nature portfolio

Peer Review File

Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This research manuscript presents a comprehensive and ambitious effort to leverage AI algorithms for the design and virtual screening of novel ionizable lipids tailored for mRNA-LNP delivery. The study's approach and findings are quite innovative and potentially impactful in the field of mRNA delivery systems. However, there are several critical questions and aspects that could be further addressed or clarified:

Page 10 and Fig. 3e) It's unfortunate that these results cannot be displayed due to the patent application. "Briefly, the mRNA delivery efficiency of 14 ionizable lipids (structures not shown due to the IP concerns) were predicted and subsequently compared to experimental data (Fig. 3e)".

Fig. 3 and Fig. 5) For an accurate determination of mRNA delivery using different LNPs (LNPs containing various ionizable lipids), the amount of luciferase mRNA, or the luc. assay, should be determined per organ (extract the 6-7 major organs including liver, heart, lung, etc., and then carefully determine the amount of delivered mRNA per organ).

Page 15 and Fig. 4D) Explain the criteria for choosing exactly these six lipids to synthesize out of the 21 lipids that had predicted better mRNA delivery. Moreover, A better explanation of Fig. 4a is needed.

Fig. 5) In Fig. 5b, it is not stated which statistical analyses are performed to compare mRNA delivery using LNPs containing the different lipids. The difference between MC3 and especially the novel lipids LQ091 (also LQ089) is quite small. Therefore, it may be incorrect to describe on page 17 that ' Impressively, all new lipids generally outperformed MC3 in terms of mRNA delivery efficiency, among which LQ089 and LQ091 exhibited superior performance'.
Moreover, regarding the results of Fig. 5, you also need to show the images of the mice that are analyzed for total luminescence using the spectrum imaging system. You can add them among the supplemental Figs.

Long-term stability and safety of the predicted lipids:
The study focuses on the efficacy of mRNA delivery, but how do these novel lipids manage in terms of long-term stability and safety? For example, by measuring the levels of inflammatory cytokines in the blood following the delivery of mRNA with various LNPs (LNPs containing different lipids), you can find out the most effective and safest lipid for mRNA delivery.
Extended Fig. 3 and 4) Information on how the ranking system for the ionizable lipid heads and tails is missing. How the ranking is performed and how it can be generalized for all ionizable lipids on the market.

In the discussion part, these issues should be addressed:

1. Dataset limitations and generalizability: The study acknowledges the modest size and diversity of the dataset used. How might this limitation affect the generalizability of the AI models to a broader range of

ionizable lipids? Would the models be equally effective in predicting the efficacy of structurally diverse lipids not represented in the training dataset?

2. Mechanistic Insights: The manuscript highlights that AI modeling advanced the screening process but did not elucidate the mechanistic structure-activity relationships within ionizable lipids. Could additional computational methods or experimental designs be implemented to gain more insight into the structural features that contribute to the efficacy of the screened lipids? For example, it has been shown previously that mRNA molecules that are neutrally charged can pass over the endosomal membrane and thus end up in the cytosol for translation (PMID: 31551417).

3. Potential bias in model training: Was there any potential bias in the training process of the AI models, especially considering the limitations of the dataset?

4. Impact of lipid structure on specificity and versatility: How does the structure of these novel lipids influence their specificity for different types of mRNA, e.g. short and long mRNAs? Do they work equally efficiently (do endosomal escape equally well) for all types of mRNA to be translated?

Addressing these questions would strengthen the study, providing a more robust framework for the application of AI in the rational design of ionizable lipids for mRNA delivery.

Reviewer #2 (Remarks to the Author):

Review: Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery

In this report, Wang et al. describe an AI materials discovery platform for screening ionizable lipids as mRNA delivery agents in lipid nanoparticles (LNPs). The authors (1) ingest literature structural data from mRNA LNP studies and patents to build and validate apparent pKa and mRNA delivery efficient models, (2) screen ~20 million virtual structures using the models to down-select 3 lipids to synthesize and test in mice, and (3) optimize the lipid head/tail features in another round of screening/validation with 666 candidates before proceeding to generate 6 lipids for testing. The in vivo performance was evaluated with whole-body luminescence of the animal, which was compared to positive controls MC3 and SM-102. The authors conclude by asserting that the novel lipids developed through this process outperform MC3 and SM-102 in mRNA delivery, thereby offering immense potential for further use to advance nonviral gene delivery as a whole.

Overall, while this work clearly has taken considerable time and effort to develop, there are significant concerns that I have with the data collection and cleaning process, choice of AI algorithms used to evaluate datasets to make machine versus human decisions, and synthesis/testing of the lead lipid compounds. Each of these sections are described with points below for the authors to consider. Although the topics of AI and LNPs are clearly relevant to not only the scientific community but also the general public, there are serious scientific shortcomings in the current version of the manuscript that need to be corrected so that the conclusions and claims of the work are supported by the data. At this

point, it is unconvincing that any of this can be reproduced or used to inform future LNP design. If structures cannot be shown "due to IP concerns," then this work cannot be published in an academic journal. There are some general conclusions of what lipid features may be important for mRNA delivery, but I am left unclear of what structure-property relationships are actually realized, or how to actually apply this to translation of therapeutic effects for different genetic diseases. In terms of bioperformance, the end result of total body luminescence as a readout of mRNA delivery is also underwhelming, with no evidence of organ-specific delivery or long-term evaluation of toxicity. SM-102 was a lipid developed and optimized for addressing the COVID19 pandemic, so outcompeting it in a completely unrelated luminescence assay with a model mRNA is not that impactful. Unfortunately, this manuscript should not be published by either Nature Communications or other journal until these fundamental flaws are sufficiently addressed by the research team conducting this work.

- The authors claim that the training data were "meticulously gathered from literature sources and patents" with some structures shown in Fig S-1. This is vague and unclear… were values manually copied over by humans? Or was data scraped using some automated tools? Fig S-1 only shows 6 structures with a caption that is inappropriate for scientific publication ("Some example ionizable lipids collected in the dataset. Some of them have commercial name"). The data and code availability statement by the authors suggest that they have no intention of adhering to the FAIR Principles that has been championed by those working in the AI space (see "The FAIR Guiding Principles for scientific data management and stewardship" by Wilkinson et al. 2016).
- The quality of the data is critical—critical, to the efficacy of AI models in their predictions. If the foundational data is not made available to reviewers or to the readers, the value of the predictions (even if they produce reasonable R squared values) is not meaningful. There is no description of how the training data was curated—that is, how was it decided whether a lipid structure used for mRNA delivery was to be included or not in training the models? Or if the biological readout of delivery was consistent? It is well recognized that mixed datasets and quantification methodologies jeopardize their utility in machine learning, even leading to misleading conclusions. See Corral et al. "Quantifying nanoparticle delivery: challenges, tools, and advances" 2024 as a comprehensive example that argues doing this blindly is akin to comparing apples to oranges.
- The authors describe a method to calculate the "apparent pKa" of lipids. What is the difference between this and the actual pKa of a molecule? Was any work done to evaluate how closely apparent pKa values are to actual analytical measurements done, e.g., titrations?
- The first model uses only ECFP, so it is unclear why SHAP was used since it does not offer any additional information qualitatively or quantitatively to materials design. Each component of ECFP cannot be mapped directly to something on the molecule.
- Furthermore, the regression model for the apparent pKa prediction uses this value as a proxy for classification, assuming efficiency will be good within a certain range set by the authors. By doing this, prediction of virtual molecules filters out candidates before efficiency and synthesizability are even considered. Doesn't this bias the AI model?
- The authors do not sufficiently describe the down-selection process for the first round of lipid synthesis. They state that synthesizability and diversity are important to consider, but no details are provided on how likely hits were whittled down to 3 for synthesis.
- Was any structural similarly analysis done on the virtual lipids? It is unclear how the authors define "diversity."

- In the second round of virtual screening, the authors intentionally lock in the head group as a synthesis capability factor before down-selection. This reportedly shrank the pool of lipids considerably. The subsequent discussion of how each model performs is extremely confusing. Even upon rereading several times, I cannot explain how the selection process resulted to the final six compounds.
- I find it challenging to believe that N/P or weight ratio of mRNA (processing conditions) has negligible effect on the biological readout of delivery for LNPs. Processing LNPs is crucial to their delivery efficacy, all else equal, especially for in vivo settings, serving as a competitive advantage for biotech and pharmaceutical companies that must be reported once they reach pre-clinical and clinical studies. Nearly any LNP review emphasizes this point. Right away, this puts the quality of the dataset used to train both models into question. If they are not available in the majority of the original sources, in my view they should not be used to train any AI model.
- Why are the images of the total luminescence in mice not shown for Round 2? This needs to be in the Supporting Information at minimum.
- Several of the references are missing volumes and page numbers.
- Finally, in both the main manuscript and the supporting information, no materials characterization is provided on the synthesized lipids other than a protocol and a yield. This is unacceptable. Any scientist cannot claim to have made novel molecules for use in something as important as in vivo applications without definitively showing that compound is pure and actually follows literature synthesis procedures. Millions of hypothetical structures generated from machines are useless if they cannot be synthesized by material chemists in the lab.

Reviewer #4 (Remarks to the Author):

This is an interesting and timely study and could add new knowledge to the field and be of value to many researchers in the field. There are some comments the authors should address:
1. Ionisable lipids for vaccines will be different for those for therapy, give the balance of immunogenicity. I think the authors should consider this in their discussion.
2. The model looks at luciferase and hEPO, this then focuses on therapeutics rather than vaccines. Protein expression levels would not be the only factor in vaccine efficacy.
3. Did the model consider the type and biological sex of the mice? This may impact on the results.
4. There is a heavy emphasis on pKa, and this is well known that around 6 to 6.8 is the best range. Does the model consider structure of the lipids, and configuration within the LNPs. This is more important than the pKa.
5. Figure 5. The area under the curve ploy is not particularly useful for a 3 point graph and when looking at it, I could not follow then I spotted the colour coding does not match. This should be corrected and the authors should consider again if AUC is the best way to look at this.
6. Figure 5. The number of mice used is low and looks to be in a single study. Given the closeness of the data and the variability in mouse studies, the study should be replicated again. Even with low numbers (e.g. 2 mice per group).
7. Was a ‘negative’ control tested e.g. a lipid that should only give low responses tried to validate the model. This should be added in to the replicate study.
8. How do the authors consider the lack of data translation from mouse to NHPs to humans, this is a recognised problem. The authors should comment.

9. I would question the validity of using both IV and IM data as these are very different data sets and IV is thought to rely on ApOE binding and then delivery to the liver whilst IM is thought to involve resident immune cells at the injection site or infiltrating immune cells to the injection site.
10. The discussion should add more into how the reader can gain from this model so a wide audience can apply this learning. This would add more significance to the work and the paper.

Reply to comments for Manuscript "Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery"

First, we would like to express our sincere thanks to the reviewers for the constructive comments and suggestions on our manuscript "Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery" (Research Article, **NCOMMS-23-53992**). We have addressed these comments with corresponding changes and answers. The point-by-point responses to the referees were listed below this letter. We earnestly appreciate editors/reviewers' dedicated work and hope that the revisions address all concerns. We believe that we have improved the quality of the manuscript and we are submitting our revised version for publication in the journal "Nature Communications".

Reviewer #1 (Remarks to the Author):

This research manuscript presents a comprehensive and ambitious effort to leverage AI algorithms for the design and virtual screening of novel ionizable lipids tailored for mRNA-LNP delivery. The study's approach and findings are quite innovative and potentially impactful in the field of mRNA delivery systems. However, there are several critical questions and aspects that could be further addressed or clarified:

Page 10 and Fig. 3e) It's unfortunate that these results cannot be displayed due to the patent application. "Briefly, the mRNA delivery efficiency of 14 ionizable lipids (structures not shown due to the IP concerns) were predicted and subsequently compared to experimental data (Fig. 3e)".

Answer: We sincerely thank you for your approval and helpful suggestions. Structures of the 14 ionizable lipids have been added to the supplementary materials as Supplementary Fig. 1.

Fig. 3 and Fig. 5) For an accurate determination of mRNA delivery using different LNPs (LNPs containing various ionizable lipids), the amount of luciferase mRNA, or the luc. assay, should be determined per organ (extract the 6-7 major organs including liver, heart, lung, etc., and then carefully determine the amount of delivered mRNA per organ).

Answer: We appreciate for your constructive advice. To determine mRNA delivery, the curve of luminescence over time and AUC thereof can monitor the expression of mRNA within a long period to describe the long-term performance of LNP, however, this method does miss the organ distribution of mRNA expression at the highest point. In our revised manuscript, we supplemented the organ distribution of luciferase expressed by LNP-loaded mRNA administered both intravenously and intramuscularly (Fig. 6 and 7). Luciferase mRNA encapsulated by the new LNPs via intravenous injection was

mainly expressed in the liver, while the intramuscularly injected one was mainly expressed at the injection site of leg.

Page 15 and Fig. 4D) Explain the criteria for choosing exactly these six lipids to synthesize out of the 21 lipids that had predicted better mRNA delivery. Moreover, A better explanation of Fig. 4a is needed.

Answer: Thanks for your reminder. We aimed to select lipids with more diversity in their tails, including those with two long branches, dendric branches, and cyclohexyl group. The AI model outputs the probability of a LNP being positive in mRNA delivery efficiency, and lipids with high probability were considered. The probabilities were listed in Supplementary Table 6. The 21 lipids are now supplemented in Supplementary Fig. 8.

The following statement was added to the caption of Fig. 4a, *"Data representation of Model 2 predicting mRNA delivery efficiency. Compared to the representation method of Model 1, the positive criterion was set as 2-fold the delivery efficiency of the standard MC3 LNP"*.

Fig. 5) In Fig. 5b, it is not stated which statistical analyses are performed to compare mRNA delivery using LNPs containing the different lipids. The difference between MC3 and especially the novel lipids LQ091 (also LQ089) is quite small. Therefore, it may be incorrect to describe on page 17 that ' Impressively, all new lipids generally outperformed MC3 in terms of mRNA delivery efficiency, among which LQ089 and LQ091 exhibited superior performance'.

Answer: Thank you for your careful reminder. The Fig. 5 has been merged into Fig. 4. Statistical significance in our study was analyzed by one-way ANOVA (GraphPad Prism 9). We have added this information to the caption of Fig. 4 and other figures demanding.

We have changed the sentence to *"Impressively, all new lipids are well-performed in terms of mRNA delivery efficiency, among which LQ089, LQ091-LQ093 exhibited significantly higher efficacy than MC3"*.

Moreover, regarding the results of Fig. 5, you also need to show the images of the mice that are analyzed for total luminescence using the spectrum imaging system. You can add them among the supplemental Figs.

Answer: Thanks for your reminding. The Fig. 5 has been merged into Fig. 4. The bioluminescence images supporting Fig. 4 were listed in Supplementary Fig. 6 of the revised version.

Long-term stability and safety of the predicted lipids:

The study focuses on the efficacy of mRNA delivery, but how do these novel lipids manage in terms of long-term stability and safety? For example, by measuring the levels of inflammatory cytokines in the blood following the delivery of mRNA with various LNPs (LNPs containing different lipids), you can find out the most effective and safest lipid for mRNA delivery.

Answer: Thanks for your comments. As your suggestions, safety, effectiveness, and stable and controllable quality are three important attributes of medicines. For LNPs containing new ionizable lipids, the safety and long-term stability are as important as delivery efficiency. Following up your helpful suggestion, we supplemented long-term storage stability tests and acute toxicity tests on the LNPs formed of the newly generated lipids. The freezing-thawing stability of LNPs at -20℃ and -80℃, and the stability of LNPs stored at 4℃, -20℃ and -80℃ for 30 days, were evaluated for pharmaceutical properties and in vivo efficiency of mRNA expression (Fig. 8). The results showed that the leading lipids, LQ089 and LQ092, exhibited comparable pharmaceutical stability to MC3 and SM-102. Besides, we performed an acute toxicity test on the leading lipids. The LNPs containing new lipids showed good safety in mice at an acute dose of 5 mg/kg (mRNA) of which the lipid dose was about 75 mg/kg (Fig. 9 and 10).

Supplementary Fig. 3 and 4) Information on how the ranking system for the ionizable lipid heads and tails is missing. How the ranking is performed and how it can be generalized for all ionizable lipids on the market.

Answer: Thanks for your comments. Supplementary Fig. 3 and 4 now are numbered as Supplementary Fig. 4 and 5. The tails and heads were ranked based on their "positive rate". How the ranking system is performed has been added to the method section:

"In virtual screening, the general performance of a candidate tail or head segment was judged by positive rate. A generated virtual ionizable lipid would be marked as positive if its predicted mRNA delivery efficiency was better than the standard MC3 LNP and apparent pKa was between 6.0 to 7.0. Since the lipids were generated by combining different head and tail segments, therefore, for each segment, the positive rate can be defined as:

$$Positive\ rate\ of\ segment = \frac{Number\ of\ postive\ lipids\ containing\ the\ segment}{Number\ of\ lipids\ containing\ the\ segment} \quad (1)$$

Segments with high positive rate means they are more compatible in ionizable lipids to increase the LNP performance."

The positive rate for a segment is calculated based on a large number of combinations between it with other various parts of lipids, thus the positive rate is a general indication of the segment quality. This ranking system is applied to lipid segments, not to whole ionizable lipids. For lipid molecule, we possibly use probability of being positive in mRNA delivery efficiency or ECFP score to quantify their qualities.

In the discussion part, these issues should be addressed:

1. Dataset limitations and generalizability: The study acknowledges the modest size and diversity of the dataset used. How might this limitation affect the generalizability of the AI models to a broader range of ionizable lipids? Would the models be equally effective in predicting the efficacy of structurally diverse lipids not represented in the training dataset?

Answer: Thanks for your comments. AI models would perform better for molecules closer to those in the training dataset, exemplified as LQ085-087 and LQ089-094. In other two AI application studies, new designed lipids are still like the training data. We think strategy like pre-training model with large molecule database, representing molecules with more refined model, or mechanistically modeling like molecular dynamic (MD) simulation may help in alleviating this problem. This has been added to the discussion:

"--- generalizing the model to a broader formulation design space is challenging, like novel lipid structures and different mRNA sequences. In this work, predictions on LQ089-094 are more accurate than those on LQ085-087, and the former are closer to the majority of lipids in the dataset. In other AI modeling work, the newly designed molecules are similar to their training data[27,54]. This limitation might be alleviated through data augmentation, introducing more diverse data, or adopting a pre-train and fine-tune model building workflow. Besides, mechanistic modeling is a promising way to break through the generalization limitation, such as molecular dynamic (MD) simulation. The simulated LNP and the interaction between RNA and lipids have been reported many times[26,55–57], with no limitation to ionizable lipid structure."

2. Mechanistic Insights: The manuscript highlights that AI modeling advanced the screening process but did not elucidate the mechanistic structure-activity relationships within ionizable lipids. Could additional computational methods or experimental designs be implemented to gain more insight into the structural features that contribute to the efficacy of the screened lipids? For example, it has been shown previously that mRNA molecules that are neutrally charged can pass over the endosomal membrane and thus end up in the cytosol for translation (PMID: 31551417).

Answer: Thanks for your comments. We added structure-activity relationship analysis as Supplementary Fig. 3 (correlation between ECFP score and mRNA delivery efficiency) and Fig. 5 (effect of change in tail structure on the performance of lipids).

As mentioned above, we think MD simulation is a powerful tool to elucidate interaction between lipids and mRNA. RNA escape from endosomal membranes is still the interaction between lipids and nucleic acids by nature. MD can be used to construct any lipids and mRNA. Our lab is devoted to the research of molecular dynamics (MD) in the modeling of lipid nanoparticles (LNPs). We have established an LNP model that is close to the actual size (~60 nm), and the relevant manuscript is being edited and will be submitted in soon. Based on this model, we can continue to study the interaction between LNPs and endosomes.

Besides, we think that contribution of ionizable lipids to the process like RNA escape from endosomal can be quantified and clearly described by physiologically-based pharmacokinetics (PBPK) and quantitative systems pharmacology (QSP) models. Both of them are based on ordinary differentiation equations and are mechanistic model. It takes the advantage of integrating various data measured in vivo and in vitro. Actually, we have developed a such model for RNA-LNP transportation in Hela cells. The rates of trafficking along the processes of endocytosis, endosomal maturation, excretion outside cells, release from endosomes were determined. The calculated rate of RNA released from endosomes therefore is pure result of interaction between LNP and endosomal membranes, excluding other confounding physiological factors. This model also predicted the fraction of RNA released into the cytoplasm as around 2 percent. This work has been accepted by Acta Pharmaceutica Sinica B, entitled as “Modeling on in vivo disposition and cellular transportation of RNA lipid nanoparticles via quantum mechanics/physiologically-based pharmacokinetic approaches”.

In the last paragraph of discussion, we added:

“--- other advanced modeling methods such as physiologically-based pharmacokinetics (PBPK) and quantitative systems pharmacology (QSP) models[59–62] *are very useful. PBPK is specialized in inferring the fate of drugs across different species. This inference is based on the properties of the drug and the physiological conditions of the subject, and therefore such extrapolation is mechanistically based. QSP is also mechanistic, predicting dynamic changes in signal pathways, biomarkers, and even therapeutic effects. For a complex system such as immune response, QSP is promising to address it*[63,64]*. Further, the association of the two models can integrate various in vitro and in vivo data, being able to quantify rates of critical processes in nucleic acid delivery such as RNA escape from endosomes*[65]*”*

3. Potential bias in model training: Was there any potential bias in the training process of the AI models, especially considering the limitations of the dataset?

Answer: Thank you for your comments. The small data size can indeed introduce bias in models by restricting the structural diversity of lipids. Additionally, the model has been trained exclusively with data from mice and intravenous administration. Consequently, these factors make the model biased to be more compatible with this

certain situation but it could exhibit reduced predictive accuracy for different scenarios.

4. Impact of lipid structure on specificity and versatility: How does the structure of these novel lipids influence their specificity for different types of mRNA, e.g. short and long mRNAs? Do they work equally efficiently (do endosomal escape equally well) for all types of mRNA to be translated?

Answer: Thanks for your comments. Thank you for your advice. You are proposing a very interesting and forward-looking direction. Previous studies in mRNA delivery have paid little attention to the special requirements of different mRNA types for ionizable lipids. Recently, a study focused on the influence of lipid tail length on delivery efficiency of mRNA of varying length with C12-200 as a model lipid, and concluded that shorter tails of lipids might lead to higher transfection of LNPs encapsulating larger mRNAs, and that longer tails might be more efficient for smaller mRNA cargos (Mrksich, K. et al. J Biomedical Materials Res jbm.a.37705 (2024)).

In our task, the coding sequence length of the luciferase mRNA is approximately 1600, while the length for hEPO is around 600. The mRNA wrapped in LNP is longer than the coding sequence by around 270 nt (https://www.apexbt.com/ez-captm-cy5-firefly-luciferase-mrna-5-moutp.html#gfdescription, https://www.apexbt.com/ez-captm-epo-mrna-psutp.html). Few lipids in our dataset were tested in both mRNA. A general trend is that if a lipid shows better delivery efficiency for luciferase mRNA, it will also achieve better performance for hEPO. Besides, it seems that LNP loading hEPO mRNA is more likely to get better delivery efficiency than MC3. However, this is a possibly biased conclusion due to the sample size. The following content has been added to the discussion.

"MD simulation should also facilitate the understanding of the lipid specificity to different mRNA sequences. In our work, data of luciferase and hEPO mRNA were merged, but only nearly 10 lipids were tested using both mRNA. The delivery efficiencies for the two mRNA show a consistent trend, but using hEPO seems to be more likely to obtain positive results."

Besides, as mentioned above, PBPK model can be used to judge the efficacy in critical process such as endosomal escape. In our accepted work, we compared the RNA release rate of MC3, L319, and C12-200.

Addressing these questions would strengthen the study, providing a more robust framework for the application of AI in the rational design of ionizable lipids for mRNA delivery.

Answer: Thanks again for your wonderful comments and helpful suggestions, which are of great significance to the improvement of our manuscript.

Reviewer #2 (Remarks to the Author):

Review: Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery

In this report, Wang et al. describe an AI materials discovery platform for screening ionizable lipids as mRNA delivery agents in lipid nanoparticles (LNPs). The authors (1) ingest literature structural data from mRNA LNP studies and patents to build and validate apparent pKa and mRNA delivery efficient models, (2) screen ~20 million virtual structures using the models to down-select 3 lipids to synthesize and test in mice, and (3) optimize the lipid head/tail features in another round of screening/validation with 666 candidates before proceeding to generate 6 lipids for testing. The in vivo performance was evaluated with whole-body luminescence of the animal, which was compared to positive controls MC3 and SM-102. The authors conclude by asserting that the novel lipids developed through this process outperform MC3 and SM-102 in mRNA delivery, thereby offering immense potential for further use to advance nonviral gene delivery as a whole.

Overall, while this work clearly has taken considerable time and effort to develop, there are significant concerns that I have with the data collection and cleaning process, choice of AI algorithms used to evaluate datasets to make machine versus human decisions, and synthesis/testing of the lead lipid compounds. Each of these sections are described with points below for the authors to consider. Although the topics of AI and LNPs are clearly relevant to not only the scientific community but also the general public, there are serious scientific shortcomings in the current version of the manuscript that need to be corrected so that the conclusions and claims of the work are supported by the data. At this point, it is unconvincing that any of this can be reproduced or used to inform future LNP design. If structures cannot be shown "due to IP concerns," then this work cannot be published in an academic journal. There are some general conclusions of what lipid features may be important for mRNA delivery, but I am left unclear of what structure-property relationships are actually realized, or how to actually apply this to translation of therapeutic effects for different genetic diseases. In terms of bioperformance, the end result of total body luminescence as a readout of mRNA delivery is also underwhelming, with no evidence of organ-specific delivery or long-term evaluation of toxicity. SM-102 was a lipid developed and optimized for addressing the COVID19 pandemic, so outcompeting it in a completely unrelated luminescence assay with a model mRNA is not that impactful. Unfortunately, this manuscript should not be published by either Nature Communications or other journal until these fundamental flaws are sufficiently addressed by the research team conducting this work.

Answer: Thanks for your comments. These help a lot to improve our article.

In the method section, we described how the data was cleaned to improve its intrinsic homogeneity. Specifically, it is:

"---(1) removed data that was not measured in mice; (2) removed data where the LNP was not administrated intravenously; (3) removed data of mRNA expression level which was not measured as the luminescence signal or concentration of the luciferase or the human erythropoietin (hEPO) induced by mRNA delivery; (4) removed data of the luminescence signal of luciferase that was not measured for whole-body of subject animals or livers; (5) maintained the data where the mRNA expression levels of LNPs could be transformed as the fold-change based on a standard LNP formulation. The standard LNP formulation was composed of MC3 (the ionizable lipid), DSPC (the helper lipid), cholesterol, and PEG2000-DMG (the PEG lipid) at the molar ratio of 50/10/38.5/1.5, which is commonly used as the control since it is the LNP formulation of the first approved siRNA drug[67]. The standard expression level of this formulation included: (1) luciferase concentration at 198 ng/g liver tissue at 4 h after administration of 0.3 mg/kg mRNA[38]; (2) luciferase luminescence flux at 2.57E+9 p/s in livers at 6 h after administration of 0.5 mg/kg mRNA (for data from the institution of Moderna) [35]; (3) luciferase luminescence flux at 8.66E+8 p/s in the whole-body at 6 h after administration of 0.5 mg/kg mRNA (for data from the institution of Tufts University)[31]; (4) plasma hEPO concentration at 1570, 1830, 810 ng/mL at 3, 6, 24 h respectively, after administration of 0.5 mg/kg mRNA[35]. The value of concentrations of expressed proteins was comparable among different institutions, while the value of luminescence flux was not since the measurement of the flux is the signal after amplification via the photomultiplier, which is dependent on the experimental instrument of the institute."

We have added the structures of the 14 lipids in the external validation related to "IP concerns" to the revised manuscript as Supplementary Fig. 1.

We have enhanced the examination of the correlation between the efficiency of mRNA delivery and the ECFP code of ionizable lipids. Additionally, we offer a separate structure-activity analysis for a specific class of lipids that have hydroxy groups in their heads. We demonstrate the impact of factors such as chain length and linker position on the performance of the tails. Our aim is that these findings will advance the design of ionizable lipids. While the application to specific diseases is not within the purview of this research, our current goal is to improve the screening process of ionizable lipids, primarily using the most commonly reported mRNA.

We have provided result of organ-specific delivery (Supplementary Fig. 9, 10, and 12) and toxicity (Fig. 9 and 10, Supplementary Fig. 13, 14, 15, and 16) data to the revised manuscript.

Although SM-102 is used in COVID-19 vaccine, it truly exceeds MC3 in delivering hEPO and luciferase mRNA. Our goal is to obtain new lipids with high mRNA delivery efficiency, therefore, we tested whether our lipids outperform not only the baseline MC3 but also the superior SM-102.

- The authors claim that the training data were “meticulously gathered from literature sources and patents” with some structures shown in Fig S-1. This is vague and unclear… were values manually copied over by humans? Or was data scraped using some automated tools? Fig S-1 only shows 6 structures with a caption that is inappropriate for scientific publication (“Some example ionizable lipids collected in the dataset. Some of them have commercial name”). The data and code availability statement by the authors suggest that they have no intention of adhering to the FAIR Principles that has been championed by those working in the AI space (see “The FAIR Guiding Principles for scientific data management and stewardship” by Wilkinson et al. 2016).

Answer: Thank you for your reminding. Structures of ionizable lipids were extracted from resources with InDraw AI chemical structure recognizer and transformed to SMILES string. Other data was extracted by manually copying and checked. These has been added to the method section.

We have re-edited the Figure S1 as Supplementary Table 1 with a clearer format now.

We have provided our training data and model file for review. We will make our models publicly available as our FormulationAI platform (https://formulationai.computpharm.org/. Dong, J., et al. FormulationAI: a novel web-based platform for drug formulation design driven by artificial intelligence. Briefings in Bioinformatics 25, bbad419 (2024).)

- The quality of the data is critical—critical, to the efficacy of AI models in their predictions. If the foundational data is not made available to reviewers or to the readers, the value of the predictions (even if they produce reasonable R squared values) is not meaningful. There is no description of how the training data was curated—that is, how was it decided whether a lipid structure used for mRNA delivery was to be included or not in training the models? Or if the biological readout of delivery was consistent? It is well recognized that mixed datasets and quantification methodologies jeopardize their utility in machine learning, even leading to misleading conclusions. See Corral et al. “Quantifying nanoparticle delivery: challenges, tools, and advances” 2024 as a comprehensive example that argues doing this blindly is akin to comparing apples to oranges.

Answer: Thanks for your careful reminding. We have provided our training data and model file for review. We will make our models publicly available as our FormulationAI platform (https://formulationai.computpharm.org/. Dong, J., et al. FormulationAI: a novel web-based platform for drug formulation design driven by artificial intelligence. Briefings in Bioinformatics 25, bbad419 (2024).)

The article published by See Corral et al introduces three methods to represent the

delivery efficiency of nanoparticle, named functional readouts, nanocarrier tracking, and cargo tracking. In our data, the mRNA delivery efficiency was measured through functional-readout method, since the signal is sourced from the proteins translated from the mRNA delivered. The two mRNA, luciferase and hEPO are mainly expressed in liver because it is the main target of current LNPs. Trends in expression of the two mRNA should be consistent since they share the same mechanism of mRNA release and translation. We especially care about the consistence in data and whether data is comparable between different studies. For hEPO, it is the expressed protein concentration in blood that was measured. This is a straightforward method measuring mRNA expression, therefore this readout is comparable. For luciferase, if it is protein concentration that was measured, it is also comparable; otherwise, if luminescence was measured, as mentioned in See Corral et al, it is influenced by many other factors, thus, only data with comparison to standard control LNP was maintained. In other words, luminescence measured from different institutions are not comparable. In the method section, we mentioned the data curation process in detail as:

"For the analysis of nanomedicine from multiple data sources, ensuring internal consistency within the data is crucial[66]. For AI models predicting the in vivo mRNA delivery efficiency of LNPs, subsequent data processing work was conducted to improve its internal consistence and maintain as large data as possible: (1) removed data that was not measured in mice; (2) removed data where the LNP was not administrated intravenously; (3) removed data of mRNA expression level which was not measured as the luminescence signal or concentration of the luciferase or the human erythropoietin (hEPO) induced by mRNA delivery; (4) removed data of the luminescence signal of luciferase that was not measured for whole-body of subject animals or livers; (5) maintained the data where the mRNA expression levels of LNPs could be transformed as the fold-change based on a standard LNP formulation. The standard LNP formulation was composed of MC3 (the ionizable lipid), DSPC (the helper lipid), cholesterol, and PEG2000-DMG (the PEG lipid) at the molar ratio of 50/10/38.5/1.5, which is commonly used as the control since it is the LNP formulation of the first approved siRNA drug[67]. The standard expression level of this formulation included: (1) luciferase concentration at 198 ng/g liver tissue at 4 h after administration of 0.3 mg/kg mRNA[38]; (2) luciferase luminescence flux at 2.57E+9 p/s in livers at 6 h after administration of 0.5 mg/kg mRNA (for data from the institution of Moderna) [35]; (3) luciferase luminescence flux at 8.66E+8 p/s in the whole-body at 6 h after administration of 0.5 mg/kg mRNA (for data from the institution of Tufts University)[31]; (4) plasma hEPO concentration at 1570, 1830, 810 ng/mL at 3, 6, 24 h respectively, after administration of 0.5 mg/kg mRNA[35]. The value of concentrations of expressed proteins was comparable among different institutions, while the value of luminescence flux was not since the measurement of the flux is the signal after amplification via the photomultiplier, which is dependent on the experimental instrument of the institute."

We used this process trying our best to make the data uniform. For pKa model, no

special curation process was applied. We utilized all collected pKa data.

- The authors describe a method to calculate the "apparent pKa" of lipids. What is the difference between this and the actual pKa of a molecule? Was any work done to evaluate how closely apparent pKa values are to actual analytical measurements done, e.g., titrations?

Answer: Thanks for your comments. The apparent pka is the pKa of a lipid nanoparticle, measured by TNS method. The actual pKa of a molecule is the pKa of ionizable lipid itself, which is usually calculated theoretically (*Carrasco, M. J. et al. Ionization and structural properties of mRNA lipid nanoparticles influence expression in intramuscular and intravascular administration. Commun Biol 4, 1–15 (2021); Eygeris, Y., Gupta, M., Kim, J. & Sahay, G. Chemistry of Lipid Nanoparticles for RNA Delivery. Acc Chem Res 55, 2–12 (2022).*). It is found that the actual pKa is 2-3 units higher than the apparent pKa when it is incorporated into LNP.

Since the actual measurement such as titration method requires sample solved in water, it is not applicable to ionizable lipid molecules. A compromise approach is to prepare a water-soluble compound with the same theoretical pKa, and then measure the pKa in experiments. (*Carrasco, M. J. et al. Ionization and structural properties of mRNA lipid nanoparticles influence expression in intramuscular and intravascular administration. Commun Biol 4, 1–15 (2021)).* If use NMR to measure the chemical shift of proton near ionizable group in different pH, the resulted pKa is close to the theoretically calculated one.

- The first model uses only ECFP, so it is unclear why SHAP was used since it does not offer any additional information qualitatively or quantitatively to materials design. Each component of ECFP cannot be mapped directly to something on the molecule.

Answer: Thanks for your comments. The first model uses ionizable lipid ECFP and composition information of LNP. SHAP here was used to attribute the model prediction to the contribution of each feature, including all ECFP bits, to quantify the feature importance and effect. In different molecules, the same substructure leads to the same bit set as 1. For example, in the following two molecules, three substructures -CN(C)C-, -O-, and -CCC=C- correspond to the same number of bit (652, 695, and 731), which set the three bits in both two molecules as 1. This implies that ECFP explicitly include substructure information though some confliction in bit folding occasionally happens.

Using SHAP values for ECFP, we could identify the direction and extent of contribution of each bit, which helped us to calculate the ECFP score for ionizable lipids. We found that the ECFP score was positively correlated to the mRNA delivery efficiency ($r > 0.6$). This was a surprise since the model had not accepted any quantified delivery efficiency data when training. Since the outcome is very convincing, we think the use of ECFP and SHAP is justified. This content was added to the revised manuscript in sections of result and method (Equation 5 and 6).

- Furthermore, the regression model for the apparent pKa prediction uses this value as a proxy for classification, assuming efficiency will be good within a certain range set by the authors. By doing this, prediction of virtual molecules filters out candidates before efficiency and synthesizability are even considered. Doesn't this bias the AI model?

Answer: Thanks for your comments. The prediction of apparent pKa was independent of and in parallel with mRNA delivery efficiency. In this view, prediction of pKa does not bias the AI model. We set the criterion of "good" lipids as within a pKa range between 6-7, and with positive mRNA delivery efficiency just during the virtual screening stage, in which the AI model was fixed and not changed.

We used two conditions to screen lipids because this can lower the risk of picking

falsely positive candidates than just using single model. The pKa has been proven to significantly affect the delivery efficiency. Our pKa range was enlightened by the collected data. In fact, this pKa range is broader than others reported in articles (they have been cited in the discussion part), because we wanted to search good candidate from a larger space.

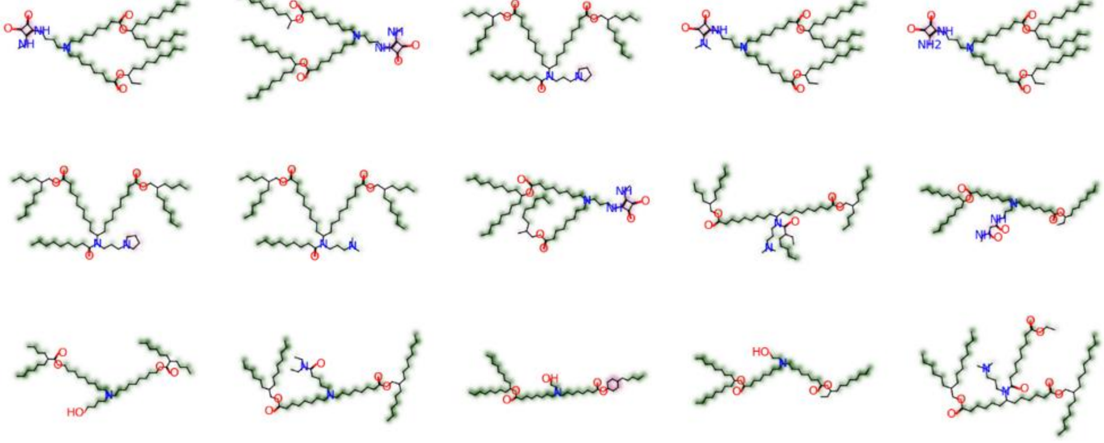
- The authors do not sufficiently describe the down-selection process for the first round of lipid synthesis. They state that synthesizability and diversity are important to consider, but no details are provided on how likely hits were whittled down to 3 for synthesis.

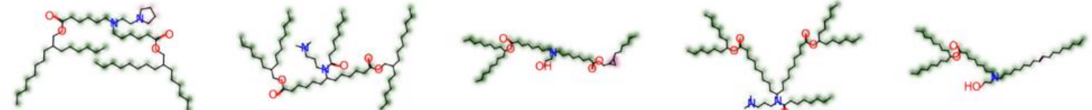
Answer: Thanks for your reminding. When AI model predicts mRNA delivery efficiency, it can output the probability of being positive in delivery efficiency, so we utilized this information and screened out lipids with high probability to be synthesized. This section was revised as:

"The choice of lipid tails was influenced by our synthesis capabilities. Our tail library allowed the synthesis of 666 possible lipids for each head type. They were predicted for pKa and delivery efficiency, and the probability of being positive in delivery efficiency was output by the model. Consequently, lipids with desired pKa, high probability, and positive mRNA delivery efficiency were chosen for synthesis: LQ085, LQ086, and LQ087 (Fig. 2c). Their predicted pKa and positive probability are listed in Supplementary Table 4."

- Was any structural similarly analysis done on the virtual lipids? It is unclear how the authors define "diversity."

Answer: Thanks for your kind advice. We provide a visualization of similarity analysis for 40 lipids with the highest ECFP score in Supplementary Fig. 3. For each lipid molecule, we show its summarized similarity to all other lipids. The distinct substructures are colored in red while similar part in green. For example:





We can see that nearly all cyclic structures are in red and therefore show diversity to other lipids. This is why we emphasize cyclic structures in our generated lipids, like LQ085, 086, 087, 089, and 091.

- In the second round of virtual screening, the authors intentionally lock in the head group as a synthesis capability factor before down-selection. This reportedly shrank the pool of lipids considerably. The subsequent discussion of how each model performs is extremely confusing. Even upon rereading several times, I cannot explain how the selection process resulted to the final six compounds.

Answer: Thanks for your comments. We built a Model 2 increasing the criteria of positive mRNA delivery to 2-fold to the original level, attempting to train a stricter model. The following section was largely rewritten:

"Compared to Model 1, the performance of Model 2 was defective in validation using the collected data (Fig. 4b). This defective performance was also evidenced by the external validation, in which the number of wrong predictions increased from three to six (Fig. 4c). However, all the mistakes happened to be that truly positive lipids were falsely predicted as negative, while truly negative lipids were predicted correctly. In other words, Model 2 showed a stricter criterion when assessing mRNA delivery efficiency, which was also reflected by the increasing precision index if validated against the original data (Supplementary Table 5). Higher precision means fewer false positive predictions.

Combining the ethanolamine head group and our tail library, 666 lipids were constructed, which included the molecules in the external validation set. Among the 666 lipids, Model 1 predicted 94 positive lipids, from which Model 2 predicted 21 positive lipids (Supplementary Fig. 8), while the other 645 molecules were negative in delivery (Fig. 4c). Therefore, the 21 lipids were more likely to have better delivery efficiency and worth exploring. Like the first round, lipids with desired properties and diverse tail structures, such as two long branches, dendritic branches, and cyclohexyl groups, were

preferred. Eventually, six of them (Fig. 4d) were selected for synthesis and evaluation. Their predicted pKa and probability of being positive in mRNA delivery efficiency were reported in Supplementary Table 6."

- I find it challenging to believe that N/P or weight ratio of mRNA (processing conditions) has negligible effect on the biological readout of delivery for LNPs. Processing LNPs is crucial to their delivery efficacy, all else equal, especially for in vivo settings, serving as a competitive advantage for biotech and pharmaceutical companies that must be reported once they reach pre-clinical and clinical studies. Nearly any LNP review emphasizes this point. Right away, this puts the quality of the dataset used to train both models into question. If they are not available in the majority of the original sources, in my view they should not be used to train any AI model.

Answer: Thanks for your careful reminding. N/P ratio is truly an important formulation factor influencing the LNP performance. We did not include this parameter because many patents provided the ratio as ranges instead of clear values. However, based on these ranges, we can estimate their N/P ratio. For example, N/P ratio of lipids in Supplementary Table 1 are shown below:

ID of lipid	Chemical structure of the ionizable lipid	Formulation of the LNP	Ref.
MC3		Lipid: DSPC: Chol: DMG-PEG (mol%) =50:10:38.5:1.5, N/P=6.	[1-8]
Amino Lipid 13 (LP01)		Lipid: DSPC: Chol: PEG-c-DMA (mol%) =54.6:10.9:32.8:1.6, lipid: mRNA (wt/wt) =12:1, N/P=6.5.	[1]
Compound of formula (IV)		Lipid: DOPE: Chol: DMG-PEG (mol%) = 40:20:38.5:1.5, N/P=5.67.	[2]
Compound 25 (SM-102)		Lipid: DSPC: Chol: DMG-PEG (mol%) =50:10:38.5:1.5, N/P=5.67, lipid: mRNA (wt/wt) ≈20:1.	[3,4]

Compound 9		Lipid: DSPC: Chol: PEG-DMA (mol%) =50:10:38.5:1.5, lipid: mRNA (wt/wt) =10:1-30:1, N/P= N/P=2.1-6.2.	[5]
Compound 2		Lipid: DSPC: Chol: PEG-DMA (mol%) =50:10:38.5:1.5. lipid: mRNA (wt/wt) =10:1-30:1, N/P=2.9-8.6	[6]
Compound 5		Lipid: DSPC: Chol: PEG-DMA (mol%) =50:10:38.5:1.5. lipid: mRNA (wt/wt) =10:1-30:1. N/P= N/P=4.5-13.6	[7]
Compound 3 (ALC-0315)		Lipid: DSPC: Chol: PEG-DMA (mol%) =50:10:38.5:1.5, total lipid: mRNA (wt/wt) =10:1-30:1, N/P=2.3-6.8.	[8]
OF-02		Lipid: DOPE: Chol: C14-PEG (mol%) = 35:16:46.5:2.5, lipid: mRNA (wt/wt) =10:1, N/P=4.9.	[9]
306Oi10		Lipid: DOPE: Chol: C14-PEG (mol%) = 35:16:46.5:2.5, lipid: mRNA (wt/wt) =10:1, N/P=9.6.	[10]
306-O12B		Lipid: Chol: DOPC: DMG-PEG (mol%) =50:38.5:10:1.5, lipid: mRNA (wt/wt) =7.5:1, N/P=4.9.	[11]

Most lipids in our dataset were derived from MC3 or SM-102, and the formulation ratio of the lipids in LNP was approximately 50:10:38.5:1.5, with N/P ratio of about 6 or lipid/mRNA weight ratio of about 20. Other types of ionizable lipids also have N/P ratios around 6.

Several studies compared the mRNA delivery efficiency with different N/P ratios. The result is shown such as Fig. 7e in Carrasco, M. J. et al. Commun Biol 4, 1–15 (2021) and Fig. 2b and 2c in Kauffman, K. J. et al. Nano Lett 15, 7300–7306 (2015). We can find that the effect of N/P ratio or lipid/RNA weight ratio is within one magnitude.

But, as our data shown in Supplementary Fig. 2, the efficiency of LNPs is spanning across several magnitudes. Therefore, lipid structure does much more difference to mRNA delivery. The lipid structure can be evaluated as the ECFP score that is found to be linearly correlated to mRNA delivery efficiency ($r = 0.6819$). This indicates that our model identifies the main influencing factors in LNP despite of the missing information of N/P ratio. Note that, our model did not receive specific delivery efficacy values, it was trained based on classification labels. Thus, this evolved knowledge is the result of machine learning instead of information leaking.

To summarize, the N/P ratios in our collected LNPs is concentrated to around 6. We did not represent this variable because we did not want to introduce unnecessary manual adjustment. The trained model has been validated against quantified delivery efficiency, which is beyond its expectation. This AI model should be convincing and reliable.

- Why are the images of the total luminescence in mice not shown for Round 2? This needs to be in the Supporting Information at minimum.

Answer: Thank you for reminding us. The images of the bioluminescence in mice in the second round of screening were supplemented in Supplementary Fig. 6 of the revised version.

- Several of the references are missing volumes and page numbers.

Answer: Thanks for your comments. The volumes and page numbers of cited articles have been added to the references.

- Finally, in both the main manuscript and the supporting information, no materials characterization is provided on the synthesized lipids other than a protocol and a yield. This is unacceptable. Any scientist cannot claim to have made novel molecules for use in something as important as in vivo applications without definitively showing that compound is pure and actually follows literature synthesis procedures. Millions of hypothetical structures generated from machines are useless if they cannot be synthesized by material chemists in the lab.

Answer: Thanks a lot for your kindly reminding. The NMR result of 9 ionizable lipids were added to the supporting information along with their synthesis routes.

Reviewer #4 (Remarks to the Author):

This is an interesting and timely study and could add new knowledge to the field and be of value to many researchers in the field. There are some comments the authors should address:

1. Ionisable lipids for vaccines will be different for those for therapy, give the balance of immunogenicity. I think the authors should consider this in their discussion.

Answer: Thanks for your comments. You have raised a very important opinion. Immunogenicity is truly an important factor for mRNA therapy. In protein supplement therapy, the immunogenicity is desired to be low, but in vaccines, moderate immunogenicity is hoped to enlarge the immune response. Currently, our AI model ensures that the screened lipids have relatively high expression level for in vivo applications, therefore it does not consider the immunogenicity. It is a good attempt to train a model for this goal in the future.

Additionally, we think other computational modeling tool such as physiologically-based pharmacokinetic (PBPK) and quantitative system pharmacology (QSP) model is helpful to simulate the complex signal pathway such as immunogenicity of LNP. The two model are based on ordinary differentiation equation and are mechanistic model. This has been added to the discussion of the manuscript.

"Lastly, the goal of this work is to construct lipids with generally high mRNA delivery ability, not specific for any organ, disease, or therapy. Therefore, only data of luciferase and hEPO mRNA delivery in mice were collected, as this is a basic screening method. However, models tailored according to therapeutic objectives or types of diseases are more appealing. For example, maximizing protein expression level is the priority in mRNA therapy supplementing missed proteins, but in mRNA vaccines against viruses, immunogenicity of the formulation needs additional consideration[58]. Developing models predicting immunogenicity is important for mRNA delivery. Likewise, another iteration direction of the model will be to screen out lipids with high-level expression of mRNA in organs other than the liver to meet the needs of a variety of diseases. Additionally, prediction in primates and even humans for specific diseases is profound for clinical translation. AI modeling methodology is still possible to handle these tasks only if data supports. However, other advanced modeling methods such as physiologically-based pharmacokinetics (PBPK) and quantitative systems pharmacology (QSP) models[59–62] are very useful. PBPK is specialized in inferring the fate of drugs across different species. This inference is based on the properties of the drug and the physiological conditions of the subject, and therefore such extrapolation is mechanistically based. QSP is also mechanistic, predicting dynamic changes in signal pathways, biomarkers, and even therapeutic effects. For a complex system such

as immune response, QSP is promising to address it[63,64]. Further, the association of the two models can integrate various in vitro and in vivo data, being able to quantify rates of critical processes in nucleic acid delivery such as RNA escape from endosomes[65].”

2. The model looks at luciferase and hEPO, this then focuses on therapeutics rather than vaccines. Protein expression levels would not be the only factor in vaccine efficacy.

Answer: Thanks for your comments. As mentioned above, immunogenicity is as important as mRNA expression in the application scenario of mRNA vaccines. A certain high efficiency of mRNA delivery is the prerequisite for LNPs to be applied in vivo, particularly in applications other than vaccines. We think that AI models would be better tailored for different diseases and therapies, but this might be limited by data size. Therefore, we think mechanistic model such as PBPK and QSP is helpful in simulating immune response pathway. This method can integrate various in vitro immune effect data and in vivo drug exposure data to make prediction. Protein expression predicted by AI models and immune effect from mechanistic model construct a valuable estimation of mRNA vaccines.

3. Did the model consider the type and biological sex of the mice? This may impact on the results.

Answer: Thank you for your reminding. Before this study, we had already established a stable platform for in vivo evaluation of ionizable lipids. Herein, we have again verified the stability and reliability of the method as shown in Supplementary Fig. 7. The results showed that there was no difference in luciferase mRNA expression between male and female mice.

4. There is a heavy emphasis on pKa, and this is well known that around 6 to 6.8 is the best range. Does the model consider structure of the lipids, and configuration within the LNPs. This is more important than the pKa.

Answer: Thanks for your comments. Yes, we consider the structure of lipids and the composition in LNP. This is shown in the method section as below:

“In this study, the ionizable lipid structure was represented by ECFP converted via the RDKit package in Python[55]. The ECFP radius was set to 9, and the number of bits was set to 1024. Each ionizable lipid had a unique ECFP sequence. The involved three helper phospholipids, DSPC, DOPE, and DOPC were represented by two ‘0-1’ category variables (‘DS’ or ‘DO’, ‘PC’ or ‘PE’). The PEG-lipids were represented by single multiple-category variable. Only one type of cholesterol lipid was included in our data, so it was not represented. Molar ratios of the four types of lipids in LNP were represented as numeric variables between 0 and 1.”

The three-dimensional structure or shape of LNP was not considered in the model.

Shape or structure of LNP is significantly affected by lipids used and their molar composition. Since our model included the lipid structure, the shape factor should be implicitly considered by the model.

5. Figure 5. The area under the curve ploy is not particularly useful for a 3 point graph and when looking at it, I could not follow then I spotted the colour coding does not match. This should be corrected and the authors should consider again if AUC is the best way to look at this.

Answer: We appreciate for your suggestion. Figure 5 has been merged into Fig. 4. In the revised manuscript, we validated the method for evaluating the lipids (Supplementary Fig. 7). We plotted the bioluminescence-time curve over eight time points (1, 2, 4, 6, 8, 10, 24, and 48h). Then, AUC was calculated by 3-point method (4, 24, 48h) and 8-point methods, resulting in no difference between the two methods. For the determination of mRNA delivery, the curve of luminescence over time and AUC thereof can monitor the expression of mRNA within a long period to describe the long-term performance of LNP. Since AUC analysis miss the organ distribution of mRNA expression at the highest point, we supplemented the organ distribution of luciferase expressed by the LNP-loaded mRNA administered, both intravenously and intramuscularly (Fig. 6 and 7) to comprehensively present the performance of lipids

6. Figure 5. The number of mice used is low and looks to be in a single study. Given the closeness of the data and the variability in mouse studies, the study should be replicated again. Even with low numbers (e.g. 2 mice per group).

Answer: Thank you for your reminder. We have replicated the in vivo mRNA expression of intravenous administration and increased the number of mice to six, three females and three males (Fig. 6). The order of the intensities between different lipids matched with the AUC method of three mice.

7. Was a 'negative' control tested e.g. a lipid that should only give low responses tried to validate the model. This should be added in to the replicate study.

Answer: Actually, the LQ085 is a negative lipid tested in the first round of screening (Fig. 3) and it was included in the replicate study (Fig. 6 and 7).

8. How do the authors consider the lack of data translation from mouse to NHPs to humans, this is a recognised problem. The authors should comment.

Answer: Thanks for your comments. AI models face challenges in extrapolation across species, but other computational tools such as PBPK model are suitable for this task. PBPK predicts drug exposure based on both formulation properties and physiological parameters, such as organ volumes and organ blood flows, etc. If formulation properties can be determined in vitro or calculated by fitting to mouse data, they can be associated

with NHP or human physiological data to make prediction. For example, PBPK is helpful to inform the design of first-in-human study (Miller NA, Reddy MB, Heikkinen AT, Lukacova V, Parrott N. Physiologically Based Pharmacokinetic Modelling for First-In-Human Predictions: An Updated Model Building Strategy Illustrated with Challenging Industry Case Studies. Clin Pharmacokinet. 2019 Jun;58(6):727-746.). The following content has been added to the discussion:

"Additionally, prediction in primates and even humans for specific diseases is profound for clinical translation. AI modeling methodology is still possible to handle these tasks only if data supports. However, other advanced modeling methods such as physiologically-based pharmacokinetics (PBPK) and quantitative systems pharmacology (QSP) models[59–62] *are very useful. PBPK is specialized in inferring the fate of drugs across different species. This inference is based on the properties of the drug and the physiological conditions of the subject, and therefore such extrapolation is mechanistically based. QSP is also mechanistic, predicting dynamic changes in signal pathways, biomarkers, and even therapeutic effects. For a complex system such as immune response, QSP is promising to address it*[63,64]*. Further, the association of the two models can integrate various in vitro and in vivo data, being able to quantify rates of critical processes in nucleic acid delivery such as RNA escape from endosomes*[65]*."*

9. I would question the validity of using both IV and IM data as these are very different data sets and IV is thought to rely on ApOE binding and then delivery to the liver whilst IM is thought to involve resident immune cells at the injection site or infiltrating immune cells to the injection site.

Answer: Thanks for your kind advice. We reevaluated our data and found the IM samples make up a minimal proportion of the whole data set (10 out of 397). Keeping them may increase data heterogeneity without significant benefit. Thus, we have reevaluated our model based on only IV data and reported the new performance in the revised manuscript. Compared to the previous performance, the influence is minimal.

IV and IM are both important routes, we conducted the mRNA expression test via IV and IM (Fig. 6 and 7) to learn more about our lipids.

10. The discussion should add more into how the reader can gain from this model so a wide audience can apply this learning. This would add more significance to the work and the paper.

Answer: Thanks for your comments. Besides the AI models and the selected lipids, we emphasize three points for broader applicability. First, the defined ECFP score is positively correlated with mRNA delivery efficiency. This shows that our model develops a refined quantitative relationship based on rough qualitative data, proving the validity of model building method. This content has been added to discussion as:

"Although the AI model was trained solely on data samples labeled with categories of mRNA delivery efficiency, it unexpectedly developed the ability to quantify this efficiency, which is a fortunate discovery. The efficiency is positively correlated with the newly defined ECFP score. Some lipids containing squaramide were reported to show remarkable delivery efficiency[15], and they were also given high ECFP scores (Supplementary Fig. 3). The score is derived from lipid ECFP bits and their corresponding SHAP values. The SHAP algorithm, providing quantified assessments of feature contributions, has proven to be highly effective in explaining the outputs of AI models[43,51]. From this view, the model we have developed is interpretable in terms of its structure-activity relationship."

Second, in the second screening model, we combined two models to evaluate lipid performance. Model 1 performed better in general prediction accuracy but is less efficient in selecting candidates. Model 2 was trained with a classifying criterion doubling the original one. It was less performed in general but showed high precision, filtering out falsely positive prediction effectively. Proper combination and leveraging the advantages of different models led to 6 positive lipids successfully, and this strategy is valuable for other AI applications. This was added to the discussion section as:

"The second round of virtual screening focused on lipids containing the ethanol amine head and trained a stricter model (Model 2) to facilitate the screening. The evaluation process was actually an association of Model 1 and Model 2. Model 1 performed better in general prediction accuracy but still predicted too many lipids with positive mRNA delivery efficiency. Model 2 was performed less well in general but showed high precision, making it effective in filtering out false positive predictions. This round of screening yielded six ionizable lipids, LQ089-094, all of which were equal to or superior to MC3, proving the validity of the screening strategy. Properly combining and leveraging the advantages of different models is important for AI applications.
"

Third, we performed a comprehensive structure-activity relationship analysis for a specific type of ionizable lipids. The result is straightforward, showing how small structural changes affect performance of lipids. This experience can guide the design of lipids. But this method is not flawless, because it can only be used to evaluate molecules with specific structural patterns. The diversity of structures is so high-dimensional that it is unrealistic to present intuitively and comprehensively the structure-activity relationship. This content in discussion is like:

"Lipids with head groups containing hydroxyl are commonly tested but show varied performance. The structure-activity relationship in this type of lipid has not been described in detail. However, with a well-trained AI model, this relationship can be comprehensively explored (Fig. 5). It can be observed that although the lipids, after ECFP transformation, produce high-dimensional and discontinuous features, gradually changing the structure of the lipids, such as gradually extending the length

of the carbon chain and shifting the position of the linker, results in a continuous trend in the AI model's predictive performance. To obtain well-performing lipids, all chain segments in tails should have harmonious length. The linker position should be compatible with the whole tail length, and the length threshold is dependent on the linker type. Ionizable lipids like SM-102, ALC-0315[17]*, and our selected LQ092, LQ093, and LQ094 all belong to the area with a relatively high positive rate. The structure-activity relationship represented in this way is easy to be understand and applicable to guide molecule design. However, this method can only show the relationship in a narrow region of the molecule design space."*

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I am satisfied with the responses, modifications, and new experiments that the authors have included in the latest version of the manuscript.

Reviewer #2 (Remarks to the Author):

The authors have resubmitted their manuscript on "Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery". Substantial major revisions were provided from all reviewers' feedback, and the revised manuscript is significantly improved from the original submission.

This work is an ambitious attempt to leverage AI to design lipids for mRNA delivery in vivo. The changes to the manuscript address most of the points brought up, but the results still show the fundamental limitations of this approach in informing future LNP design. I am severely concerned that insufficient chemical and materials characterization of the nanoparticles before bioadministration will limit repeatability of the results, which show typical LNP accumulation to the liver regardless of structural changes in lipid design. The use of AI to design virtual lipids as a product is not novel for Nature Communications.

A more recent work by Dong et al. "Multicomponent Synthesis of Imidazole-Based Ionizable Lipids for Highly Efficient and Spleen-Selective Messenger RNA Delivery" provides a good example of rigorous characterization and understanding that is connected to in vivo RNA delivery with AI as a feature, not the main product of the manuscript.

Reviewer #2 (Remarks on code availability):

The authors do not provide any of the training data used for the AI models. This does not comply with Nature's policy of code availability and support the FAIR Guiding Principles for scientific data management and stewardship (https://www.nature.com/articles/sdata201618)

Reviewer #4 (Remarks to the Author):

The authors have nicely addressed all the comments raised. No further suggestions from me.

Reply to comments for Manuscript "Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery"

First, we would like to express our sincere thanks again to the reviewers for the constructive comments and suggestions on our manuscript "Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery" (Research Article, **NCOMMS-23-53992**). We have addressed these comments with corresponding changes and answers. The point-by-point responses to the referees were listed below this letter. We earnestly appreciate reviewers' dedicated work and hope that the revisions address all concerns. We believe that we have improved the quality of the manuscript and we are submitting our revised version for publication in the journal "Nature Communications".

Reviewer #1 (Remarks to the Author):

I am satisfied with the responses, modifications, and new experiments that the authors have included in the latest version of the manuscript.

Answer: Thanks for your kind comment.

Reviewer #2 (Remarks to the Author):

The authors have resubmitted their manuscript on "Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery". Substantial major revisions were provided from all reviewers' feedback, and the revised manuscript is significantly improved from the original submission.

This work is an ambitious attempt to leverage AI to design lipids for mRNA delivery in vivo. The changes to the manuscript address most of the points brought up, but the results still show the fundamental limitations of this approach in informing future LNP design. I am severely concerned that insufficient chemical and materials characterization of the nanoparticles before bioadministration will limit repeatability of the results, which show typical LNP accumulation to the liver regardless of structural changes in lipid design. The use of AI to design virtual lipids as a product is not novel for Nature Communications.

A more recent work by Dong et al. "Multicomponent Synthesis of Imidazole-Based Ionizable Lipids for Highly Efficient and Spleen-Selective Messenger RNA Delivery" provides a good example of rigorous characterization and understanding that is connected to in vivo RNA delivery with AI as a feature, not the main product of the manuscript.

Answer: Thanks for your kind comment. For chemical and materials characterization, we first used chromatography to purify synthesized ionizable lipids and then use NMR to verify their chemical structures. All ^{1}H NMR, ^{13}C NMR, and MS spectra of synthesized lipids were provided in the revised supplementary information. We measured particle size, zeta potential, entrapping efficiency, and apparent pKa of mRNA-LNP after their preparation. For LNP stability, we tested their in vivo performance after storage at a wide range of temperature for up to 30 days. In terms of safety, we measured the weight change of subject mouse and their main organs after dosage administration; blood samples were detected for inflammation (white blood cells), liver impairment (ALT), and kidney impairment (creatinine); and toxicity in organ was inspected using H&E staining. All characterization proofs for the newly designed ionizable lipids were as comprehensive as possible, aligning with the article you recommended (Dong et al.) for the characterization of lipid nanoparticles. The recommended study has also been cited in our revised manuscript.

To make sure of the repeatability of the result, firstly, we repeated the validation experiment of the newly formulated LNPs, which were administrated through intravenous and intramuscular routes in enlarged number of animals (Fig. 6, 7 and Supplementary Fig. 10-13). Secondly, we have conducted quality control and detailed characterization (particle size, PDI, and entrapping efficiency) for each batch of LNPs before animal testing. They are presented in Supplementary Table 4 and 6 for the first batch and Supplementary Fig. 9 for the second batch. Finally, both the formulation characterizations and the long-term stability tests (Figure 8) demonstrate that the LNPs maintain stable properties, and the process is consistent and repeatable, fully satisfying the criteria for in vivo administration.

As for hepatic accumulation, we have to stress that organ preference is not the concern of the presented work. Liver is the target organ for most existing ionizable lipids and current formulation screening heavily relies on expression of delivered mRNA in livers. Our designed new lipids mainly target the liver because they agree with the lipids used for training model.

The AI method introduced in this study is specifically designed to screen ionizable lipids that induce a generally high level of mRNA expression, which is fundamental for in vivo applications. A super-large ionizable lipid library, comprising nearly 20 million candidates, was evaluated. Following two rounds of AI-driven generation and screening, 7 out of 9 newly designed lipids demonstrated positive delivery efficiency compared to MC3. This outcome provides compelling evidence of the accuracy and practicality of our AI models. Thus, our research underscores the novelty and immense potential of AI modeling in the development of mRNA-LNPs.

We completed our study in mid-2023 and submitted the manuscript to Nature Communications on **November 8, 2023**. Dong et al.'s study, titled *'Multicomponent*

Synthesis of Imidazole-Based Ionizable Lipids for Highly Efficient and Spleen-Selective Messenger RNA Delivery,' was **submitted on Janunary 10, 2024 and published on May 22, 2024.** Despite this study serves as an excellent example of LNP characterization, it is unlike our work. Dong et al.'s study utilized an internal chemical library to produce ionizable lipids and conducted a total in vivo screening without employing AI for performance prediction, molecule design, or LNP characterization and screening.

Reviewer #2 (Remarks on code availability):

The authors do not provide any of the training data used for the AI models. This does not comply with Nature's policy of code availability and support the FAIR Guiding Principles for scientific data management and stewardship (https://www.nature.com/articles/sdata201618)

Answer: Thanks for your comment. In the first-round response, we have shared the training data and models via Supplement Data for the purpose of review. Now the files can be downloaded from https://figshare.com/s/ad928807e1b4795b9b5e. Furthermore, we plan to build the AI-LNP module in our FormulationAI platform for freely public access (https://formulationai.computpharm.org/, Dong, J., et al. FormulationAI: a novel web-based platform for drug formulation design driven by artificial intelligence. Briefings in Bioinformatics 25, bbad419 (2024)).

The data and code availability have been changed to as:

*"**Data availability***
The data that support the findings of this study are available on request from the corresponding author D.O.
Code availability
The code that support the findings of this study are available on request from the corresponding author D.O."

Reviewer #4 (Remarks to the Author):

The authors have nicely addressed all the comments raised. No further suggestions from me.

Answer: Thanks for your kind comment.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

I am satisfied with the authors' responses to reviewer #2's questions and comments.

Reply to comments for Manuscript "Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery"

First, we would like to express our sincere thanks again to the reviewers for the constructive comments and suggestions on our manuscript "Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery" (Research Article, **NCOMMS-23-53992B**). We have addressed these comments with corresponding changes and answers. The point-by-point responses to the referees were listed below this letter. We earnestly appreciate reviewers' dedicated work and hope that the revisions address all concerns.

Reviewer #1 (Remarks to the Author):

I am satisfied with the authors' responses to reviewer #2's questions and comments.
Answer: Thanks for your kind comment.