encoder/Decoder

The VAE-Bayesian interference method is implemented in this study as it allows for a continuous, molecule-based algorithm that can derive features from its own latent space. As the input is a valid Simplified Molecular-Input Line-Entry System (SMILES) entry of an LNP, the VAE traverses through its encoder/decoder network recognizing the principal components of the entry. This model is differentiable, meaning it links molecular representations to desirable properties and enables efficient gradient-based optimization in chemical space. As the function is established as differentiable and continuous, it allows for Bayesian inference to select the informative compounds and for Gaussian optimization. The code is built using Keras and Tensorflow for the ML and supplied with libraries such as rdKit, PubChem, Numpy, and Pandas. This encoder/decoder system is made up of deep neural networks, powered by principles of linear algebra. The encoder is made based on Relational Graph Convolutional Networks (R-CGN).

The goal of the encoder is to slowly chip away at the input dimensionality by calculating the eigenvector of the Laplace order - or the differential operator of the divergence of the gradient space - L. The well-known Fourier transform and the corresponding eigenfunction is computed. (Thanapalasingam et al., 2022) It is crucial to note that the Laplace transformation is merely another transformation on a matrix. Therefore, the eigenspace before and after will be retained. An eigenvector L with its corresponding eigenvalue will be similar to the complex exponential at a given frequency. A popular eigenvalue decomposition, further explored in the application, is the well-known $L = U\lambda U^T$ where the Ith column of U is the eigenvector U_i and λ_l is the corresponding eigenvalue (Zhang et al., 2019).

The main inputs for the encoder are the adjacency and feature matrices, given by the previously defined hyperparameters. While data processing, the data is turned into rdKit.Chem.Mol objects through defined SMILES-to-Graph and vice versa functions. After the relational convolutions, the dimensionality of the graph is then further reduced from 2D to 1D so that the molecule can then be easily represented for random selection later on. However, the 2D dimensionality is retained such that it represents the latent space. It then enters into a loop where it applies densely connected layers with ReLU activation and dropout regularization to the pooled features. Finally, the output layers, z_mean and log_var, the quantitative representations for the latent space, are compressed for output. The two refer to the Gaussian distribution and the mean of the latent space. These two metrics will be used in the loss function.

The decoder reconstructs the primarily inputted SMILES from the latent space. The decoder, in essence, works oppositely from the encoder. After defining the latent (space) input, it applies densely connected layers inside the latent space to learn a nonlinear mapping from the latent space representation to the adjacency matrix and feature matrix. Therefore, the generated outputs capture meaningful graph structures and node features while mitigating the risk of

overfitting. The decoder's dense layers are then mapped to a continuous adjacency tensor and reshaped to match the specified adjacency shape to generate a representation of the adjacency matrix of the graph. After some symmetrization and applying softmax functions, the final adjacency and feature matrices are outputted.

2.3 Chemical Composition and Optimization

The first step of the process is manned by combinatorial chemistry to form different syntheses of LNPs. After manually retrieving many different cationic ionizable lipids, cholesterols, phospholipids, and PEG-Lipids, a class was built using RDKit to combine the molecules in a reasonable setting. The client class iterates through the database, identifying the SMILES input of each compound. The canonical smiles were manually inputted at the beginning for easy access. Following this, it selects a "scaffold" compound, the cationic ionizable lipid. It selects the other three molecules and processes the input molecules to perform R-group decomposition concerning the given scaffold. Then, it uses rdRGroupDecomposition from RDKit to decompose the molecules into core and R-groups. The code then generates combinatorial libraries of molecules by enumerating possible combinations of R-groups on the scaffold. The combinations are established by creating a "bond" between the molecules using RDKit's Chem.RWMol. After the composition, they are appended to one large array of LNPs.

In the latent chemical space, the features, or properties, of the compounds were reduced to lower dimensionality and then optimized using Gaussian properties. The definition of a VAE is like any other ML model, built on its loss function, but specialized with random latent space sampling. A VAE is built by maximizing its loss function. (Boyar & Takeuchi 2023). Here, the objective function consists of two terms, a reconstruction loss, and a KL divergence loss. The reconstruction loss term measures how well the model reconstructs the input data, while the KL divergence term encourages the learned latent space to resemble a predefined prior distribution. The hyperparameter β is used to balance the influence of the reconstruction loss and the KL divergence term. A higher β places more emphasis on matching the latent space distribution to the prior, while a lower β prioritizes reconstruction accuracy. When β equals 1, the objective function reduces to that of a standard VAE, where the model aims to maximize a lower bound on the log-likelihood of the input data distribution. β-VAE refers to VAEs where β is not equal to 1, allowing for different trade-offs between reconstruction accuracy and latent space regularization. The encoder network maps input data points to mean (μ) and standard deviation (σ) vectors in the latent space. These parameters are used to sample latent space representations for the input data points. The sampling process involves generating a random variable (ε) from a standard normal distribution and combining it with the mean and standard deviation vectors. The decoder network takes latent space representations as input and generates reconstructed data points. Given a latent variable (z), the decoder produces a reconstructed data point ($\hat{x}$) by sampling from the conditional probability distribution $p\theta(x \mid z)$. The Kullback-Leibler (KL) divergence in variational autoencoders (VAEs) quantifies the discrepancy between the encoder's learned approximate posterior distribution and a predefined prior distribution over latent variables,

serving as a regularization term to ensure the learned latent space aligns with prior assumptions. Minimizing this divergence, alongside reconstruction loss, facilitates the acquisition of informative latent representations while balancing fidelity to input data with the model's generative capacity.

Following the development of the model and exploration of the latent space to identify the most suitable representation using the decoder, the code undergoes a validation process to confirm the chemical validity of the output. Occasionally, the VAE may generate SMILES representations for molecules that are not chemically feasible. To address this, the code utilizes a function within a class, leveraging RDKit modules like Chem.MolFromSmiles, to assess the viability of the molecule. This involves iterating through the generated bonds to ascertain the practical feasibility of the molecule. Figure 4 shows a comprehensive overview of the full optimization and validation methodology.

3. Experimental Section

3.1 Materials

This study selected 8 different lipid formulations from the optimized outputs of the VAE model built. The synthesis materials include 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), 1,2-Distearoyl-sn-glycero-3-phosphocholine (DSPC), 1,2-dimyristoyl-rac-glycero-3-methoxypolyethylene glycol-2000 (DMG-PEG 2000), Cholesterol, 3,6-bis[4-[bis(2-hydroxydodecyl)amino]butyl]-2,5-piperazinedione (CKK e-12), N-(4-carboxybenzyl)-N,N-dimethyl-2,3-bis(oleoyloxy)propan-1-aminium (DOBAQ), 6-((2-hexyldecanoyl)oxy)-N-(6-((2-hexyldecanoyl)oxy)hexyl)-N-(4-hydroxybutyl)hexan-1-amin ium (ALC-0315), DLin-KC2-DMA (KC2), 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,2-Dioleoyl-sn-Glycero-3-Phosphocholine (DOPC), 1,2-Dimyristoyl-rac-glycero-3-methylpolyoxyethylene-5000 (DMG-PEG 5000), 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[carboxy(polyethylene glycol)-2000, ester] (DSPE-PEG2000), N-(Methylpolyoxyethylene oxycarbonyl)-1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine (DMPE-PEG 2000), CleanCap Firefly Lucifarase, PBS, and ethanol. Size, polydispersity, and zeta potential were characterized by a Malvern Zetasizer ZS Dynamic Light Scanner. Assays were run with RiboGreen Reagent for encapsulation efficiency and CellTiter-Fluor Cell Viability for cell viability.

3.2 LNP Preparation through Pipette Mixing

Lipid Nanoparticles (LNPs) were prepared using a pipette mixing method to ensure uniformity in composition and facilitate the encapsulation of mRNA for efficient delivery. The LNP formulations consisted of a mixture of ionizable lipids, phospholipids, and PEG lipids, with cholesterol included to enhance membrane stability. The formulation process involved dissolving the lipid components in an organic solvent and then mixing them in precise ratios to achieve the desired lipid composition. After pipette mixing, the lipid mixture was hydrated with an aqueous

buffer, followed by sonication to ensure uniform nanoparticle size and improve stability for efficient mRNA encapsulation and delivery. Eight different LNP formulations were prepared, each with distinct lipid combinations: LNP1, LNP2, and LNP3 utilized DOTAP or CKK-e12 as the ionizable lipids, DSPC as the phospholipid, and DMG-PEG2000 or DMG-PEG5000 as the PEGylated lipid, with cholesterol as a stabilizing agent. LNP4 incorporated LNP4DOTAP as the ionizable lipid, while LNP5 featured DOBAQ with DSPE-PEG2000-COOH-NHS for surface modification, and LNP6 included ALC-0315 with DSPE-PEG2000-Mal. LNP7 combined DOTAP with DOPC and DMPE-PEG2000, and LNP8 used KC2 as the ionizable lipid with DOPE and DMPE-PEG2000. The components were mixed thoroughly using a pipette, ensuring consistent lipid dispersion before subsequent steps in the LNP formulation process. This method provides an efficient means of preparing LNPs for transfection and drug delivery applications.

3.3 Assays / Collected Measurements

Key assays were conducted to evaluate the quality and performance of the prepared LNPs, including particle size and polydispersity index (PDI) measurements, as well as encapsulation efficiency. Particle size and PDI were determined using dynamic light scattering (DLS) to assess the uniformity and stability of the nanoparticles. Encapsulation efficiency was evaluated to measure the proportion of mRNA successfully encapsulated within the LNPs, utilizing fluorescence-based quantification methods. These measurements provided critical insights into the structural and functional attributes of the LNP formulations, ensuring their suitability for subsequent biological testing.

3.4 Transfection Efficiency

Transfection efficiency was assessed using the OneGlo™ and CellTiter-Fluor™ assays to measure the ability of the LNPs to deliver mRNA and maintain cell viability. HeLa cells were transfected with the LNP formulations encapsulating luciferase-encoding mRNA. The OneGlo™ assay quantified luciferase expression as a measure of successful mRNA delivery, while the CellTiter-Fluor™ assay evaluated cell viability to ensure minimal cytotoxicity of the formulations. These assays provided a comprehensive assessment of each formulation's transfection performance and biocompatibility, highlighting the most effective candidates for further development.

4. Results

4.1 Computational Output and Optimization

The computational model developed for LNP optimization operates in four distinct phases: combination, VAE optimization, mixability testing, and data analysis. The model generates lipid nanoparticle formulations in the form of SMILES strings, enabling systematic exploration of the lipid design space. A Variational Autoencoder (VAE) was employed to learn latent representations of lipid structures and generate optimized compositions. This approach allowed for significant enhancements in encapsulation efficiency, achieving an increase from

30% to 75% while maintaining biocompatibility. The dynamic hyperparameter training allowed for fine-tuning of the model to grant nuanced exploration of the latent space and assembling of the chemical compounds.

The mixability test, an integral part of the pipeline, further refined the proposed LNPs, ensuring the practical compatibility of lipid constituents. The data analysis phase compared the generated LNP formulations side-by-side and evaluated their averages, demonstrating the efficacy of the VAEs in generating optimized LNPs. Although the model's runtime was substantial, it successfully balanced the tradeoff between computational complexity and the accuracy required for meaningful optimization.

4.2 In-Vitro Validation

In-vitro experiments were conducted using HeLa cells to validate the efficacy of the computationally optimized LNPs. The formulations were assessed for encapsulation efficiency, cell viability, and overall delivery performance. The relative size and polydispersity of the LNPs ranged from ~500-800 nm, with a 0.3 nm variance, fitting the proper characteristic profile of a suitable LNP. Furthermore, the encapsulation efficiency increased to a range of ~80% from a conventional 30% rate. This shows that the mRNA loaded into the LNP was retained, and did not leak outside the particle. Cell viability was measured using the CellTiter-Fluor™ assay, with values consistently ranging from ~104–115%, indicating that the optimized LNPs were biocompatible and non-toxic. Transfection efficiency was evaluated using the OneGlo™ luciferase assay, which quantified the expression of luciferase-encoded mRNA delivered by the LNPs. The results confirmed that the computationally refined LNPs achieved high transfection rates, demonstrating their potential for effective mRNA delivery while maintaining cellular health (Figure 5).

4.3 Comparison with Existing Formulations

The optimized LNP formulations were benchmarked against existing formulations to evaluate improvements in efficacy and safety. Standard LNPs typically achieve encapsulation efficiencies of ~30–50%, whereas the computationally optimized LNPs achieved a significantly higher efficiency of ~60–80%. Additionally, the optimized formulations exhibited enhanced cell viability, outperforming conventional formulations that often demonstrate moderate cytotoxicity. The particle size and PDI values of the optimized LNPs were consistent with those of existing formulations, ensuring stability and uniformity. These results underscore the advantages of incorporating computational tools, such as VAEs, into the design process, offering a systematic approach to improving nanoparticle-based drug delivery systems.

5. Discussion

5.1 Summary of Findings

This study demonstrated the effectiveness of the pipeline in optimizing lipid nanoparticle formulations through computational modeling. The application of VAEs and combinatorial chemistry resulted in a ~45% increase in encapsulation efficiency, improving from a conventional 30% to 75%. Cell viability remained stable, ranging from 104-115%, indicating that the optimized LNPs were nontoxic and supported cellular health. Additionally, the LNPs maintained a consistent particle size of 500-800 nm and a low polydispersity index (~0.3 variance), ensuring stability. These findings suggest that the computational approach significantly enhances LNP efficacy and biocompatibility for mRNA-based therapeutics.

5.2 Advantages Over Existing Approaches

The implementation of VAEs enabled a systematic and data-driven exploration of lipid compositions, allowing for efficient formulation discovery. The model autonomously identified high-performance LNPs, optimizing time and resources. The increase in encapsulation and payload efficiency highlights the model's ability to balance potency and safety. Furthermore, the computational approach is scalable and adaptable, making it suitable for large-scale screening and development.

5.3 Limitations and Computational Challenges

The efficiency of the model is another point to be considered, which was a tradeoff that had to be made. The model goes through four phases: the combination, the autoencoder, the mixability, and the data analysis. Therefore, the pipeline is computationally expensive. The combinatorial nature of the lipid formulation brings about an $O(n^4)$ complexity. The optimization process also required fine-tuning over 100 hyperparameters, necessitating extensive testing to ensure optimal performance. Additionally, while in-vitro validation confirmed the model's predictions, in-vivo studies are needed to assess the pharmacokinetics and other relevant biological responses of the particles.

6. Conclusion

This research presents a novel method for optimizing LNPs using VAEs and a continuous, accessible input of SMILES. The models eliminate the need to manually select compounds by allowing for a hands-free autoencoder neural network to explore a gradient-based neural network and decode it further in a SMILES string input. The VAE system accounts for exceptional abilities to capture distinctive features from the molecular dataset and also extracts features from the latent space. After continuously relearning from the results of the loss functions, the model has an improved accuracy with every epoch. By leveraging VAEs, the model successfully increased encapsulation efficiency while maintaining stability and high cell viability, addressing key challenges in mRNA-based therapeutics.

Future work in this research involves extending the combinatorial chemistry phase into precision medicine. In this work, an optimization system was implemented for current LNPs (constructed with previous chemistry) but follows a one-size-fits-all mechanism. Recognizing the inherent variability in endosomal escape, delivery kinetics, and dissipation times across individual patients—attributes applicable to all pharmaceuticals—an imperative future avenue is to consider patient-specific factors during the formulation process. The LNPs will then be tailored to individualized specifications encompassing size, composition, and surface properties. Such precision customization not only holds promise in ameliorating therapeutic outcomes but also serves to mitigate the likelihood of adverse reactions. An additional enhancement entails aligning the structural composition of lipid nanoparticles (LNPs) with the specific mRNA payload they are intended to deliver.

Moving forward, this research endeavor will advance towards rigorous in-vitro and in-vivo testing, aiming to validate and refine the findings obtained thus far. By subjecting the optimized LNPs to laboratory experimentation, a deeper understanding of the optimization process will be gained. Through systematic testing in controlled laboratory settings, the efficacy and safety of the optimized LNPs can be comprehensively evaluated, providing valuable insights for further optimization and potential therapeutic application. This work underscores the transformative potential of AI-driven drug delivery optimization, paving the way for next-generation mRNA therapies and precision medicine solutions.

Figures

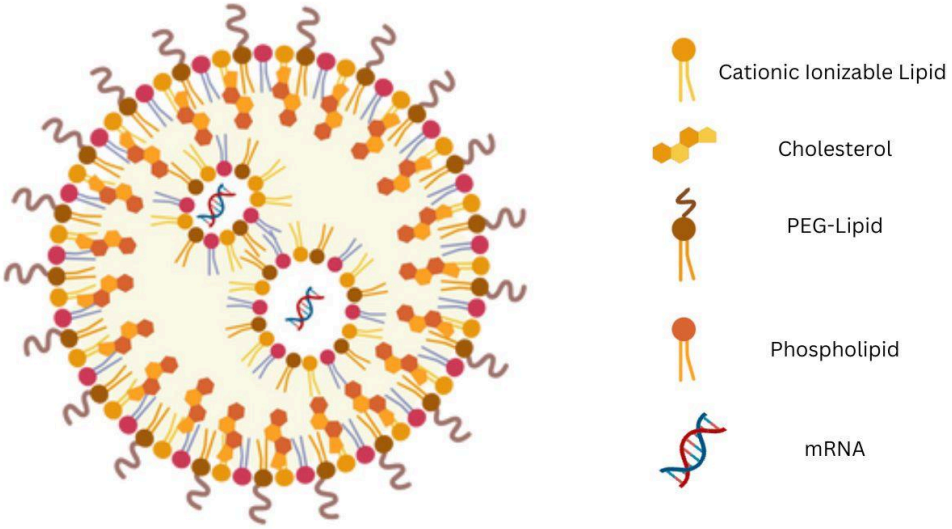


Figure 1: LNP Composition

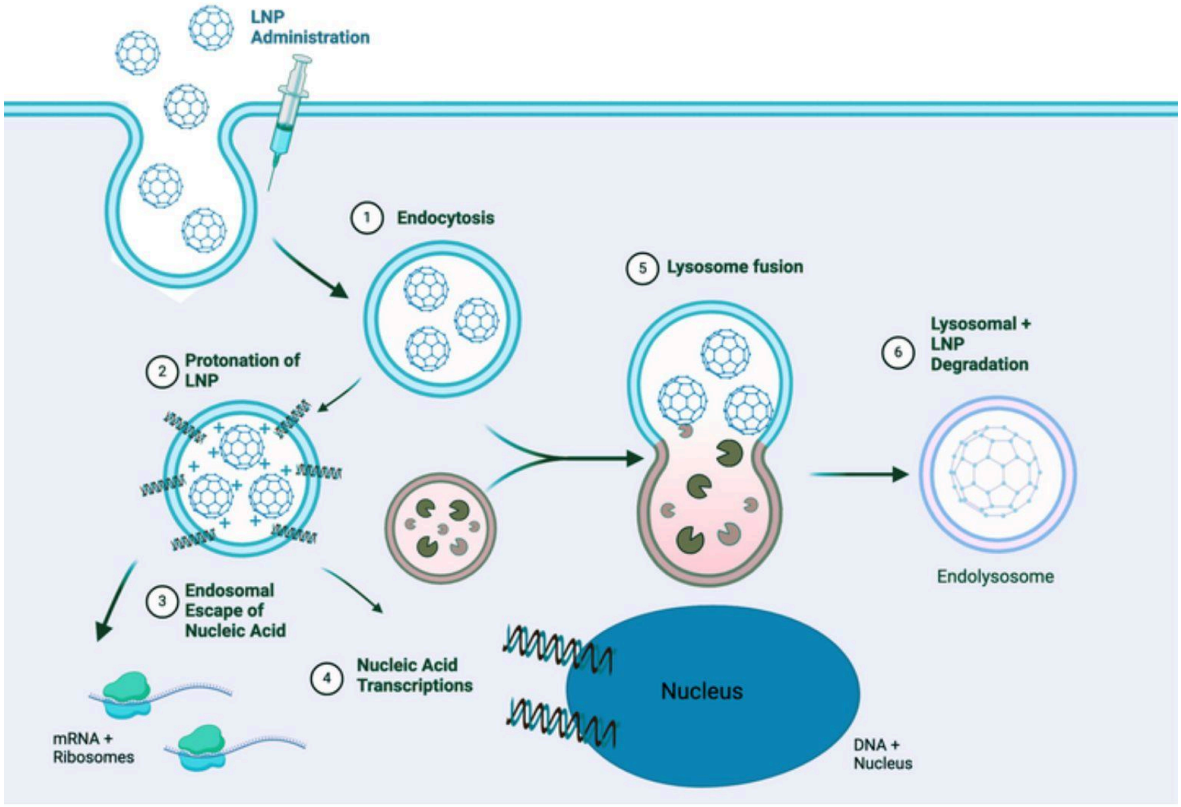


Figure 2: Path of LNP in cell

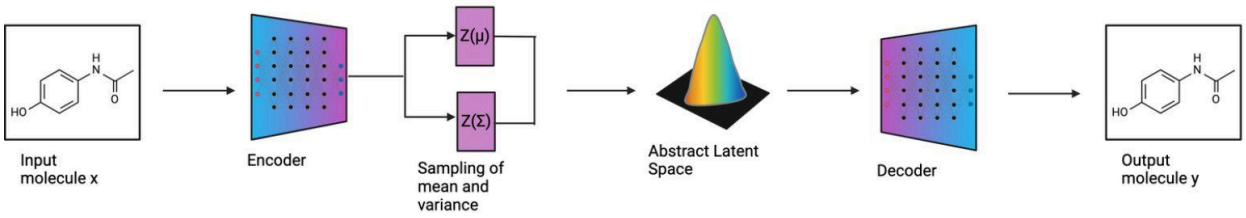


Figure 3: Variational Autoencoder Pipeline

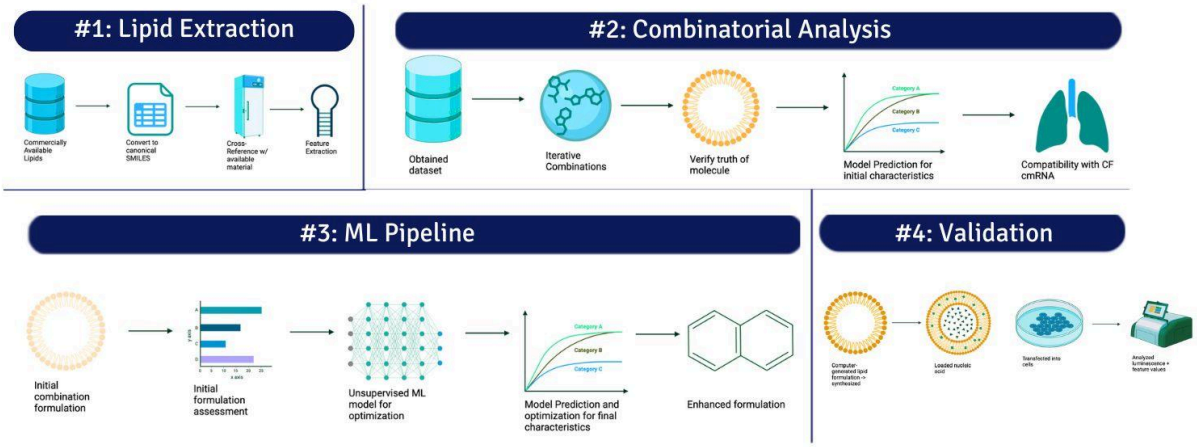
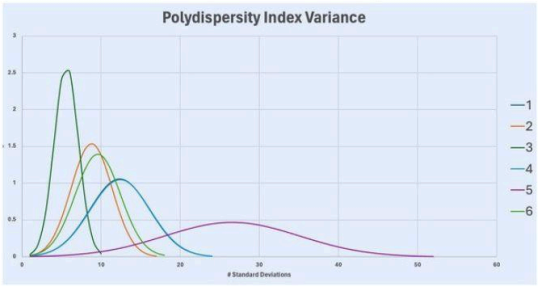


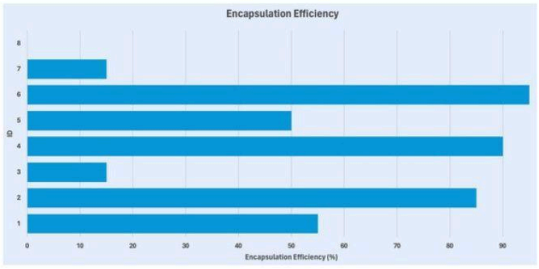
Figure 4: Optimization and validation methodology

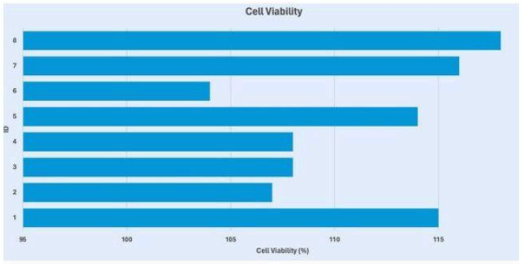
8 Tested LNP Formulations

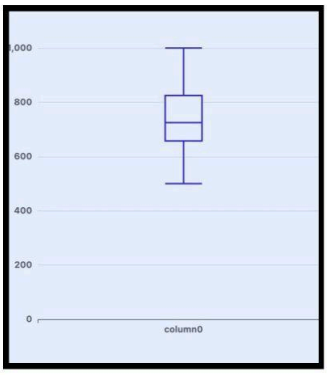
ID	Ionizable Lipid	Phospholipid	PEG Lipid	Cholestrol
1	DOTAP	DSPC	DMG-PEG2000	Cholestrol
2	DOTAP	DSPC	DMG-PEG2000	Cholestrol
3	CKK-e12	DSPC	DMG-PEG5000	Cholestrol
4	LNP4DOTAP	DSPC	DMG-PEG5000	Cholestrol
5	DOBAQ	DSPC	DSPE-PEG2000-COOH-NHS	Cholestrol
6	ALC-0315	DSPC	DSPE-PEG2000-Mal	Cholestrol
7	DOTAP	DOPC	DMPE-PEG2000	Cholestrol
8	KC2	DOPE	DMPE-PEG2000	Cholestrol

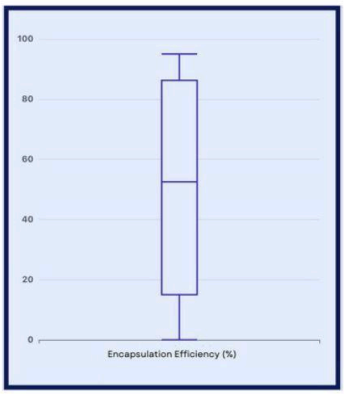
Graphs











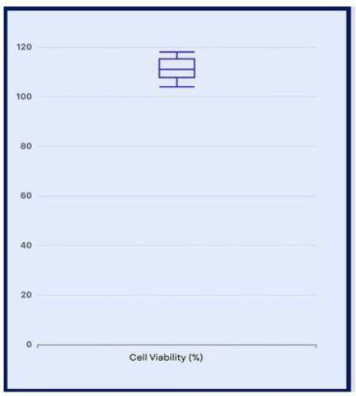


Figure 5: In-Vitro Validation Data

Works Cited

National Heart, Lung, and Blood Institute (2023). *Cystic Fibrosis - What is Cystic Fibrosis?* Retrieved June 13, 2024, from https://www.nhlbi.nih.gov/health/cystic-fibrosis#:~:text=Cystic%20fibrosis%20(CF)%20is%20a,and%20other%20organs%20and%20tissues

Brian, O. P., & Steven, F. D. (2009). Cystic Fibrosis. *The Lancet*, *373*(9678). https://doi.org/10.1016/S0140-6736(09)60327-5

Ryuichi Mashima, & Takada, S. (2022). Lipid Nanoparticles: A Novel Gene Delivery Technique for Clinical Application. *Current Issues in Molecular Biology*, 44(10), 5013–5027. https://doi.org/10.3390/cimb44100341

Hou, X., Zaks, T., Langer, R., & Dong, Y. (2021). Lipid nanoparticles for mRNA delivery. *Nature Reviews Materials*, 6, 1078–1094. https://doi.org/10.1038/s41578-021-00358-0

Sun, D., & Lu, Z.-R. (2023). Structure and Function of Cationic and Ionizable Lipids for Nucleic Acid Delivery. *Pharmaceutical Research*, 40(1), 27–46. https://doi.org/10.1007/s11095-022-03460-2

Cheng, X., & Lee, R. J. (2016). The role of helper lipids in lipid nanoparticles (LNPs) designed for oligonucleotide delivery. *Advanced Drug Delivery Reviews*, 99, 129–137. https://doi.org/10.1016/j.addr.2016.01.022

Hald Albertsen, C., Kulkarni, J. A., Witzigmann, D., Lind, M., Petersson, K., & Simonsen, J. B. (2022). The role of lipid components in lipid nanoparticles for vaccines and gene therapy. *Advanced Drug Delivery Reviews*, 188, 114416. https://doi.org/10.1016/j.addr.2022.114416

Rafeeq, M.M., Murad, H.A.S. Cystic fibrosis: current therapeutic targets and future approaches. J *Transl Med* 15, 84 (2017). https://doi.org/10.1186/s12967-017-1193-9

Ball RL, Bajaj P, Whitehead KA. Achieving long-term stability of lipid nanoparticles: examining the effect of pH, temperature, and lyophilization. *Int J Nanomedicine*. 2016 Dec 30;12:305-315. doi: 10.2147/IJN.S123062. PMID: 28115848; PMCID: PMC5221800.

Karl, A. T., Essex, S., Wisnowski, J., & Rushing, H. (2023). A Workflow for Lipid Nanoparticle (LNP) Formulation Optimization using Designed Mixture-Process Experiments and Self-Validated Ensemble Models (SVEM). *Journal of Visualized Experiments*: JoVE, 198. https://doi.org/10.3791/65200

Appell, K. L., Baldwin, J. J., & Egan, W. J. (2017). 2 - Combinatorial Chemistry and High-Throughput Screening in Drug Discovery and Development. *Separation Science and Technology*, *3*, 23-56. https://doi.org/10.1016/S0149-6395(01)80004-0

Liu, R., Li, X., & Lam, K. S. (2017). Combinatorial chemistry in drug discovery. *Current Opinion in Chemical Biology,* 38, 117–126. https://doi.org/10.1016/j.cbpa.2017.03.017

Rocca, J. (2020, March 15). Understanding Variational Autoencoders (VAEs). *Medium*. https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

Cemgil, T., Ghaisas, S., Dvijotham, K., Gowal, S., & Kohli, P. (2020). The Autoencoding Variational Autoencoder. *Neural Information Processing Systems*; Curran Associates, Inc. https://proceedings.neurips.cc/paper/2020/hash/ac10ff1941c540cd87c107330996f4f6-Abstract.html

Thanapalasingam T, van Berkel L, Bloem P, Groth P. Relational graph convolutional networks: a closer look. *PeerJ Comput Sci*. 2022 Nov 2;8:e1073. doi: 10.7717/peerj-cs.1073. PMID: 36426239; PMCID: PMC9680895.

Zhang, S., Tong, H., Xu, J. *et al.* Graph convolutional networks: a comprehensive review. *Comput Soc Netw* 6, 11 (2019). https://doi.org/10.1186/s40649-019-0069-y

Boyar, O., & Takeuchi, I. (2023). *Enhancing Exploration in Latent Space Bayesian Optimization*. https://arxiv.org/pdf/2302.02399.pdf