03.04.2025

Introducing ANNA: Turning Data into Better LNPs

Guðrún Hulda Ólafsdóttir[a)], Iain McFadyen[b)], Simon Klüpfel[a)]

[a)] Axelyf ehf., Hafnarfjörður, Iceland [b)] Applied Insight Consulting, Stoneham MA 02180, USA

Summary

This whitepaper introduces ANNA, Axelyf's novel AI model for predicting lipid nanoparticle (LNP) performance in mRNA delivery. By training on diverse and heterogeneous datasets without manual normalization, ANNA enables more flexible, real-world relevant predictions than previous approaches. We benchmark ANNA against AGILE and LiON to demonstrate its performance and generalizability.

LNPs for mRNA delivery

The field of RNA-based therapies is transforming modern medicine, from life-saving vaccines to promising treatments for cancer and genetic disorders[1]. However, the full range of applications of RNAs, especially large versions like mRNA, is currently limited by several challenges in delivery. To be effective, mRNA protection in vivo is key, so it needs to be delivered in a vehicle that protects it from degradation by nucleases while it transits through the body to the target cells. Potency of delivery then depends on uptake and efficient release of mRNA inside cells to be translated into protein. Safety in terms of vehicle effects on immune stimulation and liver accumulation of lipids are also critical to developability of products with RNA. Addressing selectively to tissue and/or cell type(s) is an ongoing challenge, in particular in the quest for non-viral delivery options. Lipid nanoparticles (LNPs) are a prime example of a non-viral delivery technology that has proven effective as a vehicle for mRNA delivery, as demonstrated by the success of the COVID-19 vaccines from Moderna and BioNTech/Pfizer[2]. However, further advancements in LNP design are needed to unlock the full range of applications of mRNA through improved targeting, enhanced efficiency, and safe repeat dosing, especially for therapeutic applications.

LNPs for RNA delivery are typically made of four key lipid components, *i.e.,* an ionizable lipid, cholesterol, a PEG lipid conjugate, and a phospholipid. Each of these components plays a distinct role in LNP performance and mRNA delivery. Ionizable lipids are particularly important and have been the focus of extensive research in the past two decades because they drive in large part the potency of the particle for delivery. Ionizable lipids, also known as amino lipids, have pH-responsive behavior that facilitates endosomal escape, reduce off-target interactions due to their neutral charge at physiological pH, and affect tissue distribution and delivery efficiency.

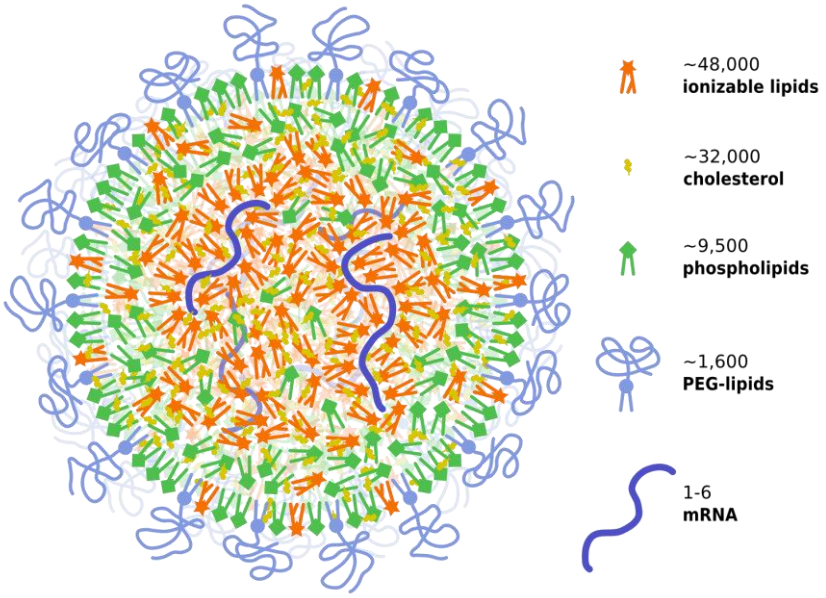

Figure 1: *LNPs are roughly spherical particles in the range of 10's to 100+ nm in diameter and made of several types of lipids and cargos. LNPs are used as delivery vehicles for multiple therapeutic molecules, from small NCEs, to small RNAs, to large mRNAs and DNA constructs encoding for functional proteins and protein complexes. The estimated numbers of each LNP component in a single particle refer to a 70-nm diameter LNP.*

Axelyf is innovating on every aspect of LNP chemistry, leveraging its best-in-class lipid library to design more effective and precisely targeted delivery systems. By improving delivery and fine-tuning each component to match specific therapeutic goals, we can dramatically expand the reach of mRNA-based medicine, reduce doses through potency improvements, while minimizing undesired side effects for patients. We are now pairing the proven expertise of our team with AI to push the edge of innovation in nanoparticle design.

Advancing LNP Design with AI

Artificial Intelligence (AI) has catalyzed huge progress in many industries, including the life sciences. From protein structure prediction with AlphaFold to the latest large language models being applied to protein and RNA design, the progress in the past decade has been breathtaking. Similarly, AI has the potential to revolutionize lipid and LNP design.

Modeling small molecules, and especially lipids used in large assemblies like LNPs, presents unique challenges due to the vast chemical space and non-linear nature of 3D molecular structure. Additionally, the biological performance of RNA-LNPs is the net result of a complex multi-step process that is not completely understood and relies on scaling multiple hurdles, from LNP assembly in the lab through to biological effects in an organism. These challenges make advances in the field more difficult to achieve but also underscore the exciting potential of using AI to uncover patterns and design principles that traditional approaches might miss.

Although the application of AI to LNPs is still nascent, more studies are published each year, and interest continues to grow due to the broad range of applications of RNA medicine.

Limitations of Current Machine Learning Approaches for LNP Design

Most existing research on LNP modeling has focused on ionizable lipids as the key variable for predicting LNP potency. These models are commonly used for *in silico* screening, helping to identify promising candidates for synthesis and testing. While such models provide a useful starting point, they often fail to fully capture the complexity of *in vivo* LNP performance. A major challenge remains in establishing a meaningful connection between virtual predictions and experimentally relevant factors, ensuring that model outputs offer actionable insights rather than broad categorizations.

Another major challenge in this field is the scarcity of large and diverse datasets of lipids or LNPs with high quality and relevant biological readouts. Existing datasets are also difficult to compare, due to a combination of factors such as lack of standardized reference samples, incomplete formulation process meta-data, differing experimental models, and incompatible measures used for the processed raw experimental results. Hence it is challenging to harmonize and aggregate existing datasets into larger meta-datasets for training AI models.

Axelyf's approach to AI driven LNP design

At Axelyf, we are dedicated to leading innovation in AI-driven lipid design by translating scientific insight into practical tools through a purposeful, pragmatic approach that serves real-world research needs. Our focus is to leverage the full informational value of existing datasets and to enable identification of truly high-performing lipids from virtual libraries by relating their potency to known reference LNPs.

To achieve this, we designed an Artificial Network for Nanoparticle Assessment model (ANNA) that predicts potency unambiguously on a common scale for all samples of the meta-dataset, or virtual screening libraries, leading to consistent and comparable predictions.

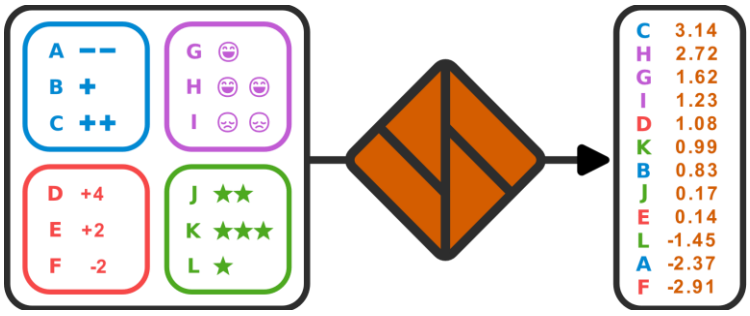

Figure 2: *Illustration of ANNA's novel approach. ANNA is trained directly on multiple, diverse datasets in their original form. This avoids an intermediate step of manually mapping all samples on a common scale, which in general requires additional information, or assumptions, to correlate samples from different datasets.*

To validate ANNA's performance, we benchmarked it against two recent modeling efforts that capture different data complexity, modeling approaches, and training strategies.

The first group[3] used a single in vitro dataset to train AGILE, a fine-tuned model designed to predict LNP potency. When benchmarked on this dataset, ANNA outperformed AGILE in identifying high-performing lipids, demonstrating strong predictive capability with limited, homogeneous data (Figure 3).

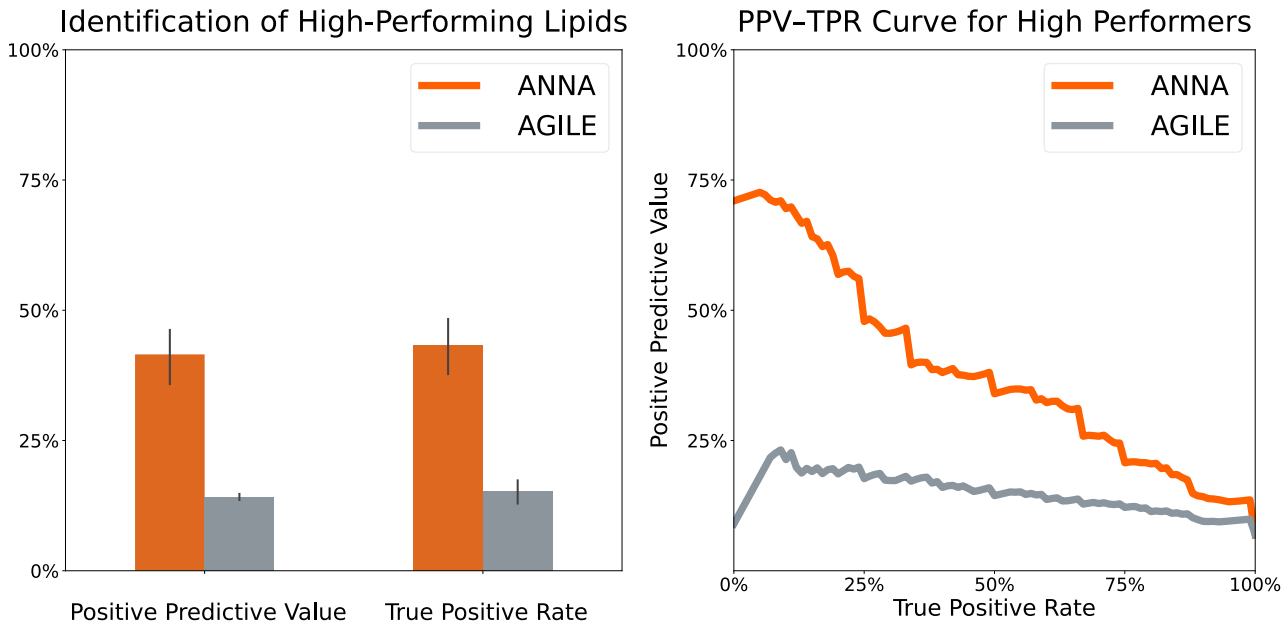

Figure 3: *Comparison of ANNA and AGILE in identifying high-performing lipids. (Left) Positive Predictive Value (PPV) and True Positive Rate (TPR) for both methods, with ANNA outperforming AGILE. (Right) PPV-TPR curve showing ANNA maintaining higher predictive performance across different true positive rates. Further details on the methodology and evaluation metrics can be found in the technical details section below.*

The second group[4] introduced a broader benchmark: a more diverse, aggregated dataset composed of both published and unpublished results. It was used to train LiON, a model built to generalize across *in vitro* and *in vivo* applications. The authors found the model to benefit from training on the full set of available data rather than exclusively on data matching the specific type of targeted predictions. To directly assess how well each model identifies potent samples, we defined a subset of high-expression samples from one of the training studies and measured predictive accuracy (Figure 4). Both ANNA and LiON achieved high positive predictive values with reasonable true-positive rates.

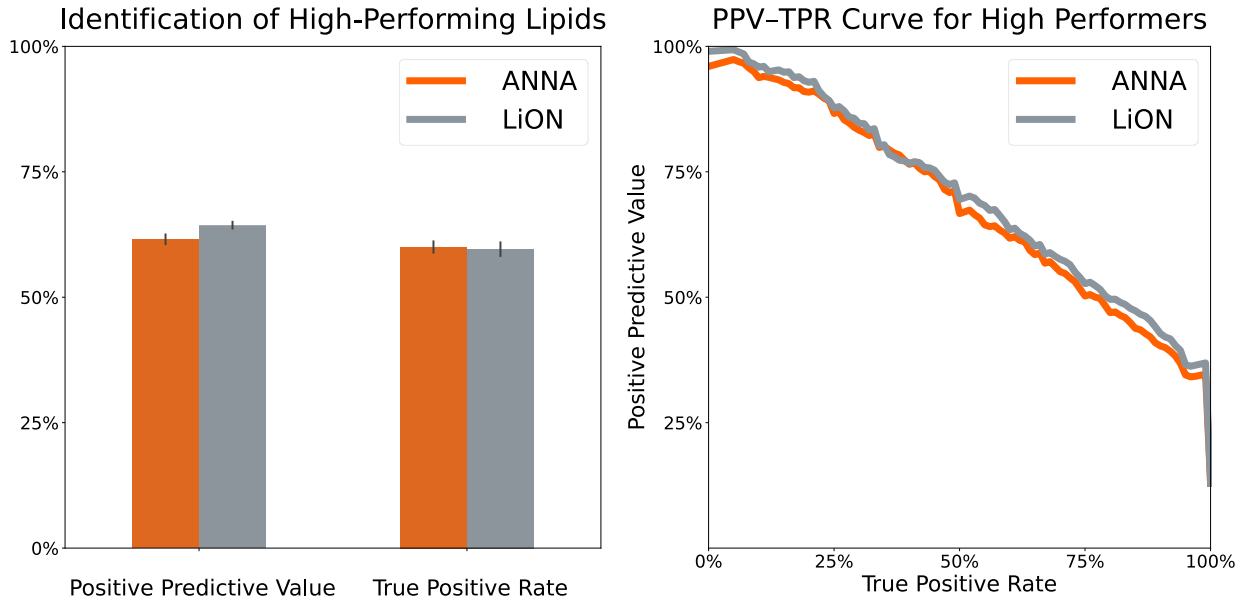

Figure 4: *Comparison of ANNA and LiON in identifying high-performing lipids in one study. The study was used in its entirety for model training and evaluation. (Left) Positive Predictive Value and True Positive Rate for both methods, showing comparable performance. (Right) PPV-TPR curve illustrating the similarity between ANNA and LiON across different true positive rates. Further details on the methodology and evaluation metrics can be found in the technical details section below.*

However, in a more realistic and challenging evaluation scenario, ANNA's advantage becomes clear. By splitting this study's samples into five smaller subsets, mimicking the common challenge of having to train a model on several smaller studies that cannot necessarily be combined using a common scale, we tested each model's ability to generalize. ANNA significantly outperformed LiON in identifying top-performing LNPs in this setting (Figure 5).

These results illustrate a key strength of our model: ANNA can make predictions on a global scale, even when only trained on smaller studies on their respective local scale. This is a particularly valuable model property given the current scarcity of large, standardized LNP datasets in the field.

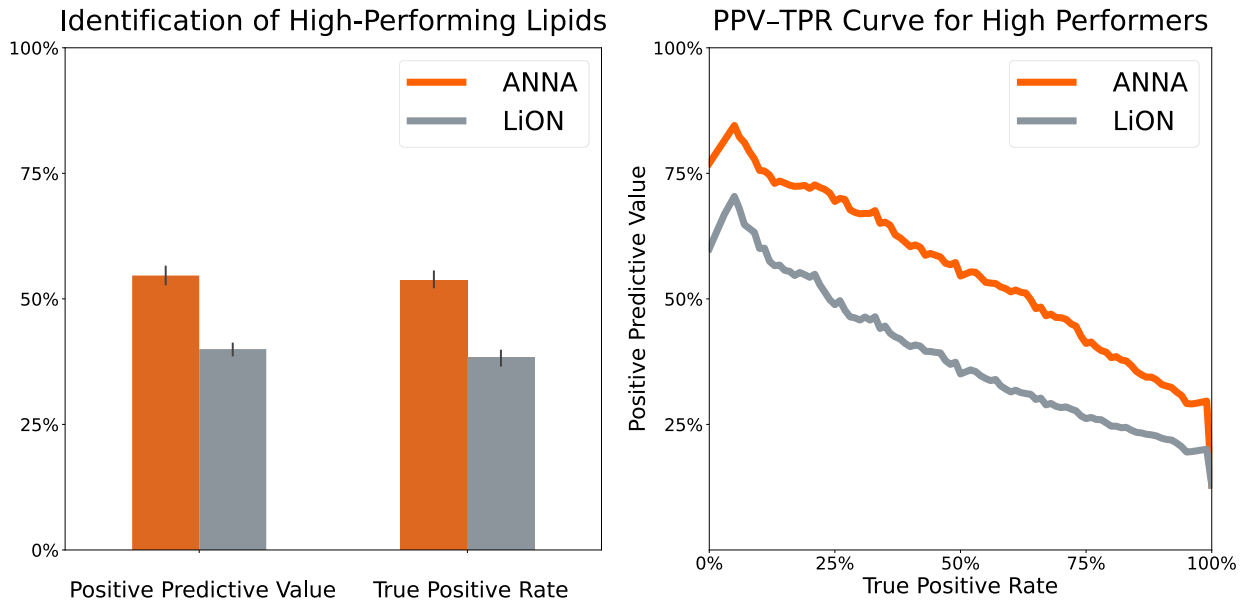

Figure 5: *Comparison of ANNA and LiON identifying high-performing lipids in one study. The study was artificially divided into five parts, treating them as separate studies to assess model generalizability. (Left) Positive Predictive Value and True Positive Rate for both methods, showing ANNA outperforming LiON. (Right) PPV-TPR curve demonstrating that ANNA maintains higher predictive performance across different true positive rates. Further details on the methodology and evaluation metrics can be found in the technical details section below.*

Conclusion

Modeling of LNP properties will play a critical role in the development of future mRNA medicines. A model capable of reliably identifying high performing lipids will reduce the efforts, time and money spent on experimental assessment of LNPs, by focusing inherently limited resources on the most promising candidates. The limited quality and quantity of the currently available experimental data hinders model development using standard techniques. By design, the ANNA model can be trained on multiple, diverse, and not directly comparable datasets, offering greater flexibility in integrating data from different sources. The results presented here demonstrate that ANNA outperforms state-of-the-art models in realistic and challenging data environments.

We are continually working to expand our own training sets and data diversity to further refine our models and benchmark against the state of the art. In the process of advancing modeling of RNA-LNP properties and performance, our approach prioritizes thoughtful model development with an emphasis on real-world usability. By addressing key challenges in data integration and model generalizability, we contribute to the long-term advancement of predictive modeling in LNP research, ensuring that these tools are not only accurate but also applicable in practical settings.

References

[1] S. Agrawal, "RNA Therapeutics Are Stepping Out of the Maze," *Trends Mol. Med.*, vol. 26, no. 12, pp. 1061–1064, Dec. 2020, doi: 10.1016/j.molmed.2020.08.007.

[2] X. Hou, T. Zaks, R. Langer, and Y. Dong, "Lipid nanoparticles for mRNA delivery," *Nat. Rev. Mater.*, vol. 6, no. 12, pp. 1078–1094, Dec. 2021, doi: 10.1038/s41578-021-00358-0.

[3] Y. Xu *et al.*, "AGILE platform: a deep learning powered approach to accelerate LNP development for mRNA delivery," *Nat. Commun.*, vol. 15, no. 1, p. 6305, Jul. 2024, doi: 10.1038/s41467-024-50619-z.

[4] J. Witten *et al.*, "Artificial intelligence-guided design of lipid nanoparticles for pulmonary gene therapy," *Nat. Biotechnol.*, pp. 1–10, Dec. 2024, doi: 10.1038/s41587-024-02490-y.

Technical details

Models and datasets were obtained for AGILE and LiON from the respective public Github repositories. Datasets were harmonized and inconsistencies corrected.

The AGILE dataset consists of 1,200 samples with ionizable lipid structure and in vitro expression in HeLa cells. The 83 samples with expression values above 9.0 were classified as high performers (6.9%).

The Lion dataset consists of 9,637 samples s from 27 individual in vivo and in vitro studies, with LNP composition, formulation, information on the experiment and "delivery" value. The 1,177 samples from the dataset with Library_ID "RM_Michael_addition_branched" and Experiment_ID "A549_form_screen" were chosen for the analysis. The 152 samples with "delivery" above 5.0 were classified as high performers (12.9%).

Model performance was evaluated using 10-fold cross-validation (CV), repeated ten times with different random seeds to ensure robustness. Results shown in the figures are aggregated from the resulting 100 different model predictions. Identical splits into training, validation, and test sets were used for the respective models in both comparisons.

For the PPV–TPR curves, PPV and TPR were evaluated from model predictions on each test set for varying classification threshold. The graph shows the average PPV values, after interpolation of the individual curves on equidistant TPR values.

For the PPV and TPR bar graph, the classification threshold was determined for each trained model, so that prevalence and predicted prevalence coincide for the whole dataset. Using the respective threshold, samples from each test set were classified. For each of the ten CV runs, the predictions of all test folds were combined and used to calculate PPV and TPR. The graph shows the average PPV and TPR values for all ten CV runs.