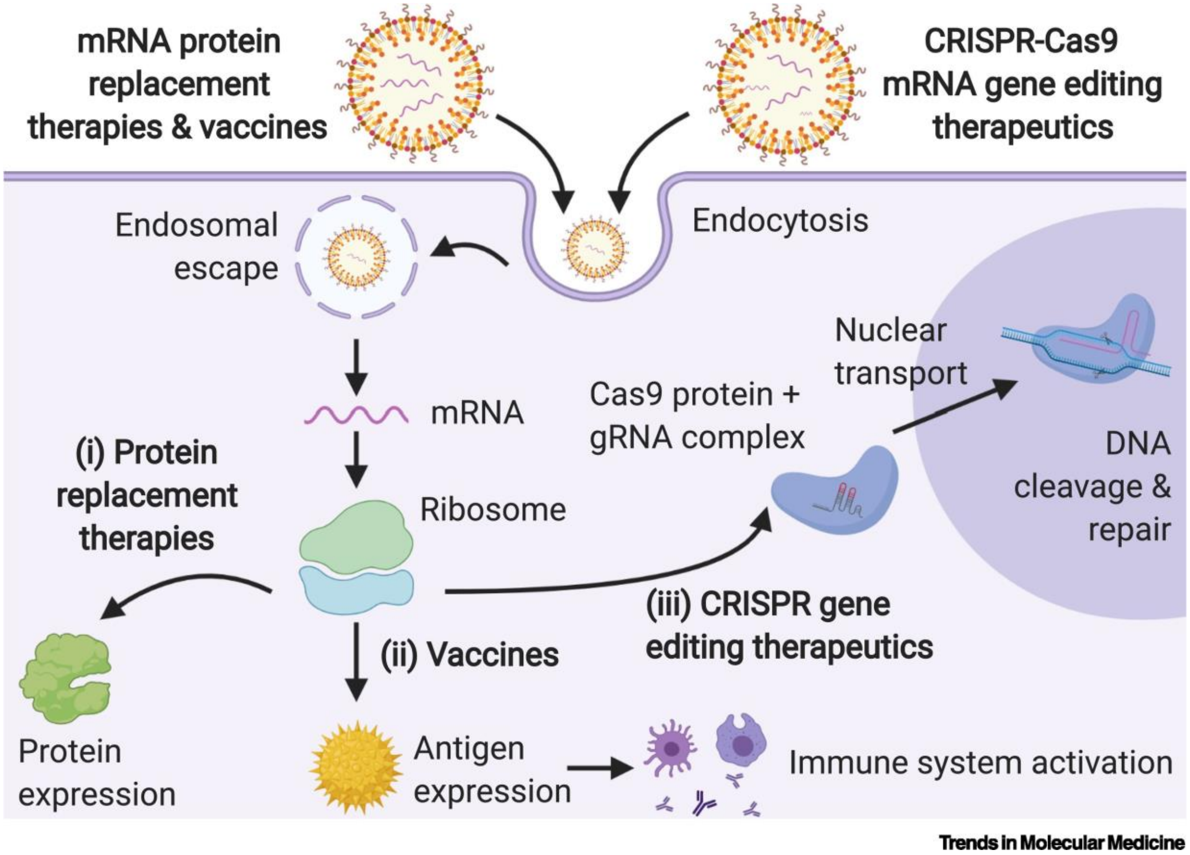
FormulationLNP: enhancing mRNA-lipid nanoparticle prediction via the language model and multi-task learning

Yiyang WU
Ph.D. candidate, Supervisor: Dr. Defang OUYANG
University of Macau, Macau SAR, China
yiyangwu2021@163.com

CRS 2025 ANNUAL MEETING & EXPOSITION
PHILADELPHIA, PA
JULY 13-18, 2025

NEXT-GENERATION DELIVERY INNOVATIONS

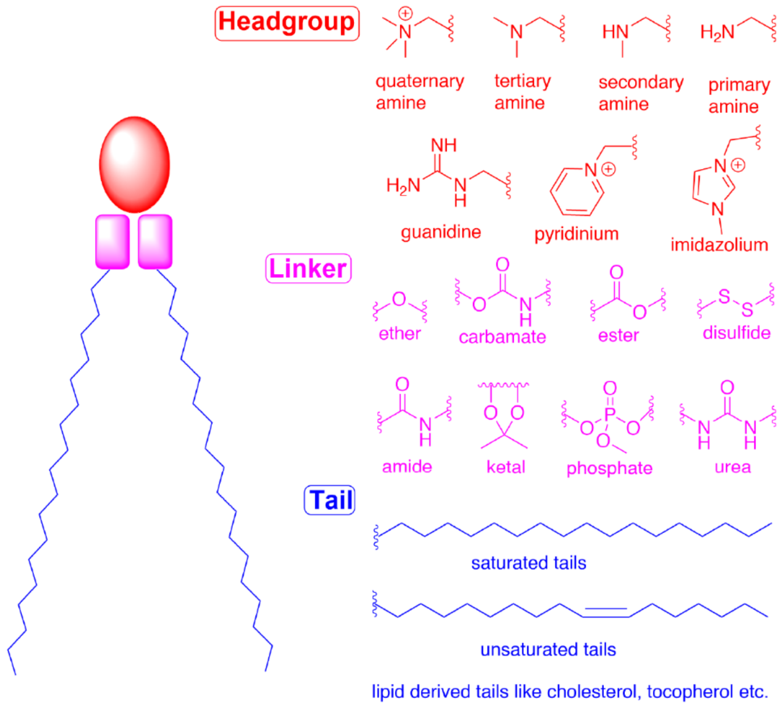
➢ Lipid nanoparticles for mRNA delivery

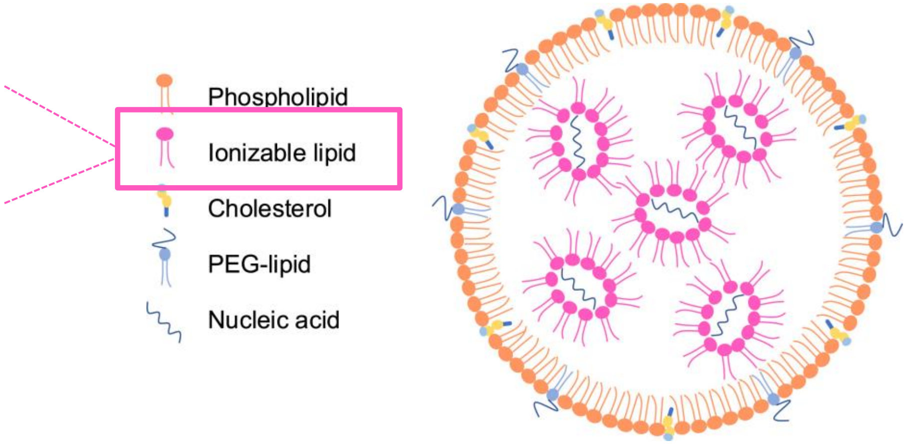


Lipid nanoparticle (LNP) is the most advanced non-viral delivery system for mRNAs

Trends in Molecular Medicine, 2021, 27(6), 616 - 617

➤ Ionizable lipids in LNPs





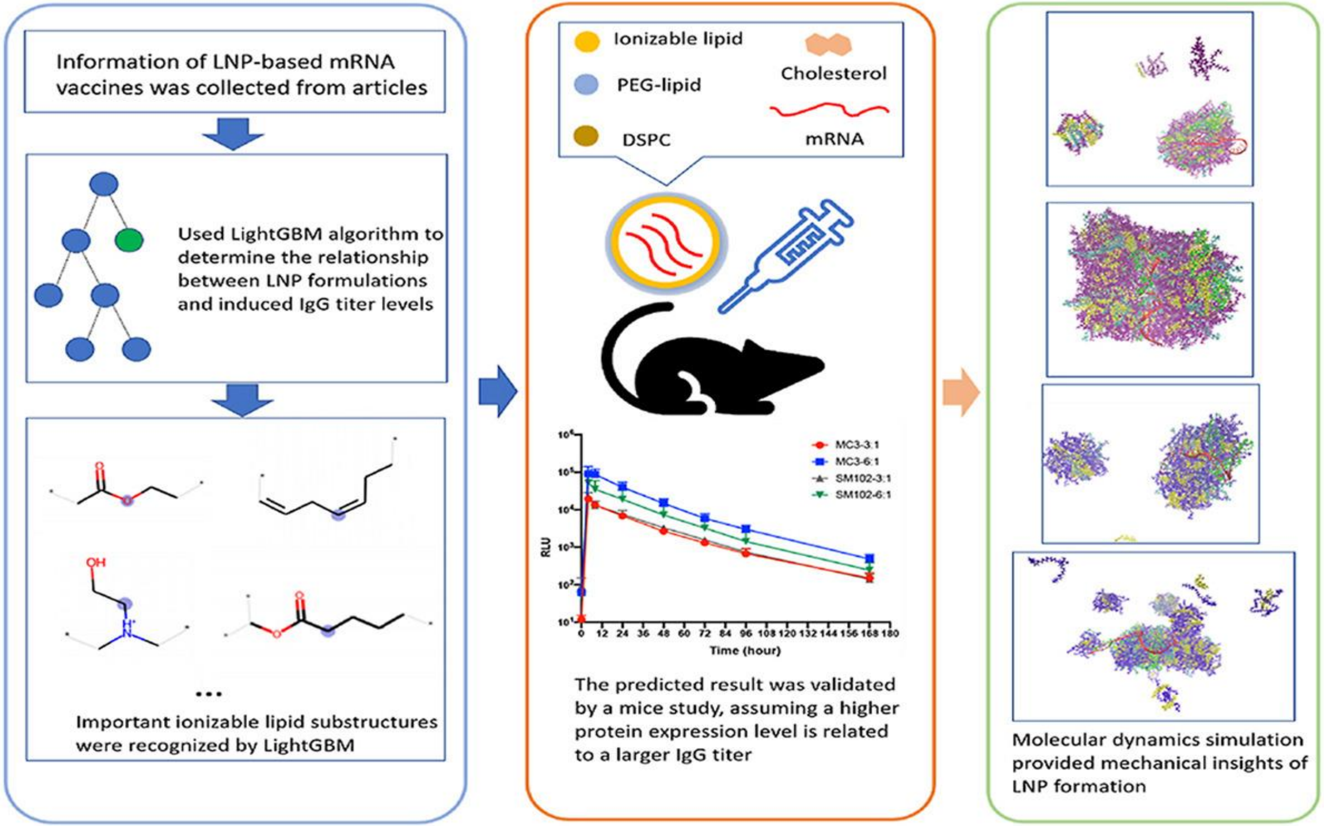
Research Limitation

Traditional trial-and-error experimental methods take a very long time to design the ionizable lipids and optimize the LNPs

Accounts of Chemical Research, 2022, 55 (1), 2-12;
Molecular Therapy Methods & Clinical Development, 33 (2), 101463

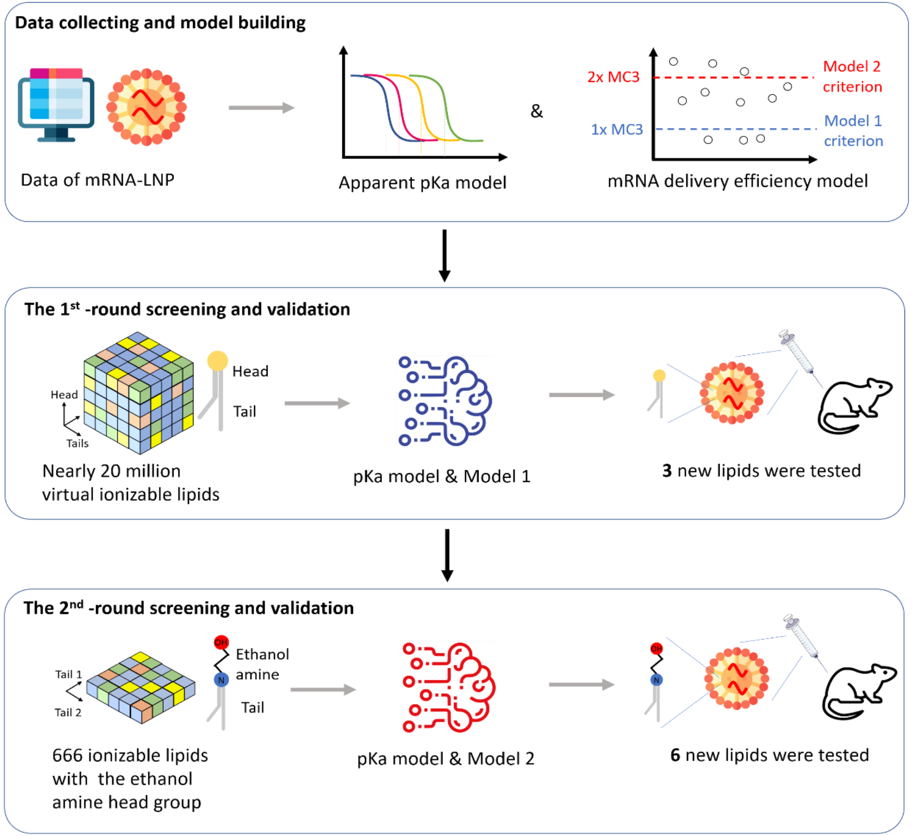
➢ Computation-driven development of mRNA-LNPs

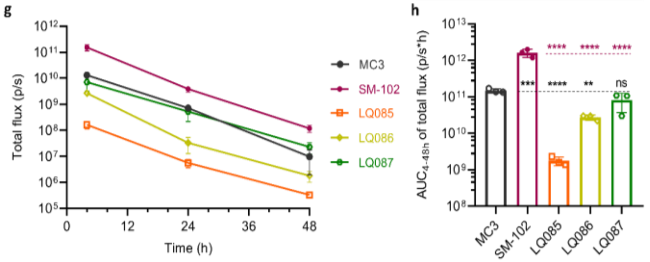
● Predicting lipid nanoparticles for mRNA vaccines by machine learning algorithm

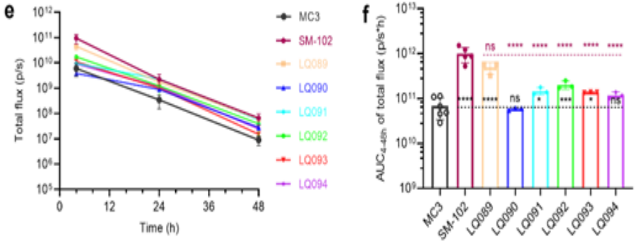


Acta Pharmaceutica Sinica B, 2022, 12 (6), 2950-2962;

● AI-driven rational design of ionizable lipids for mRNA delivery

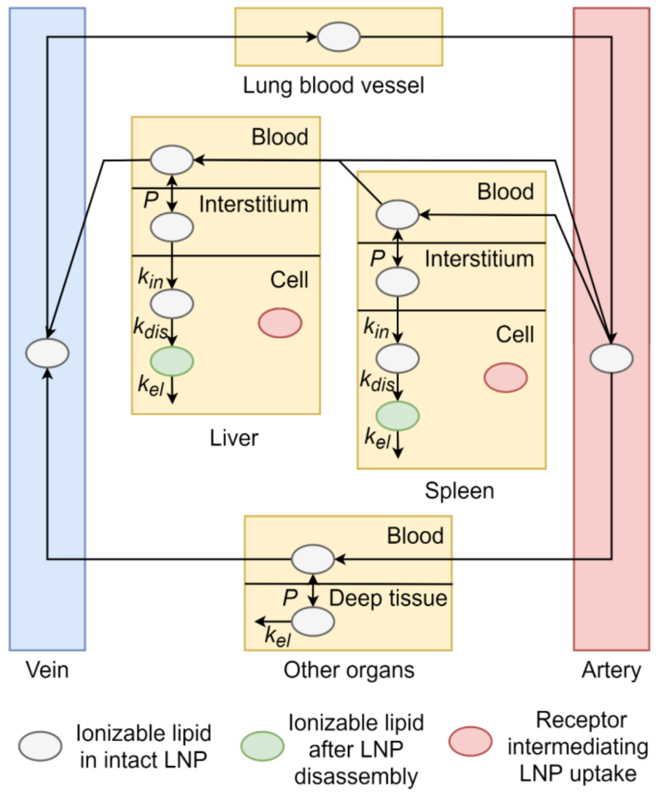


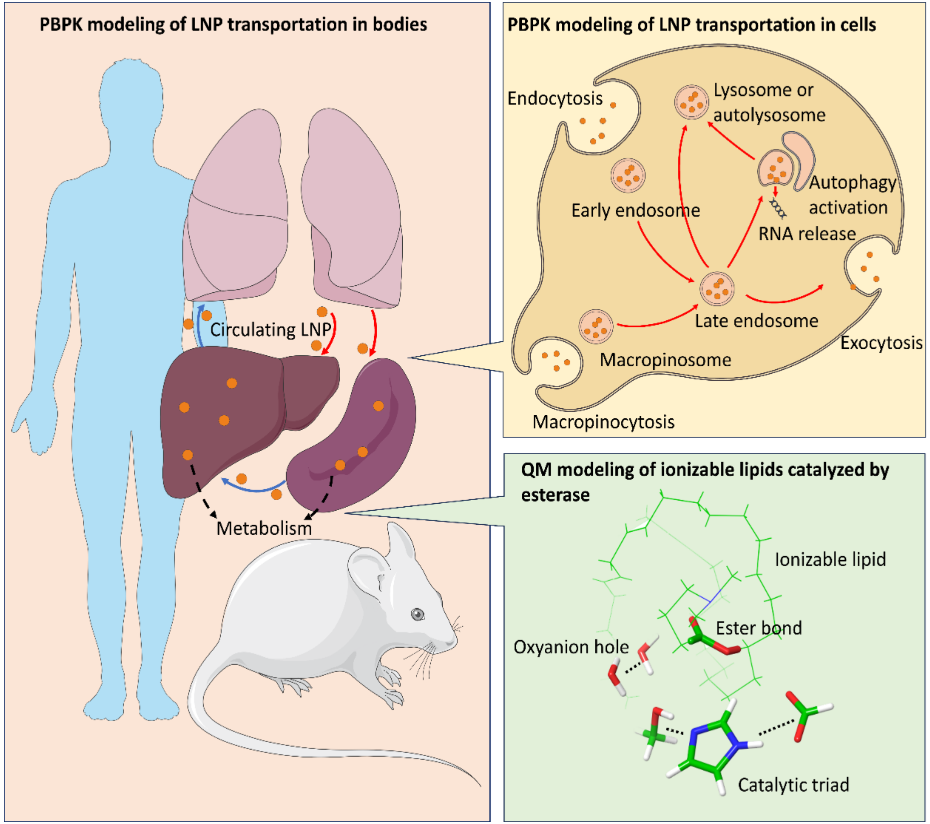




PCT patent application (PCT/CN2024/096356);
Nature Communication 2024, 15 (1), 10804

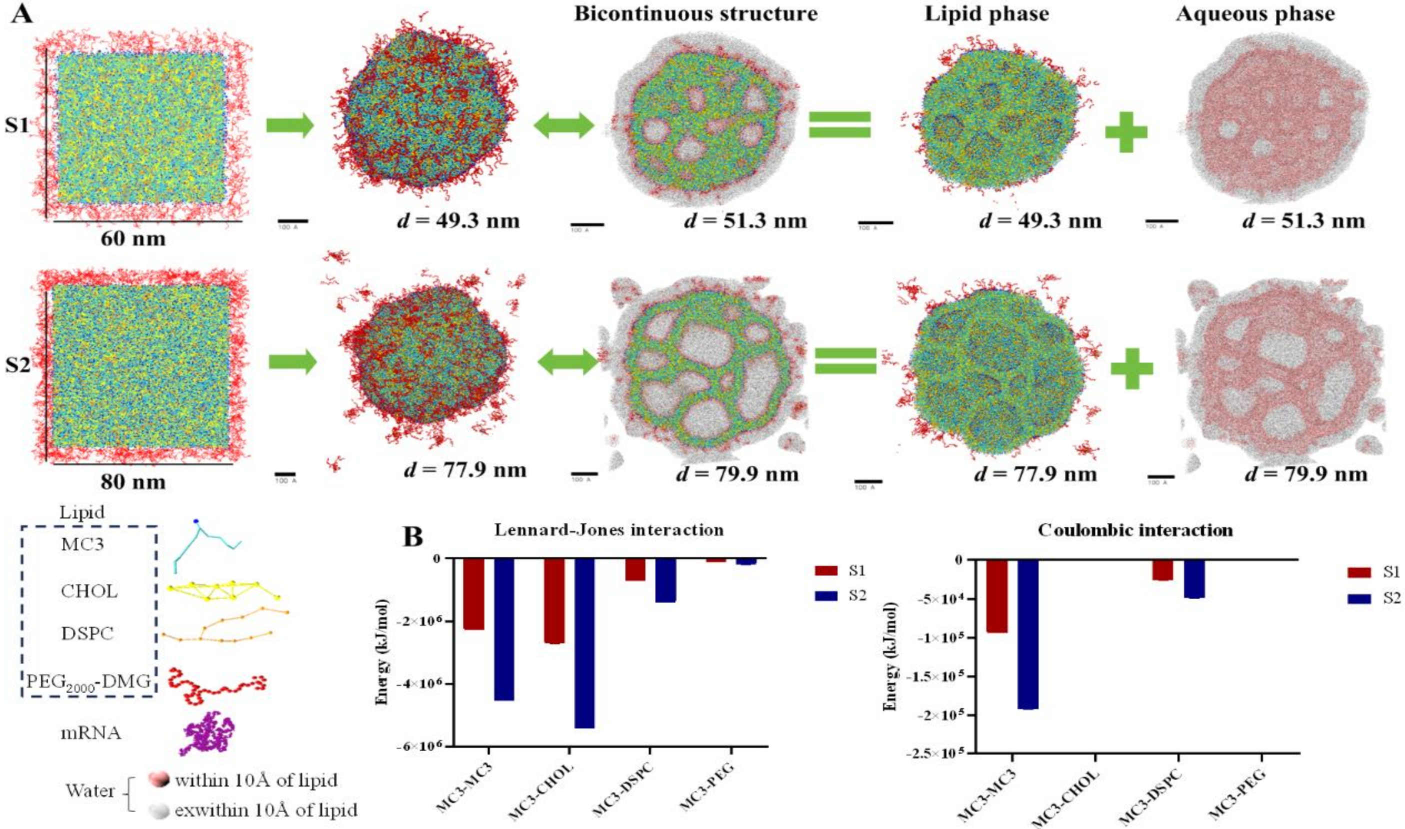
● *In vivo* model of RNA lipid nanoparticle





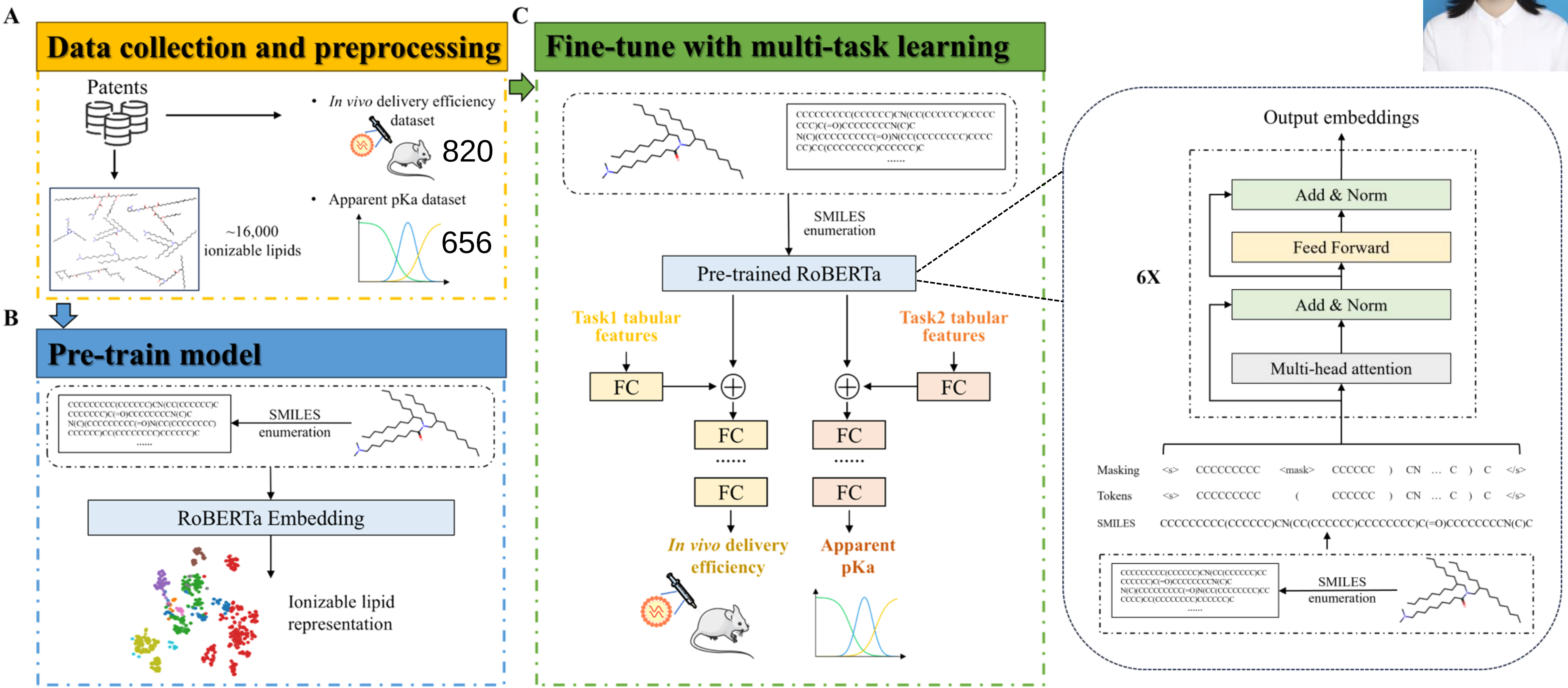
Acta Pharmaceutica Sinica B, 2024,14,4591–4607

- **Self-assembly and molecular structure of mRNA-lipid nanoparticles at real size from the ultra-large-scale simulation**



International Journal of Pharmaceutics, 2025, 670, 125114

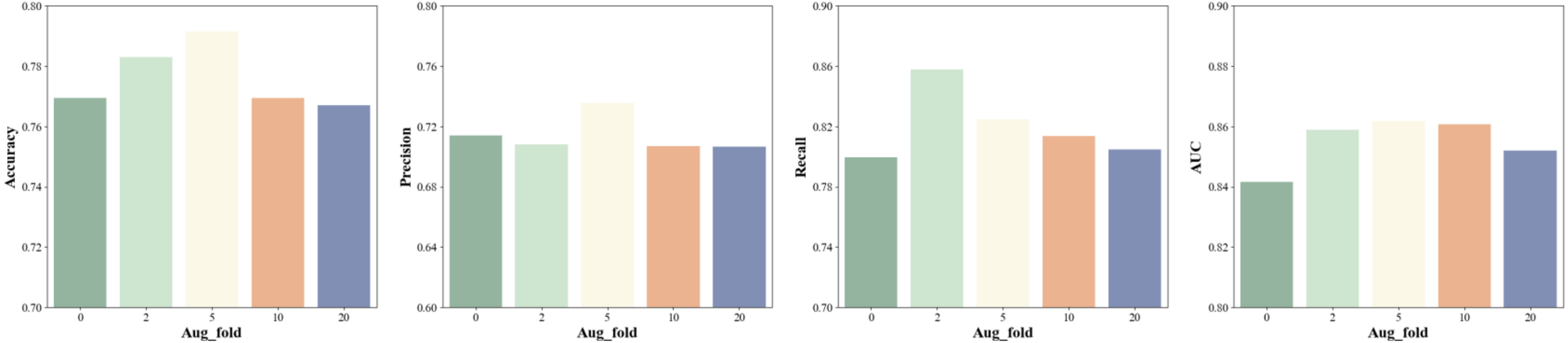
➢ **FormulationLNP** —facilitate the mRNA-LNP development in limited datasets



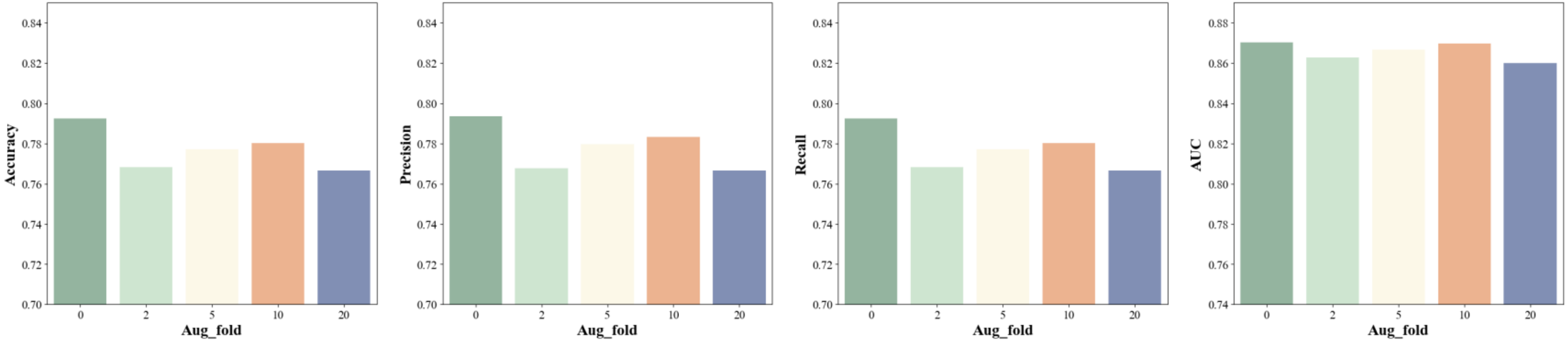
Unpublished

● Comparison of data augmentation level

A. *In vivo* delivery efficiency task



B. Apparent pKa task



Unpublished

- **Comparisons of models developed by different algorithms**

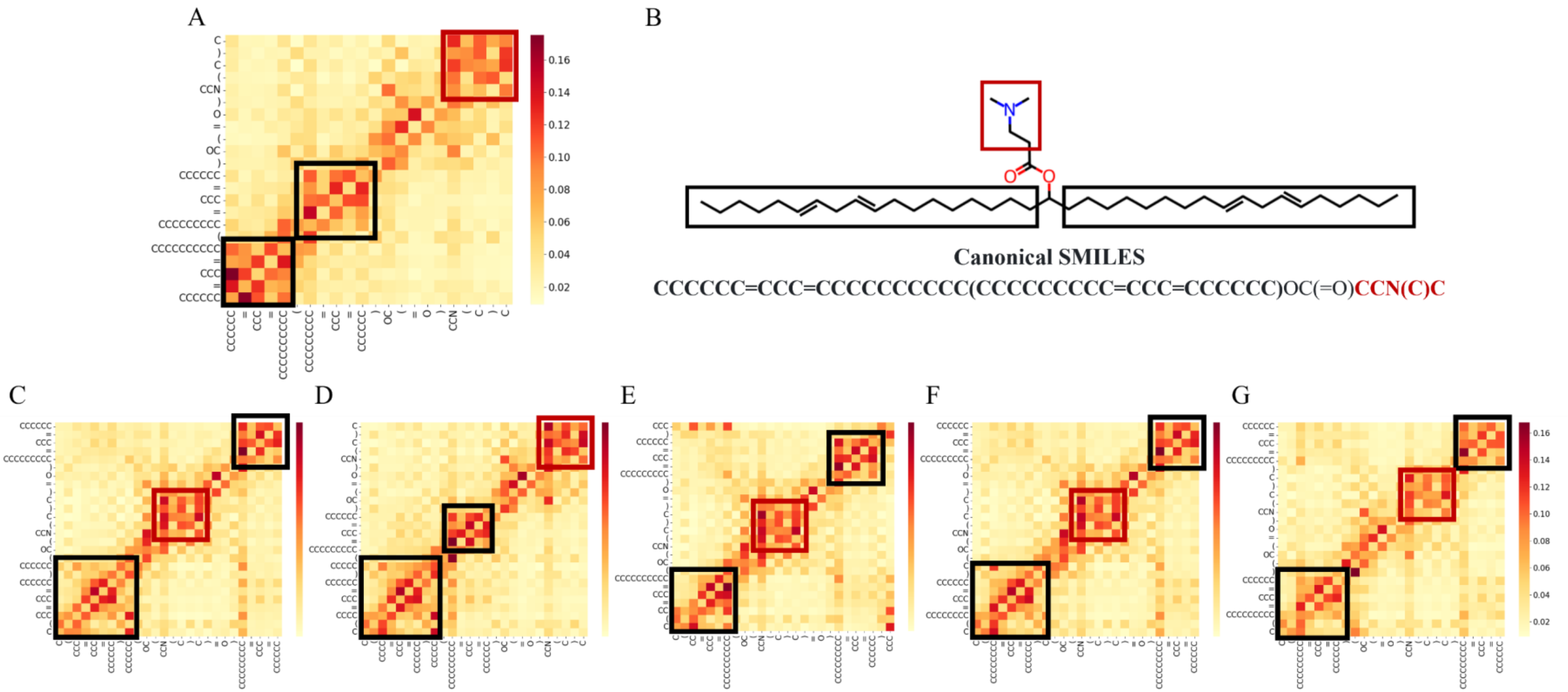
Model	*In vivo* delivery efficiency Accuracy	*In vivo* delivery efficiency AUC	Apparent pKa Accuracy	Apparent pKa AUC
LightGBM-RDKit	0.754±0.033	0.855±0.027	0.759±0.058	0.856±0.054
DNN-ECFP6	0.779±0.037	0.852±0.033	**0.780±0.037**	0.855±0.038
AttentiveFP	0.720±0.040	0.794±0.023	0.712±0.042	0.680±0.129
ChemBERTa	0.754±0.048	0.837±0.038	0.748±0.041	0.819±0.038
FormulationLNP (Ours)	**0.791±0.036**	**0.862±0.037**	0.777±0.048	**0.867±0.032**

- **Ablation study on the impact of pre-training, data augmentation and multi-task learning**

Pre-training	Data augmentation	Multi-task learning	*In vivo* delivery efficiency Accuracy	*In vivo* delivery efficiency AUC	Apparent pKa Accuracy	Apparent pKa AUC
√	√	√	**0.791±0.036**	**0.862±0.037**	0.777±0.048	0.867±0.032
√	√	×	0.767±0.036	0.852±0.039	0.773±0.058	0.875±0.035
√	×	√	0.770±0.027	0.842±0.030	**0.792±0.048**	**0.870±0.043**
√	×	×	0.761±0.022	0.844±0.030	0.789±0.043	0.859±0.049
×	√	√	0.755±0.025	0.814±0.036	0.770±0.036	0.803±0.049
×	√	×	0.743±0.032	0.802±0.032	0.720±0.054	0.800±0.041
×	×	√	0.737±0.055	0.824±0.041	0.724±0.055	0.784±0.057
×	×	×	0.737±0.043	0.807±0.029	0.709±0.079	0.777±0.061

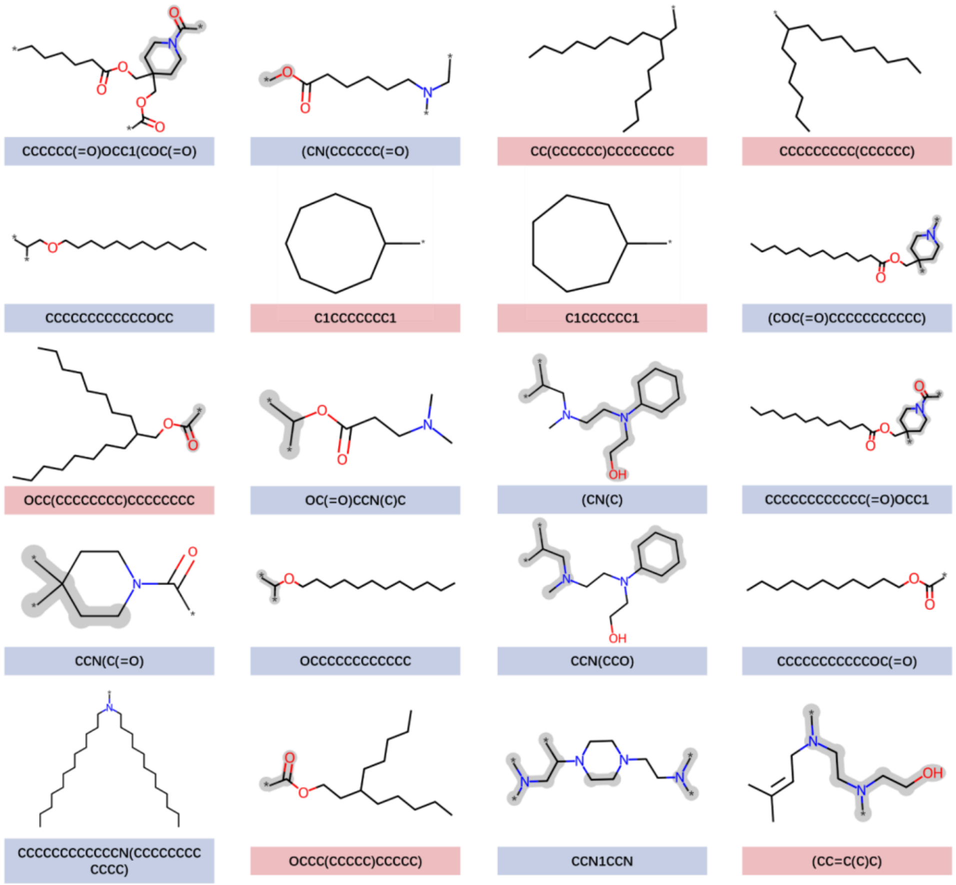
Unpublished

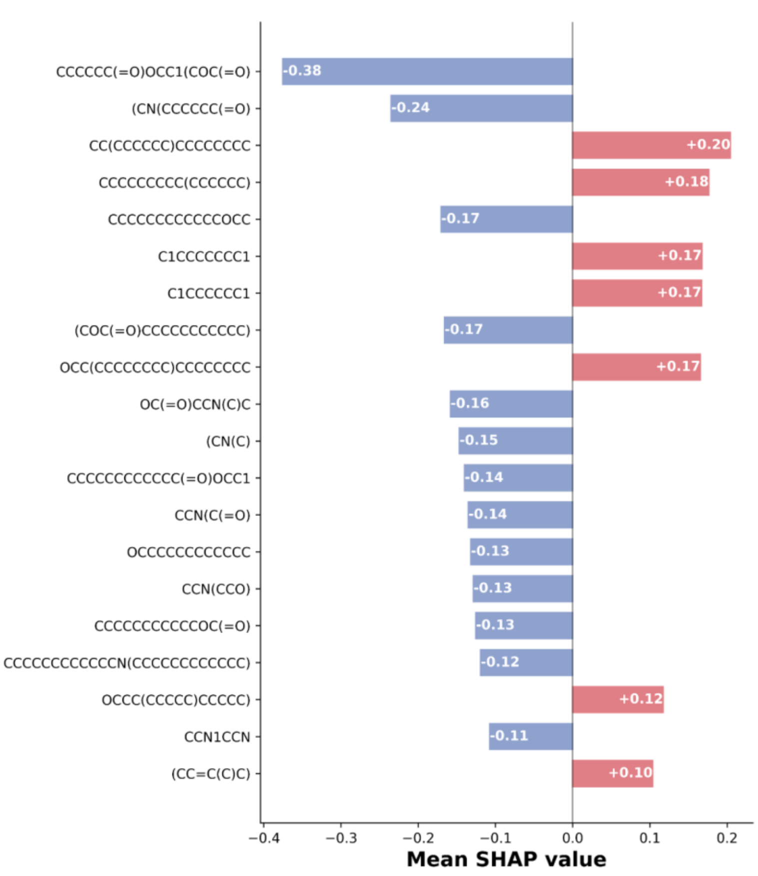
● Attention weights of an ionizable lipid represented by the SMILES strings



Unpublished

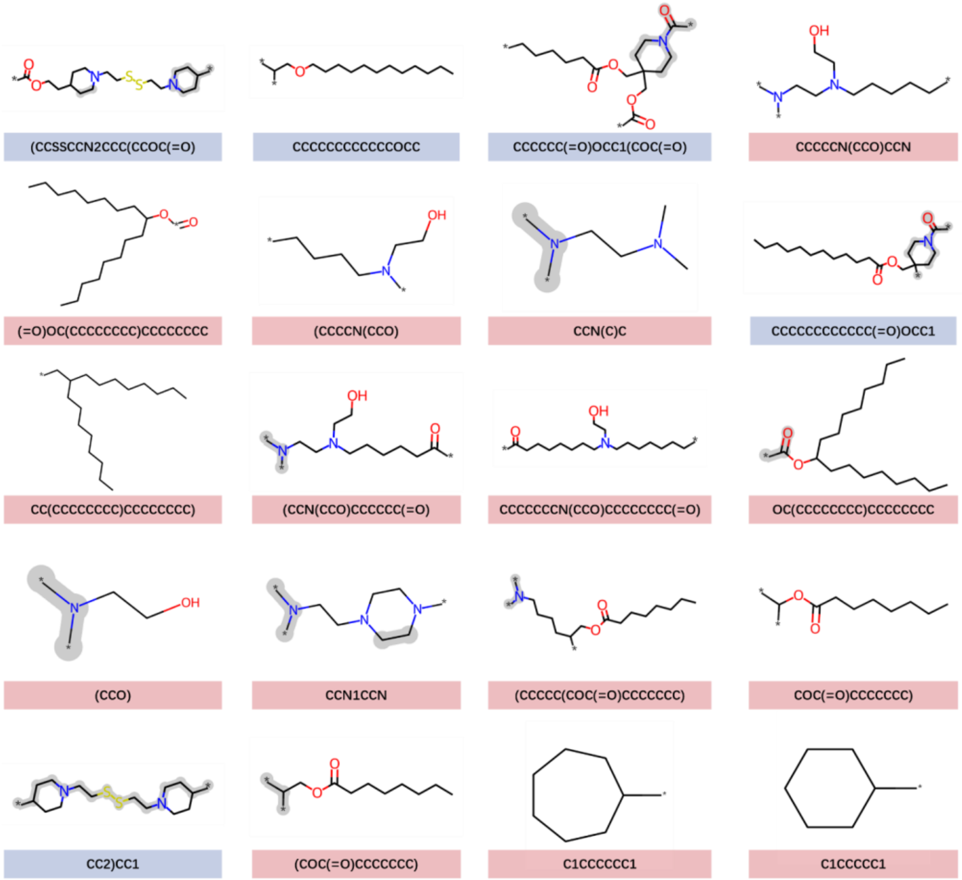
● Top 20 important substructures identified in the delivery efficiency task

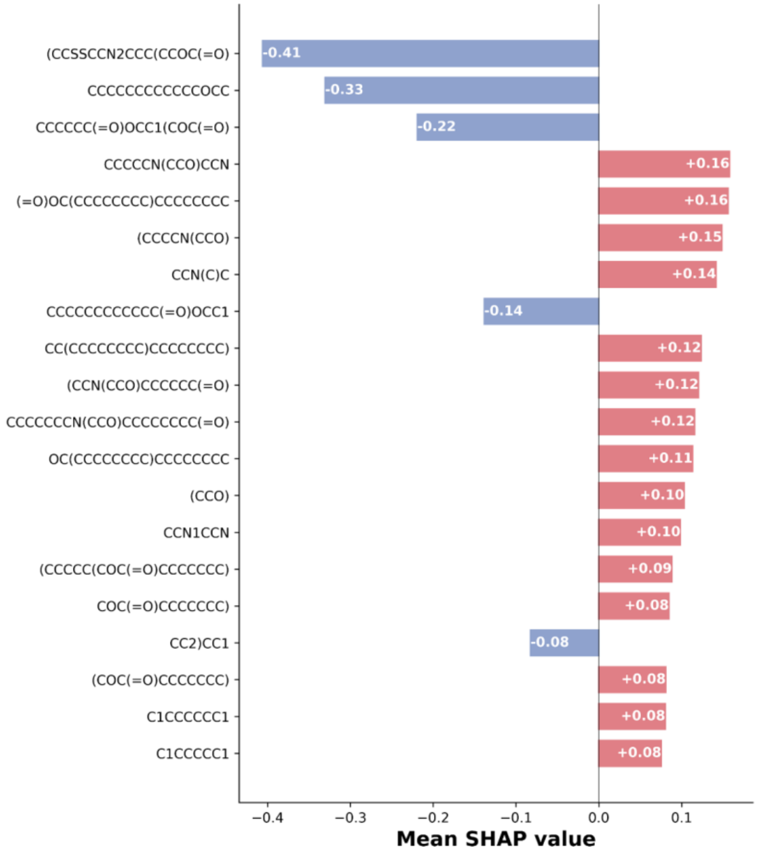




Unpublished

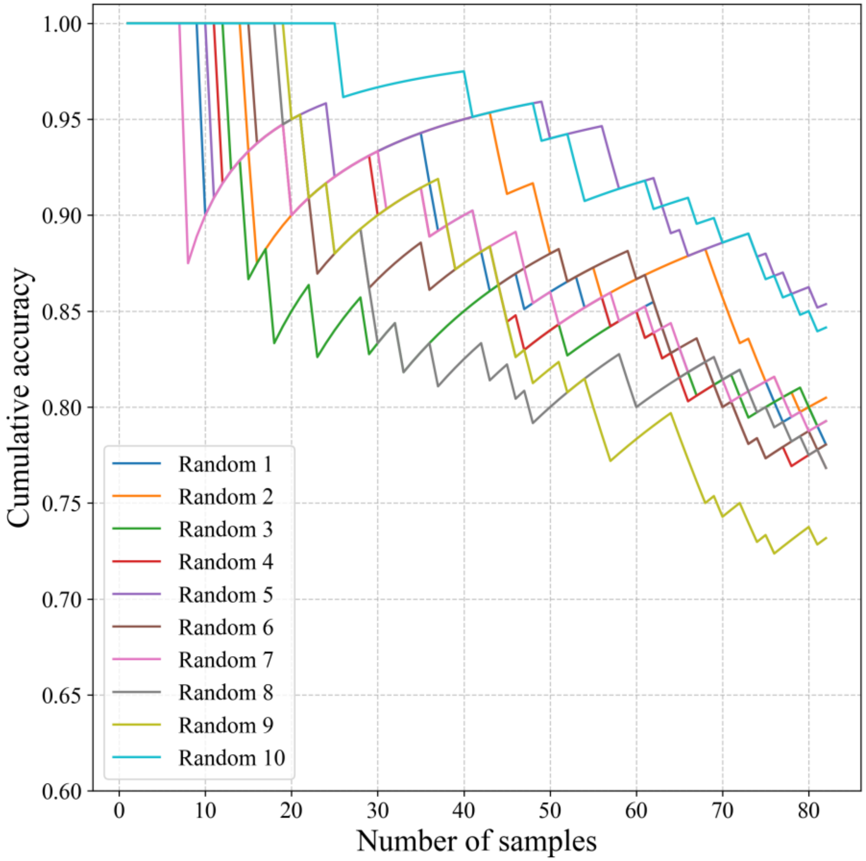
● Top 20 important substructures identified in the apparent pKa task





Unpublished

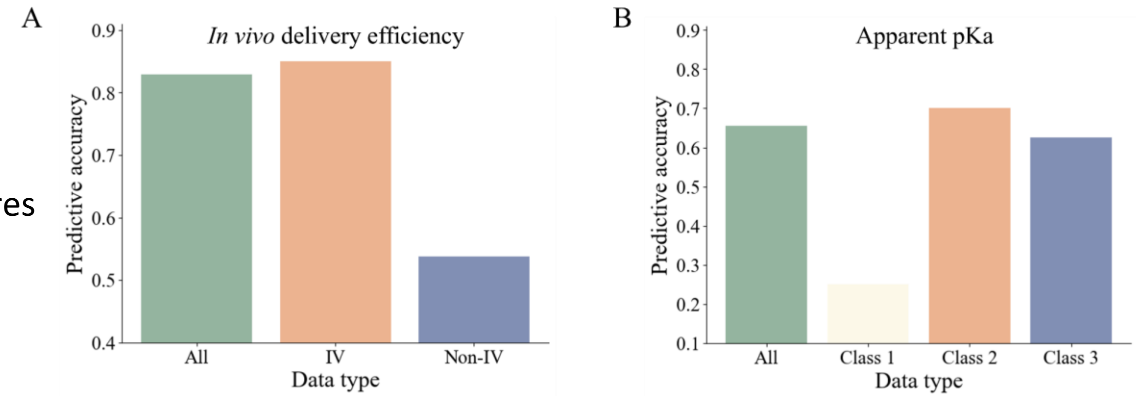
- **Prediction uncertainty quantification based on data augmentation**



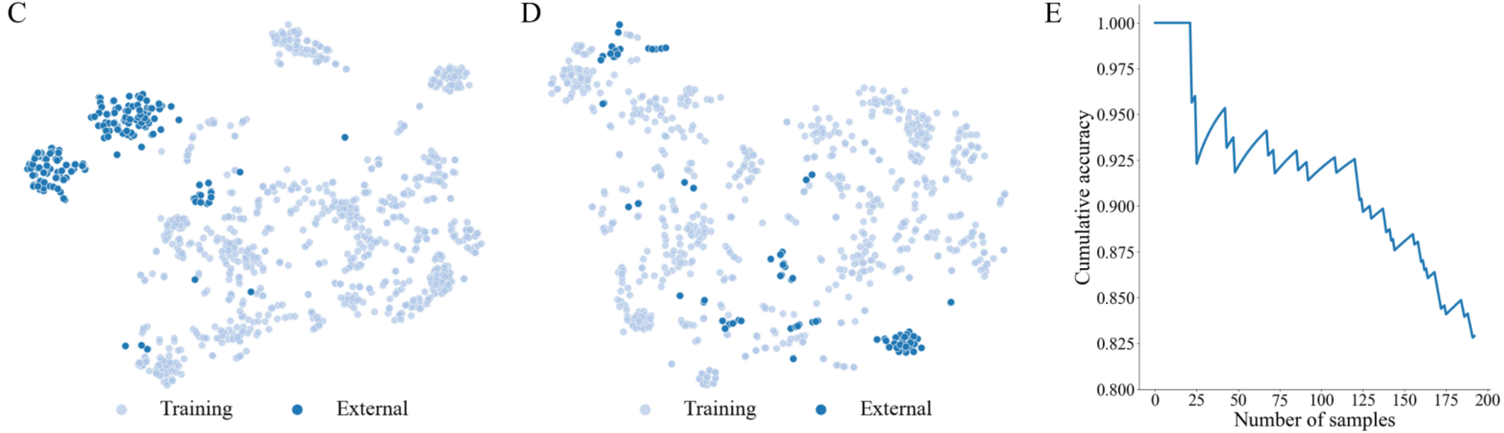
Unpublished

External validation

193 LNP formulations containing different ionizable lipid structures from 6 papers



93 LNP formulations with different ionizable lipids from 7 papers



In vivo delivery efficiency

Apparent pKa

Unpublished

➢ Deploy FormulationLNP with Streamlit

https://formulationlnp.streamlit.app/

FormulationLNP predictor

Predict in vivo delivery efficiency and apparent pKa of LNP formulations.

Model loaded successfully!

LNP Composition Parameters

Ionizable Lipid SMILES

CCCCCC=CCC=CCCCCCCCCC(CCCCCCCCC=CCC=CCCCCC)OC(=O)CCCN(C)C

LNP Component Ratios:

Ionizable Lipid Ratio (%)

50.00

Cholesterol Ratio (%)

38.50

Helper Lipid Ratio (%)

10.00

PEG-Lipid Ratio (%)

1.50

Total: 100.0%

Valid composition: Total equals 100%

Helper-Lipid Type

DSPC

PEG-Lipid Type

DMG-PEG2000

mRNA Type

Luciferase

N/P Ratio

-1.00

Mass Ratio

-1.00

Experimental Parameters

Experimental parameters don't affect the apparent pKa prediction

Mouse Strain

BALB/c

Efficiency Prediction

Predict Efficiency

Prediction Results

Mean predicted probability of delivery efficiency: 0.7533

Standard Deviation: 0.0340

Apparent pKa Prediction

Predict Apparent pKa

pKa Prediction Results

Predicted probabilities for each pKa category:

	pKa Category	Probability
0	Less than 6	0.1593
1	6 to 7	0.7962
2	Greater than 7	0.0445

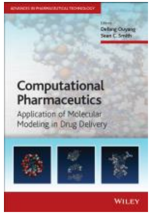
Most likely pKa category: **6 to 7** (Probability: 0.7962)

Unpublished

➢ Summary

- Explore the advanced AI techniques to facilitate the mRNA-LNP development in limited datasets
- Learn how to enhance mRNA-LNP behavior prediction with correlated properties and multi-task learning

Computational Pharmaceutics Group at University of Macau



PI: Dr. Defang Ouyang

Computational Pharmaceutics Platforms:

- FormulationAI
- FormulatuonMM
- AI-PBPK
- FormulationDT
- FormulationBCS
- PharmDE, ……

Recent Progress in AI-driven drug delivery

Preformulation prediction

- ***JCR,*** 2025, 378, 619-636
- ***Mol. Pharm.***, 2024, 22 (1), 330-342
- ***Innovation,*** 2024, 5(2): 100562;
- ***IJP***, 2021, 607, 120962;
- ***J. Cheminform.***, 2021, 13(1), 1-13

Cyclodextrin

- ***Carbohydr. Polym.***, 275 (2022): 118712;
- ***AJPS***, 2021, 16 (4), 494-507;
- ***APSB***, 2019, 9(6), 1241-1252;

Solid dispersion

- ***DDTR,*** *2024, 1-17;*
- ***IJP:X,*** *2023, 5, 100164*
- ***AAPS PharmSciTech****,2023, 24 (5), 103;*
- ***Pharmaceutics****, 202214 (11), 2257*
- ***IJP***, 2021, 604, 120705;
- ***EJPB,*** 2021, 158, 336–346;
- ***JCR***, 2019, 311-312, 16-25;

Lipid-based formulations

- ***AJPS,*** 2023, 18(3), 100811
- ***APSB***, 2021, 11(11), 3585-3594;
- ***Chem. Phys. Lett.***, 2020, 747, 137354

mRNA lipid nanoparticle

- ***Nat. Comm,*** 2024, 15 (1), 10804
- ***APSB***, 2022;12(6):2950-2962
- ***Chem. Eng. J.***, 2022, 442, 136310

Nanomedicine, microsphere and tissue engineering

- ***Burns & Trauma,*** 2024, 12, tkae053;
- ***Drug Deliv. Transl. Res.,*** 2023, 13 (4), 966-982
- ***IJP***, 2022, 628, 122302 & 122179;
- ***JCR,*** 2020, 322, 274-285;

***In vivo* prediction**

- ***Chemom. Intell. Lab. Syst.*** 2024, in press;
- ***Mol. Pharm.,*** 2024, 21 (8), 4116-4127;
- ***Brief. Bioinform***., 2024, 25 (4), bbae350;
- ***J. Cheminform.***, 2018 10, 1-11

Methodology, Platform, Review and others

- ***Brief. Bioinform***., 2024, 25 (1), bbad419;
- ***ADDR,*** 2023, 114772;
- ***Comput.-Aided Drug Des.***2023, 19 (6), 405-415;
- ***JCR***, 2021, 338, 119-136;
- ***Food Frontiers,*** 2021, 2(2), 195-207;
- ***APSB***, 2019;9(1):177–185;
- ***AJPS***, 2018,13 (4), 336-342

Acknowledgments

❖ Supervisor: Dr. Defang Ouyang

❖ Funding support institutes

❖ Lab members

❖ CRS for this opportunity

❖ Family members

Thanks for Your Attention

Questions are Welcome

Yiyang WU, Ph.D. candidate

澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

Contacts:
Ms. Yiyang WU, yiyangwu2021@163.com
Dr. Defang OUYANG, defangouyang@um.edu.mo