EasyChair Preprint

№ 13145

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

Nidhi Yadalam

April 30, 2024

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

Nidhi Yadalam

Table of Contents:

Abstract:

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 being at the forefront of innovation. However, the formulation of LNPs presents challenges concerning stability and biocompatibility, underscoring the necessity for innovative solutions. In response to these challenges, 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 to simulate CF conditions, demonstrated promising results in terms of encapsulation efficiency and cell viability. 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: Background Information

1.1 Cystic Fibrosis Background

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 (1). 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. Moreover, the viscous nature of the mucus creates an ideal environment for the retention of pathogens, making CF individuals more susceptible to recurrent respiratory infections.

The underlying cause of CF can be traced back to a mutation in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene (2). 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.

1.2 LNP Introduction

Lipid Nanoparticles (LNPs) are an effective drug encapsulation delivery system for nucleic acids (1). They are currently used in drug delivery, specifically for gene editing, cancer immunotherapy, vaccines, and other therapeutic materials. This relatively new technology was recently introduced in 1990 as an alternative to synthetic drug carriers such as polymeric nanoparticles and polymers. However, the first commercial usage was in 2018, for the siRNA therapy *Onpattro,* into the cytoplasm of hepatocytes (2). The siRNA-LNP delivery system was then optimized for encapsulation of mRNA. The most notable usage of lipid nanoparticles transfected with mRNA is in the mRNA-COVID-19 vaccine by the pharmaceutical companies Pfizer-BioNTech and Moderna.

1.3 LNP Components

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 (3). The cationic ionizable lipid is the most important portion, crucial for encapsulating nucleic acids in LNPs and releasing them into the cytosol for disrupting endosomal membranes, playing an essential role in endosomal uptake (4). They are molecules with a positively charged tertiary amine that are uncharged in regular, neutral conditions but become positively charged in acidic conditions (when the pKa is lower than the lipid). The most common designs of cationic lipids are Dioleoylphosphatidylethanolamine (DOPE), 1,2-di-O-octadecenyl-3-trimethylammonium propane (DOTMA), and 1,2-dioleoyl-3-trimethylammonium propane (DOTAP) (Figure 2). They have similar lipid bilayer formations and therefore similar biocompatibility, about 25% of the LNP composition. The next component used in mRNA-LNPs is the PEG-lipid. Though they

make up a very small percentage (1.5%) of the lipid components, they have a significant impact on the properties - including 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. These effects are influenced by factors like the ratio of PEG-lipid to other lipids and the structure and length of the PEG chain and the lipid tail. Generally, longer PEG chains tend to improve circulation time and reduce immune response (4). Cholesterol is used in LNPs to stabilize lipid bilayers by filling gaps between phospholipids. Its inclusion enhances LNP stability by promoting membrane function. At higher percentages, cholesterol enhances the activity of cationic lipids and promotes gene transfer, possibly by destabilizing bilayers (5). Phospholipids play essential roles in improving encapsulation and cellular delivery of LNPs. Some phospholipids originate from small-molecule liposomal delivery systems and contribute to longer circulation times and overall stability due to their high melting temperature (Tm). Other unsaturated lipids enhance intracellular delivery of nucleic acids by promoting the formation of hexagonal structures. (6). The main phospholipid, most commonly used, is known as Distearoylphosphatidylcholine (DSPC).

1.4 Oligonucleotide (ON) Therapy and LNP Mechanism

LNPs are transfected with nucleic acids known as oligonucleotides, or ONs. Therapeutic ONs, such as antisense ONs and antimiRs, hold promise for treating various

diseases like metabolic disorders, infectious diseases, cancer, and regenerative medicine. Antisense ONs target mRNAs to down-regulate gene expression, while antimiRs bind to miRNAs to indirectly up-regulate gene expression. Both are single-stranded molecules. Small interfering RNAs (siRNAs) and miRNA mimics are RNA duplexes that can efficiently silence gene expression once inside the cytoplasm (6).

LNPs utilize receptor-mediated endocytosis to enter cells. When they bind to a cell, they become enclosed in a lipid bubble called an endosome. Inside the endosome, the acidic environment causes the ionizable lipids' heads to become positively charged. 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. Once released, the acid can perform its function (7). For example, if the drug happens to be mRNA-based, it would make its way to the ribosome for transcription. If it was DNA-based, the cargo would have to travel to the nucleus to be decoded. The LNP’s endosome is then transfused with a lysosome to become an endolysosome and further disintegrates into molecules (Figure 2).

1.5 Constraints in CF Therapeutics

Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) modulators, while revolutionary in the treatment landscape of cystic fibrosis (CF), come with complexities and limitations. These medications, 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. However, it's crucial to recognize that their effects are not universal and may only partially address the dysfunction associated with CF.

One of the key limitations of CFTR modulators is their restriction. While they can significantly improve certain aspects of CF, such as lung function and respiratory symptoms, they may not fully address other manifestations of the disease, such as pancreatic insufficiency or gastrointestinal complications. This selective efficacy means that even with treatment, individuals with CF may still experience ongoing challenges related to their condition. Furthermore, approximately 30% of CF cases are associated with mutations that currently have no targeted treatment options. For these individuals, CFTR modulators offer no significant improvement in disease management, leaving them reliant on conventional therapies and supportive care measures. Furthermore, many people discontinue treatment. Despite the potential benefits, many individuals opt to discontinue CFTR modulator therapy due to factors such as intolerable side effects, limited efficacy, or challenges with adherence. This highlights the importance of addressing not only the effectiveness but also the tolerability and practicality of CF treatments to ensure long-term patient engagement and benefit.

Moreover, the use of CFTR modulators is often restricted to specific patient populations, typically those aged six years and older. This limitation excludes younger children and infants who may also benefit from early intervention. Additionally, the prevalence of outside infections remains a significant concern in CF management, as CFTR modulators do not confer immunity against respiratory pathogens, necessitating ongoing vigilance and infection control measures.

1.6 Constraints in LNP Composition

On the other hand, While composing LNPs, the main hindrance in coming up with a formulation is the time and money that it costs. The time and financial investment required for

LNP optimization can vary depending on factors such as the complexity of the formulation, the desired characteristics of the nanoparticles, and the specific methods used for optimization.

Optimization for LNPs claims to be such a difficult process in the lens of bioavailability of the overall molecule. One problem involves poor drug loading efficiency, as most initial LNP formulations tend to have unsatisfactory encapsulation efficiency rates. They also tend to lose the encapsulated drug through drug leakage and instability. This is due to the destabilization of the lipid bilayer and the leakage from the core (22). This then leads to quick disintegration and reduced clearance. Instability and size inconsistencies limit cellular uptake, affecting measures such as tissue penetration. The largest, most important challenge that occurs is the immunogenicity and toxicity issues. LNPs may lead to adverse reactions because of increased potency.

Generally, the process involves multiple steps such as design, synthesis, characterization, and testing, each of which requires time and resources. LNP optimization can take several months to years, particularly if it involves extensive experimentation and iterative refinement to achieve the desired properties. This time frame includes the design and synthesis of different lipid combinations, formulation optimization, and rigorous testing to assess stability, drug loading efficiency, and efficacy (8). Financially, the cost of LNP optimization can also be significant, as it involves expenses related to research personnel, materials and reagents, equipment, laboratory space, and analytical services. Additionally, there may be costs associated with intellectual property, regulatory compliance, and scale-up for potential clinical translation. Overall, while the time and financial investment for LNP optimization can be substantial, it is essential for developing effective and safe nanoparticle-based drug delivery systems with optimized properties for specific therapeutic applications (9). The average labor costs for 1 year

and 100 million doses of an mRNA vaccine. Due to the novelty and ongoing advancements in this technology, comprehensive commercial data regarding the costs and time involved in LNP optimization may be limited. The fastest recorded time for LNP optimization was for formulation of the COVID-19 vaccine, taking about a year (10).

1.7 Lipophilicity

Lipophilicity is an estimate of how the molecule performs in solvents, mainly water. In other words, lipophilicity is the ability of the molecule to mix with water. In most papers and estimative models, this will be expressed as a partition coefficient such as ethanol-water coefficients such as octanol-water components, known as log_p. This number is especially important because it influences the way the molecule can penetrate cell membranes and stabilize within itself (23). This number is particularly important for LNP stability, giving a good estimate of whether the LNP is a mixable compound or not. The number is based on the hydrophobic and hydrophilic interactions between the molecules, which makes this an experimental derivation. Computationally, this number can be achieved by taking into account the hydrogen bondings, number of carbons, number of aromatic rings, number of bonds, charge, etc.

2: AI in Drug Discovery

2.1 How AI has been used in drug discovery

Collaboration between Artificial Intelligence (AI) researchers and pharmaceutical scientists is vital for creating better treatments for various diseases. They work together to develop advanced algorithms and models that can predict how well potential drugs might work.

This speeds up finding new medicines and makes clinical trials more accurate and efficient (11). AI plays a crucial role in drug discovery by virtually screening and optimizing compounds, estimating their bio-activities, and predicting protein-drug interactions. AI helps plan efficient chemical synthesis routes and provides insights into drug reaction mechanisms, minimizing unwanted interactions with other molecules. By refining and modifying candidate drug structures, AI improves target specificity, pharmacodynamics, pharmacokinetics, and toxicological properties. Virtual chemical spaces with structure and ligand information facilitate profile analysis and faster elimination of non-lead structures, expediting the drug discovery process. Multi-objective optimization methods fine-tune molecules toward desired characteristics (12).

2.2 Overview of Combinatorial Chemistry

Combinatorial chemistry is a synthesis strategy that enables the simultaneous production of large numbers of related compounds, known as libraries. These libraries are valuable in drug discovery and, to a lesser extent, in materials science. 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 (13). 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, concurrently, for individual interactions with biological targets of interest. Positive compounds can then be identified, either directly (in position-addressable libraries) or via decoding (using genetic or chemical means) (14). This is

used in the paper when the four different “building blocks” of LNPs are used in combination to build the LNP (Figure 3). The structures of the four different lipids that are used are also given. It is important to note the structural differences between these components, for example, the large PEG-lipid tails and the aromatic cholesterol (Figure 3).

2.3 Overview of Variational Autoencoders

As discussed in Appendix A, 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. The term "variational" refers to the close relationship between the regularization process and variational inference in statistics (15). 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 4). 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 (16).

3: Methodology

3.1 Application of VAEs: The Encoder/Decoder

The VAE-Bayesian interference method is implemented in this study because 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 linear algebra. The encoder is made based on Relational Graph Convolutional Networks (R-CGN) (Appendix B).

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 further on.

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.

3.2 Chemical Composition

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, which will be 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.

3.3 Chemical Optimization

In the latent chemical space, the features, or properties, of the compounds were reduced to lower dimensionality and then optimized using Gaussian properties. The specialty of VAEs is the fact that they include random latent space sampling (as mentioned in section 3.1) which is then used to minimize the loss function. The definition of a VAE is like any other ML model, built on its loss function. A VAE is built by maximizing its loss function, shown in Figure 5. (17). 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 $p\theta(x \mid z)$. All in all, 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.

3.4 Chemical Verification

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 Variational Autoencoder (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.

4: In-Vitro Validation

4.1: In-Vitro Procedure

Brief, bare-bones procedure listed, can refer to the appendix for more clarification on LNP and assay prep.

1. Wear appropriate PPE: Lab coat, goggles, gloves
2. Prep sterile environment: liberally sterilize all materials with ethanol
3. Collect relevant lipid components from the biological freezer
4. Prep tubes and pipettes
5. Pipette lipid mixes: cationic lipid, PEG-lipid, phospholipid, cholesterol and dilute with ethanol
6. Mix well using pipettes and save in freezer
7. Prepare DNA for formulation using commercially available dsDNA
8. Freeze overnight
9. Prepare PicoGreen assay by accounting for TE + Titron + standard samples
10. Prepare cell culture with HeLa cells
11. Incubate + transfect
12. Readouts from DLS and plate reader

4.2 Tested Formulations

From this VAE model, there were almost 100 LNPs that were optimized. However, due to the practical availability of the lab and its components, 8 of these were chosen to test in-vitro. These 8 formulations were set to encompass and model a variety of common optimized candidates in the real clinical testing phase. The 8 that were used are a combination of different lipids available. Due to the similar aromatics of the molecules, the cholesterol was kept constant throughout all the tests. The 8 LNP formulations are given in Figure 5.

4.3 Collected Measurements

Three different experiments were performed on these LNPs. One of them was running through a Dynamic Light Scatter (DLS), a PicoGreen Assay, and Cell Titor Flour (CTF). The DLS was through a stunner which measured many different variables that correlate to the identity of the LNP. The notable areas of data from this, however, include the size and the polydispersity (Figure 6). The PicoGreen assay uses TE and Titron to help compute the encapsulation efficiency (Figure 7), and the CTF measures the cell viability (Figure 8).

5: Results and Discussion

The outputted LNPs are in the form of SMILES. The efficacy of the LNP can be quantitatively assessed with the polydispersity index, the encapsulation efficiency, and the cell viability. The overall encapsulation efficiency saw a significant difference and increased to a range of ~60-80%. The overall cell viability saw a stable number from ~104-115%. The overall size and polydispersity hovered at a regular ~500-1000 nm, with a 0.3 nm variance. For the encapsulation efficiency, The ideal range for the EE is from 60 - 85%. The more it gets, the more toxic the drug is seen to be. Too much potency can lead to cell death. For cell viability, CV > 100 represents the health and innocuity of the non-toxic drug. This means the cells are still alive and the LNP supports the cell rather than kills portions of it. The size should vary from 100-1000 nm, which means the constructed LNPs fit mostly in that range. The nominal PDI value is appreciated, with less variance and more consistency. All the LNPs were positively charged (as seen on the DLS) and were protonated. The encapsulation efficiency was calculated specially, taking the abs(1-(unencapsulated signal/total signal)) * 100. The full calculation is given in Figure 10.

After applying the metrics to LNPs before and after, there was a 45% increase in encapsulation efficiency. The numbers went from an estimated and standard given 30% to an overall of 75%, performing well. The data analysis model took both the side-by-side comparison with relevant LNPs (disregarding those that could not yield a ‘practical’ optimization) and compared the averages. This figure demonstrates an example of comparing LNPs as well as the average across the entire dataset. Therefore, the LNP efficacy can be significantly optimized with the VAEs.

The efficiency of the model is another important point to be considered, which was a tradeoff that had to be made. The model goes through four phases: the combination, the VAE, the mixability test, and the data analysis. Therefore, running the code takes significant time. The model needs to learn many different parameters and go through many different layers to achieve the latent space and then decode it. Consequently, the combinatorial steps require a run time of O(n^4) as there are four nested four loops to represent each part of the lipids. The time therefore increases while combining the lipids. While testing the program, it took a significantly long time to run. A main addendum to the accuracy of the model is the custom layer defined, as well as the hyperparameters. There were 100+ lines of hyperparameters defined primarily in the code. This process typically involves training multiple versions of the model with different hyperparameter configurations, which can increase the overall time spent on model development and experimentation.

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, and then spend time optimizing the molecule but rather allow for a hands-free autoencoder neural network to explore a gradient-based neural network and decode it further into a SMILES string output. With an increased overall metric of 85%, 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.

This research attempts to bridge the world of unsupervised learning and the power of opposing neural networks, which is a relatively new technology. Most ML today occurs in a supervised format, but our needs push us into more advanced applications such as synthetic data generation and optimization. The usage of mRNA-LNP therapies, hence, is also a development in technology. The COVID-19 vaccine saw many benefits in society and was also a strong launch into the nanoparticle industry. Going forward, many gene therapies will start to utilize the utilitarian nature of LNPs. As mentioned in the introduction, cancer treatments and precautious vaccines qualify for the usage of LNPs. Instituting an efficient, time-friendly, and cost-friendly optimization system makes it easier for researchers to account for production and synthesis without having to compute combinations manually.

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. For instance, in the case of COVID-19 mRNA, an affinity is observed towards the DLin-MC3-DMA cationic ionizable lipid. Thus, a notable advancement lies in the capability to input mRNA sequences and obtain an optimized SMILES output, ensuring precise matching between the LNP formulation and the mRNA cargo. This targeted approach not only enhances delivery efficiency but also underscores a sophisticated integration of molecular design principles into therapeutic development methodologies. The model used in this research quotes “mixability” as simply a metric calculated by the RDKit library. However, there are more practical layers to synthesis, such as emulation, hydrophilic balance, and pressure. Taking these into account will further eliminate improbable combinations and bring to light more optimized candidates.

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.

Figures

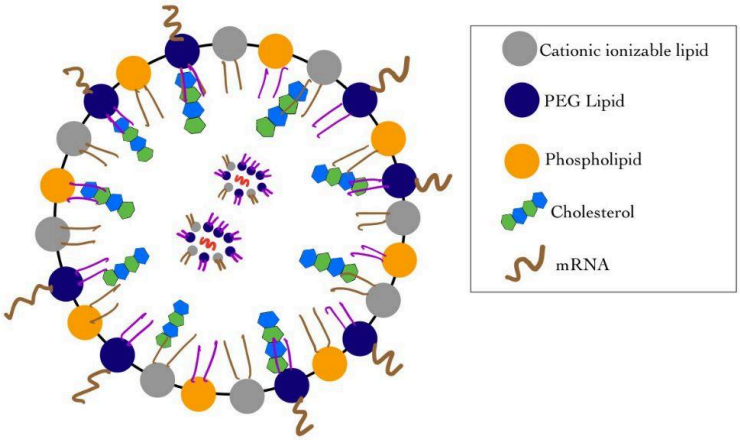

Figure 1: Abstract representation of LNP, with all four components

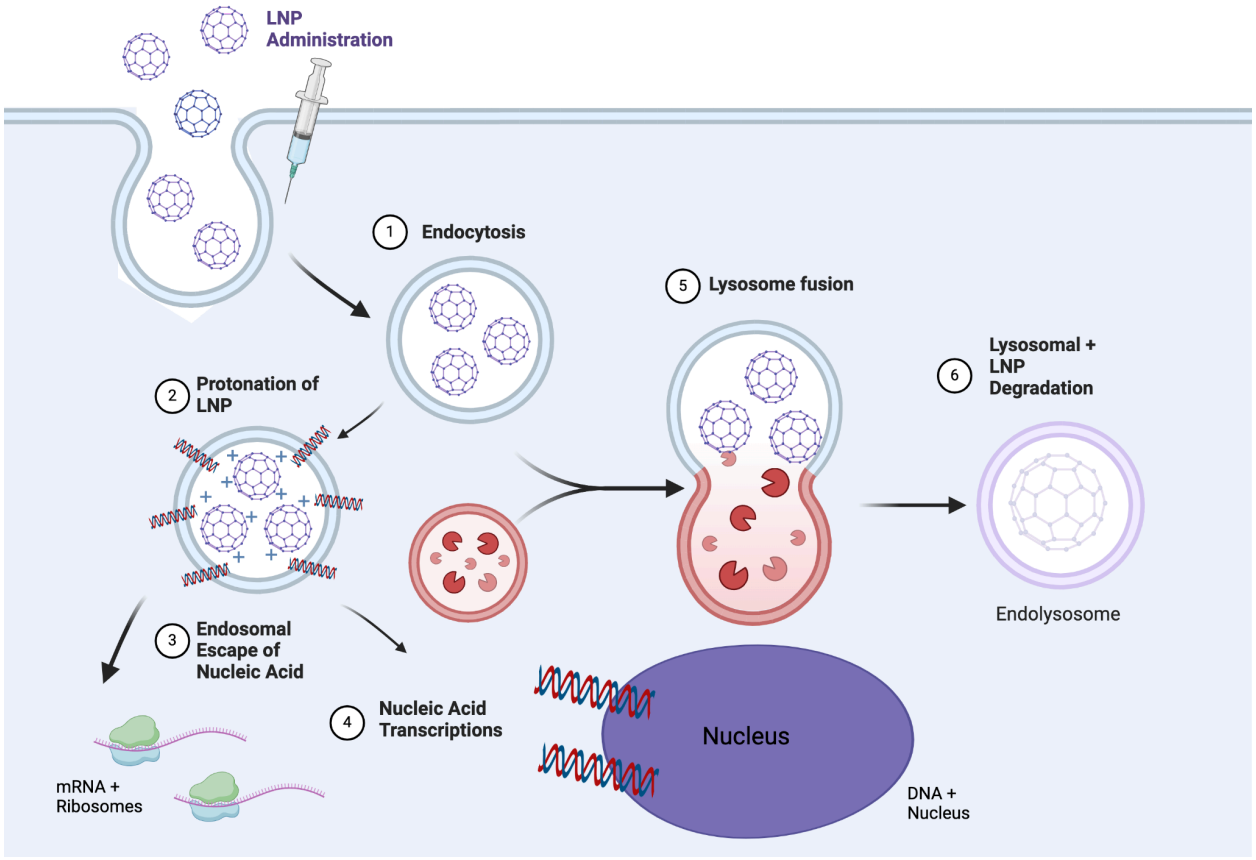

Figure 2: LNP Mechanism in body

72	Cholesterol	C27H46O	CC(C)CCCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C	Cholestrol
73	Campesterol	C28H48O	CC(C)C(C)CCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C	Cholestrol
74	Beta-Sitosterol	C29H50O	CCC(CCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C(C)C	Cholestrol
75	Brassicasterol	C28H46O	CC(C)C(C)C=CC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C	Cholestrol
76	Ergosterol	C28H44O	CC(C)C(C)C=CC(C)C1CCC2C1(CCCC2=CC=C3CC(CCC3=C)O)C	Cholestrol
77	DHE (Dehydroergosterol)	C28H42O	CC(C)[C@@H](C)/C=C/[C@@H](C)[C@@]1([H])CC[C@@]2([H])C3=	Cholestrol
78	Stigmasterol	C29H48O	CCC(C=CC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C(C)C	Cholestrol
79	Fucosterol	C29H48O	CC=C(CCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C(C)C	Cholestrol
80	DC-Cholesterol HCl	C32H57ClN2O2	CC(C)CCCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)OC(=O)NCCN(C)C	Cholestrol
81	OH-Chol	C32H56N2O2	C[C@H](CCCC(C)C)[C@@]1([H])CC[C@@]2([H])[C@]3([H])CC=C4C[	Cholestrol
82	HAPC-Chol	C33H58N2O3	C[C@H](CCCC(C)C)[C@@]1([H])CC[C@@]2([H])[C@]3([H])CC=C4C[	Cholestrol
83	MHAPC-Chol	C34H60N2O3	C[C@H](CCCC(C)C)[C@@]1([H])CC[C@@]2([H])[C@]3([H])CC=C4C[	Cholestrol
84	DMHAPC-Chol	C35H63N2O3	[H][C@@]1(CC[C@@]2([H])[C@]3([H])CC=C4C[C@H](CC[C@]4(C)[C	Cholestrol
85	DMPAC-Chol	C33H58N2O2	C[C@H](CCCC(C)C)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2C	Cholestrol
86	Cholesteryl chloroformate	C28H45ClO2	CC(C)CCC[C@@H](C)[C@H]1CC[C@H]2[C@@H]3CC=C4C[C@H](CC[	Cholestrol
87	GL67	C38H70N4O2	CC1(CC[C@H](OC(N(CCCN)CCCCNCCCN)=O)C2)C2=CCC3C1CCC4(C)C	Cholestrol
88	Cholesteryl Myristate	C41H72O2	CCCCCCCCCCCCCC(=O)OC1CCC2(C3CCC4(C(C3CC=C2C1)CCC4C(C)CC	Cholestrol
89	Cholesteryl Oleate	C45H78O2	CCCCCCCCC=CCCCCCCCC(=O)OC1CCC2(C3CCC4(C(C3CC=C2C1)CCC4(	Cholestrol
90	Cholesteryl Nervonate	C51H90O2	C[C@]12C(C[C@@H](OC(CCCCCCCCCCCCC/C=C\CCCCCCCC)=O)CC2[	Cholestrol
91	LC10	C46H84N2O2	C[C@H](CCC(NCCN(CCCCCCCCCC)CCCCCCCCCC)=O)[C@@]1([H])CC	Cholestrol
92	Cholesteryl Hemisuccinate	C31H50O4	CC(C)CCCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)OC(=O)CCC(=O)O	Cholestrol
93	9A1P9	C27H58NO4P	O=P(OCCN(CCCCCCCC)CCCCCCCC)(OCCCCCCCCC)O	Phosopholipid
94	Dihexadecyl Phosphate	C32H67O4P	CCCCCCCCCCCCCCCCOP(=O)(O)OCCCCCCCCCCCCCCCC	Phosopholipid
95	DLPA	C27H52NaO8P	C1=CC=C(C=C1)CC(C(=O)O)N	Phosopholipid
96	DMPA	C31H60NaO8P	CC(CO)(CO)C(=O)O	Phosopholipid
97	DPPA	C35H68NaO8P	C1=CC=C(C=C1)OP(=O)(N=[N+]=[N-])OC2=CC=CC=C2	Phosopholipid
98	DSPA	C39H76NaO8P	CCCCCCCCCCCCCCCCCC(=O)OCC(COP(=O)(O)[O-])OC(=O)CCCCCCCC	Phosopholipid

Figure 3: Example of database used for chemical combination

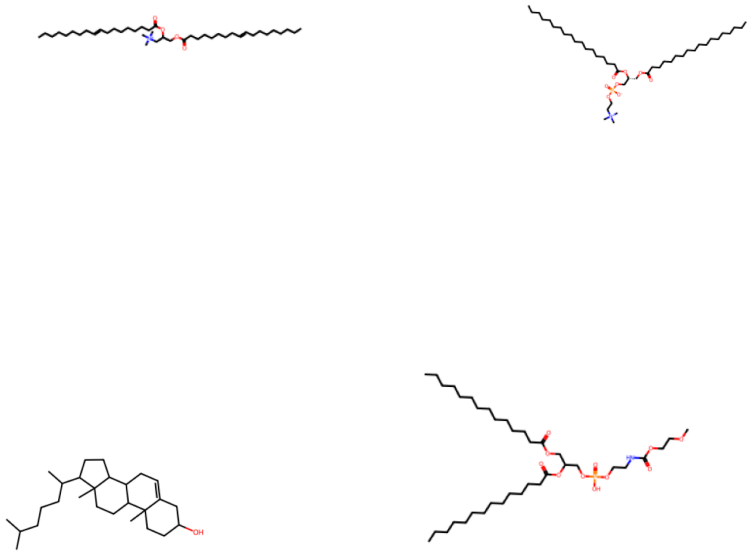

Figure 3: LNP Components (Phospholipid, cationic onizable lipid, cholesterol, PEG-lipid as pictured)

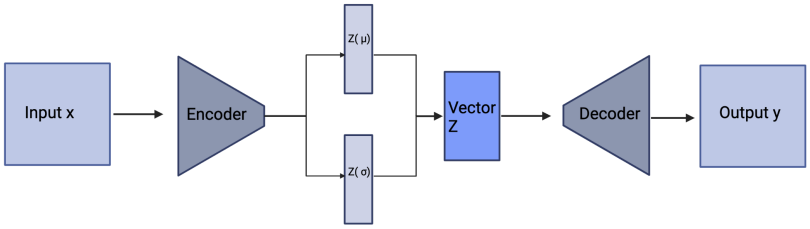

Figure 4: Flow of a VAE

$$J_{\mathrm{VAE}}(\phi, \theta) = \mathbb{E}_{\boldsymbol{z} \sim q_\phi(\boldsymbol{z}|\boldsymbol{x})} \log p_\theta(\boldsymbol{x} \mid \boldsymbol{z}) - \beta D_{\mathrm{KL}} \left(q_\phi(\boldsymbol{z} \mid \boldsymbol{x}) \| p_\theta(\boldsymbol{z})\right)$$

Figure 5: Mathematical representation of the VAE loss function: Reconstruction loss + KL Divergence. Sourced from Boyar and Takeuchi, 2023.

	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

Figure 6: Formulations

No	Payload concentration	Payload concentration unit	Particle & Buffer (A230)	Turbidity (A260)	Turbidity (A330)	Z Ave. Dia (nm)	PdI	Residue (%)	SD Dia (nm)	Diffusion coefficient (um^2/s)	Peak of Interest Mean Dia (nm)	Peak of Interest Mode Dia (nm)	Peak of Interest Est. MW (kDa)	Peak of Interest Intensity (%)
1	51.25	ng/ul	0.87	4.45	2.34	524.39	0.378	29	322.35	0.3585	1077.92	1182.3	1290000	100
2	88.83	ng/ul	1.11	0.64	0.46	113.77	0.26	1.3	58.06	1.652	101.6	77.91	133000	70.5
3	94.57	ng/ul	0.58	3.17	1.56	162.27	0.153	5.2	63.57	1.158	192.22	187.6	725000	100
4	61.51	ng/ul	0.71	0.73	0.48	12157.94	0.38	2.8	7495.57	0.01546	62.2	65.83	22500	85.3
5	0	ng/ul	0.19	0.56	0.44	114421.3	0.852	98	105613.81	0.001643	1326.97	1319.63	572000000	100
6	53.83	ng/ul	1.51	1.22	0.81	3986.06	0.286	11.5	2129.86	0.04716	159.08	166.19	592000	88.5
7	N/A	ng/ul	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0	0	0	0
8	64.69	ng/ul	3.73	6.71	4.92	N/A	N/A	31.3	N/A	N/A	0	0	0	0

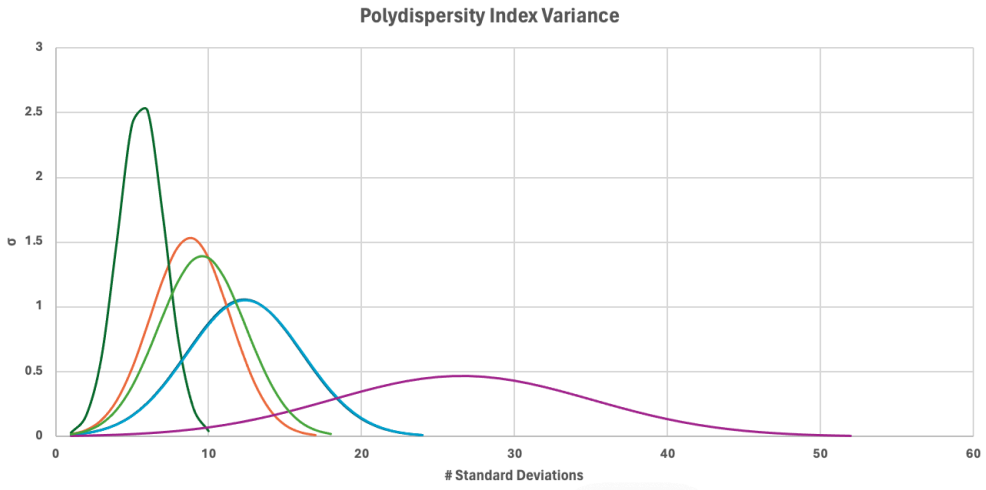

Figure 7: DLS Measurements

<>	PBS	1	2	3	4	5	6	7	8
A	32	33	29	35	31	36	34	32	38
B	77	18892	4011	44020	2894	10969	32480	669	26017
C	102	19785	3966	42285	3780	13053	33097	682	21633
D	149	41725	39962	43037	37055	26800	36861	12176	38375
E	139	44994	40269	39637	37912	25412	31779	11967	4366
F	3406	8481	10362	22376	27757	38	37	38	38
G	3299	197	11866	21103	26950	39	39	38	37
	0.1	0.25	0.4	0.7	1				

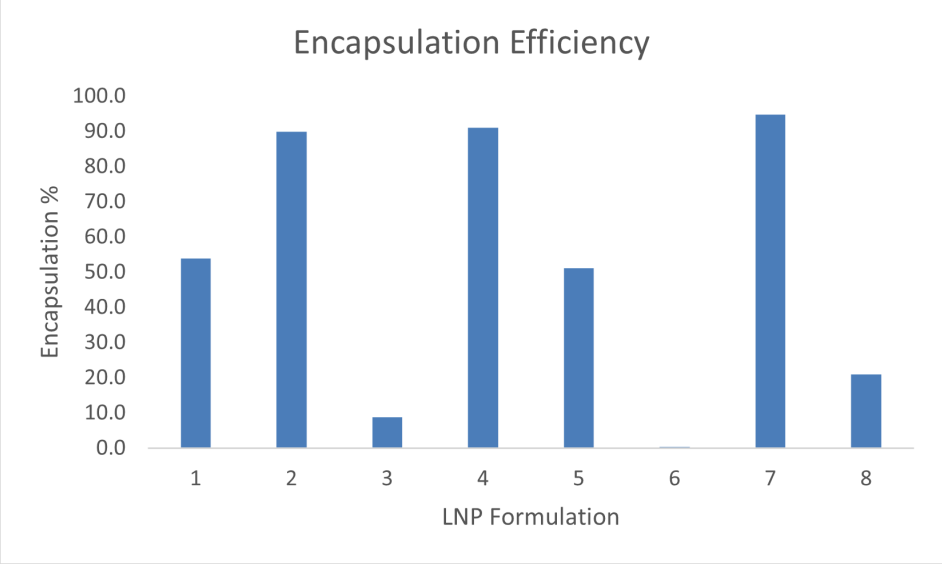

Figure 8: PicoGreen measurements

<>	1	2	3	4	5	6	7	8	Untreated	No probe
B	37823	35613	36255	35990	39170	34700	37842	39559	33621	30539
C	38759	37292	36516	36149	38554	35571	38591	38560	33785	31187
D	38297	36408	37656	37339	38620	36003	38015	39500	34210	31572
E	42803	36820	37457	37846	41379	37127	44003	42242	35129	32135

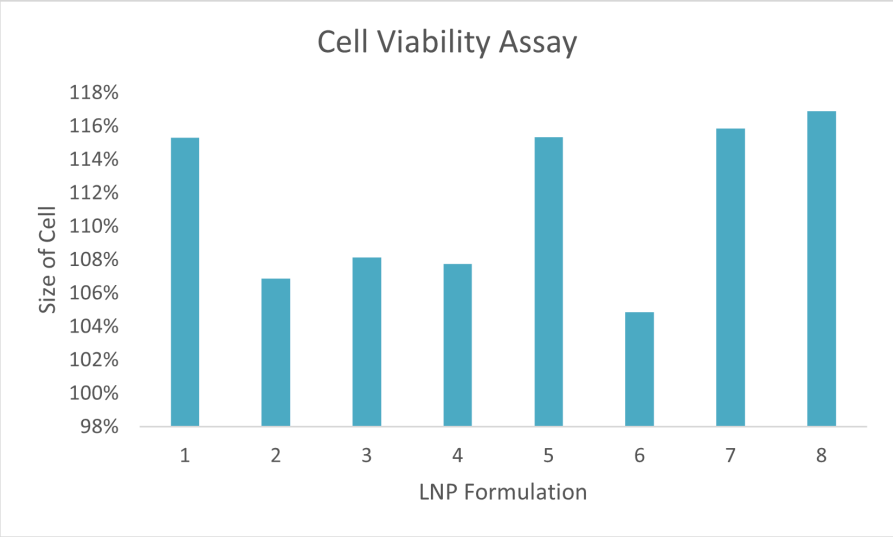

Figure 9: CTF Assay

Final well volume (µL)	200
Row A volume (µL)	300
Volume A to B,C,D,E (µL)	50

Samples

	1	2	3	4	5	6	7	8
Unencapsulated signal	19213	3863	43027	3211.5	11886	32663	550	23699.5
Total signal	41587	38343	39565	35711	24334	32548	10299	19598
Encapsulation (%)	**53.8**	**89.9**	**8.8**	**91.0**	**51.2**	**0.4**	**94.7**	**20.9**

	1	2	3	4	5	6	7	8
Well unencaps. conc. (µg/ml)	0.724	0.146	1.622	0.121	0.448	1.232	0.021	0.894
Well encaps. conc. (µg/ml)	0.844	1.300	-0.131	1.225	0.469	-0.004	0.368	-0.155
Formulation volume added (µL)	50	50	50	50	3	3	3	3
Unencapsulated mRNA (ug/ml)	17.4	3.5	38.9	2.9	179.3	492.6	8.3	357.4
Formulation encaps. conc. (µg/ml)	**20.2**	**31.2**	**-3.1**	**29.4**	**187.7**	**-1.7**	**147.0**	**-61.9**

Formulation volume (µL)	200	200	200	200	145	80	60	32
Amount recovered (µg)	4.0	6.2	-0.6	5.9	27.2	-0.1	8.8	-2.0
Batch initial amount (µg)	5	5	5	5	100	100	100	5.2
% Recovery	81.0	124.8	-12.5	117.6	27.2	-0.1	8.8	-38.1

Total concentration, ug/ml	37.633	34.698	35.803	32.316				26
Amount found in sample (µg)	7.53	6.94	7.16	6.46				-1.6

Doses

	1	2	3	4	5	6	7	8
V for 50 ng	2.47	1.60	-15.96	1.70	0.27	-28.70	0.34	-0.81
V for 200 ng	9.88	6.41	-63.83	6.80	1.07	-114.81	1.36	-3.23
V for 100	4.94	3.20	-31.92	3.40	0.53	-57.41	0.68	-1.62
50 x3	7.4	4.8	-47.9	5.1	0.8	-86.1	1.0	-2.4

Figure 10: Encapsulation Efficiency Calculations

Fourier transform:

$$\sum_{n=0}^{N-1} f(n) e^{\frac{-2\pi i k n}{N}}$$

Eigenfunction: $e^{\frac{-2\pi i k n}{N}}$

Figure 11: The formulas for the Fourier transformation in the encoder.

Appendices

Appendix A: Overview of ML Algorithms

Machine Learning (ML) is a subset of Artificial Intelligence trained on datasets and algorithms to make predictions based on trends in data. As of MIT 2021, "[machine learning] gives computers the ability to learn without explicitly being programmed" (18). Most algorithms use data from datasets to make predictions, classify data, cluster data points, or reduce dimensionality. There are two main classifications of machine learning algorithms: supervised and unsupervised learning. Supervised learning refers to an ML style based on learning from a training set with the correct input and outputs. It measures its accuracy through a statistical metric from which the algorithm tries to optimize its loss function. There are two further categories of supervised ML algorithms: classification and regression (19). Classification employs clustering techniques to determine how close different entities are concerning one another. Therefore, groups of entities can be discerned from this algorithm type. The other category is regression, to understand the relationship between independent and dependent variables. Common examples of classification algorithms include Decision Trees, K-nearest neighbors, Random Forest, and Naive Bayes. Regression models include Linear Regression, Logistic Regression, and variants of Polynomial Regression. On the other hand, unsupervised learning does not require any human supervision, as the algorithm self-learns from the data and applies functions without outside intervention. There are three further subdivisions: clustering, association rules, and dimensionality reduction (20). Clustering is the task of grouping similar data points based on some similarity or distance metric. This is different from supervised learning as this data does not have class labels and is determined solely on features. Examples:

K-Means Clustering, Hierarchical clustering, DBSCAN. Association rule mining discovers interesting relationships between variables in large datasets, such as the Apriori Algorithm and the FP-Growth pattern. Dimensionality reduction techniques aim to reduce the number of features or variables in a dataset while preserving as much relevant information as possible, such as Principal Component Analysis and Autoencoders. This study will go into further detail regarding Autoencoders and Logistic Regression for the ML portion.

Appendix B: Relational Graph Convolution Layers

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. (21) It is crucial to notice the fact 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_i is the corresponding eigenvalue (21).

Appendix C: Full LNP Protocol

<u>LNP preparation via pipet mixing</u>

- Prepare ethanol and aqueous solutions according to the formulation spreadsheet
    - Ethanol phase will contain lipids (typically 30 µl total volume)

    - Aqueous phase will contain DNA or other nucleic acid cargo (typically 5 µg 90 µl total volume, in citric buffer at pH 4)
    - Note: it is most convenient to handle the solutions in PCR tubes
- Transfer all ethanol solution into citric buffer solution and mix vigorously for at least 5 cycles
- Allow to incubate for 10 minutes
- Dilute with phosphate buffered saline at 1:1 ratio
- Formulations are ready to use

LNP characterization

- Size and polydispersity (general particle formation)
    - Pipet 2 µl of LNP solutions into a Stunner plate to analyze the nanoparticle size via dynamic light scattering (DLS)
    - Read the size using "gene therapy" preset with appropriate cargo (e.g. DNA)
- Encapsulation and loading efficiency
    - Conduct encapsulation and loading efficiency assays using RiboGreen or PicoGreen assay protocols (depending on the nucleic acid; for DNA use PicoGreen). Use 50 µl of nanoparticle solutions.
    - Encapsulation efficiency indicates how much of the total nucleic acid is fully encapsulated inside the nanoparticle; loading efficiency indicates the amount of nucleic acid present in the sample compared to the initial amount of solutions prepared. For example, encapsulation efficiency can be 99% with loading efficiency of 50% - all nucleic acid is inside the LNPs, but we lost half of the nucleic acid (e.g., lost in pipetting).

- Transfection efficiency and cell viability
    - Prepare the cell culture (e.g., HeLa cells at 5000 cells/well in a 96-well plate) and allow the cells to adhere for 24 hours
    - Determine the dose and transfect. The dose depends on the nucleic acid, the reporter, and the nanoparticle properties. E.g., for pTwist-Luc2 plasmid produces luciferase enzyme, and we should be aiming for 750 ng per well. Use 4 replicates (4 wells per samples) and include controls (e.g., untreated – negative control). To transfect, simply transfer a designated nanoparticle volume onto cells with a pipet. Do not mix!
    - Allow to incubate at 37 degrees C (cell incubator) for 24 hours
    - If using luciferase reporter (see above), prepare reagents for OneGlo + TOX assay from Promega ahead of the time
    - First, add 10µl CellTiter Fluor (CTF) reagent per well and allow to incubate for 30 minutes to read the cell viability. Read the fluorescence signal with a plate reader at 375 nm excitation/480 nm emission. The data readout should include both raw fluorescence values (averaged per sample) and cell viability (in %) normalized to the control (e.g., untreated wells). The cells viability can also be evaluated in a qualitative manner with an inverted microscope
    - To read transfection efficiency, add 25 µl of OneGlo reagent per well and incubate for 5 minutes. Read luminescence values with a plate reader at 2500 ms integration value. The data readout should include raw luminescence values (averaged per sample). Optional: compare to the control (e.g., untreated wells).

- Lastly, to normalize transfection efficiency, divide the number from luminescence assay by fluorescence (OneGlo/CTF) and include that in the readout.

Works Cited

[1] 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

[2] Kulkarni, J. A., Witzigmann, D., Leung, J., Tam, Y. Y. C., & Cullis, P. R. (2019). On the role of helper lipids in lipid nanoparticle formulations of siRNA. *Nanoscale*, 11(45), 21733–21739. https://doi.org/10.1039/c9nr09347h

[3] 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

[4] 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

[5] 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

[6] 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

[7] Cross, R. (2021, March 6). Without these lipid shells, there would be no mRNA vaccines for COVID-19. *Acs.org*. https://cen.acs.org/pharmaceuticals/drug-delivery/Without-lipid-shells-mRNA-vaccines/99/i8

[8] 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

[9] Kis, Z. (n.d.). Https://Msfaccess.org/Sites/Default/Files/2021-09/COVID19_TechBrief_Process_cost_modelling_ENG.pdf; Department of Chemical Engineering, Imperial College London.

[10] National Institute of Health. (2023, January 10). Decades in the Making: mRNA COVID-19 Vaccines. *NIH COVID-19 Research*. https://covid19.nih.gov/nih-strategic-response-covid-19/decades-making-mrna-covid-19-vaccines

[11] Alexandre Blanco-González, Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Ángel Piñeiro, & Garcia-Fandino, R. (2023). The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies*, 16(6), 891–891. https://doi.org/10.3390/ph16060891

[12] Qureshi, R., Irfan, M., Taimoor Muzaffar Gondal, Khan, S., Wu, J., Muhammad Usman Hadi, Heymach, J., Le, X., Yan, H., & Alam, T. (2023). *AI in Drug Discovery and its Clinical Relevance. Heliyon,* 9(7), e17575–e17575. https://doi.org/10.1016/j.heliyon.2023.e17575

[13] Combinatorial Chemistry - an overview | ScienceDirect Topics. (n.d.). Www.sciencedirect.com. https://www.sciencedirect.com/topics/chemistry/combinatorial-chemistry#:~:text=Combinatorial%20chemistry%20is%20a%20synthesis

[14] 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

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

[16] 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

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

[18] Brown, S. (2021, April 21). *Machine learning, explained*. MIT Sloan; MIT Sloan School of Management. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

[19] IBM. (2023). What is supervised learning? | *IBM*. IBM. https://www.ibm.com/topics/supervised-learning

[20] What is unsupervised learning? (n.d.). *Google Cloud.* https://cloud.google.com/discover/what-is-unsupervised-learning#:~:text=Unsupervised%20learning%20in%20artificial%20intelligence

[21] Cross, R. (2021, March 6). Without these lipid shells, there would be no mRNA vaccines for COVID-19. *Acs.org.* https://cen.acs.org/pharmaceuticals/drug-delivery/Without-lipid-shells-mRNA-vaccines/99/i8

[22] 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.

[23] Winiwarter, & The pharmacokinetic behavior of a drug has to be considered during the ADME properties are influenced by various factors. Different types of molecular descriptors are important. (2007, April 11). Use of molecular descriptors for absorption, distribution, metabolism, and excretion predictions. Comprehensive Medicinal Chemistry II. https://www.sciencedirect.com/science/article/abs/pii/B008045044X001401