Xiang Zhang

(+86) 131-4992-0344 zhangxiang@simm.ac.cn Github
Drug Discovery and Design Center, 647 Songtao Road, Shanghai, China

Education

Shanghai Institute of Materia Medica & NJUCM Sept. 2021 - Jul. 2025
M.Pharm. in Drug design, GPA: 3.8/4
Advisor: Prof. Dr. Mingyue Zheng and Dr. Xutong Li

Henan University Sept. 2017 - Jul. 2021
B.S. in Pharmaceutics, GPA: 3.3/4

Research Interest

AI for synthesis: Synthetic Accessibility Prediction & Retrosynthesis Planing.
Molecular property prediction: Drug's ADMET prediction, Uncertainty estimation.
Molecule generation: Multi-objective RL optimization, Multi-modal conditional generation

Research Experience

Synthetic Accessibility Prediction Jul. 2023 - Feb. 2025

- Developed SynFrag, a fragment assembly generation model for SA prediction, desired by simulating the stepwise assembly of building blocks in synthesis.
- Curated 9.2M pre-training and 800k fine-tuning dataset, contributing two application scenario test sets. SynFrag achieved SOTA while demonstrating chemical interpretability.
- Deployed SynFrag online service: sub-second & high-throughput & interpretable prediction support for in-silico to in-lab drug design.

Molecular Property Prediction Oct. 2024 - present

- Develop BioCLIP, a progressive multi-modal bootstrapping framework that address the scarcity of tri-modal data through hierarchical training, for diverse molecular property predictions. [Code]
- Integrate 1.3K tri-modal samples (Molecules, Gene Expression, Cell Painting) with 30K bi-modal pairs; using cross-modal synthesis to generate 6K high-confidence pseudo tri-modal samples.

Organic Synthesis Jan. 2022 - Sept. 2022

- Synthesis, Purification, and Analysis of PROTAC Intermediates in Prof. Chen's Lab.

Publications

[1] **Zhang, X.**, Liu, J., Chen, K. (Academician). SynFrag: Synthetic Accessibility Predictor based on Fragment Assembly Generation in Drug Discovery. *JCIM*. [Web] | [DOI] | [Code] | [PDF]
[2] Fan, Z., Yu, J., **Zhang, X**. Reducing overconfident errors in molecular property classification using Posterior Network. *Patterns*, 2024, **5**(6): 100991. [DOI] | [Code] | [PDF]

Skills

· Programming/Software: Python, HTML, CSS, LaTeX, Prism, ChemDraw, PyMOL, Matplotlib
· Cheminformatics: Pytorch, RDkit, DGL, deepchem, sklearn, TensorBoard, GNN, Transformer
· in Laboratory: Organic synthesis, purification & analysis
· Language: English (fluent, CET6: 520), Chinese (native), Cantonese (native), Hakka (native)

Conference

· Oral Presentation: The 13th Shanghai Symposium on Computer-Aided Drug Design | 2024
· Participation: World Artificial Intelligence Conference (WAIC), Shanghai | 2022, 2023, 2025

Awards & Honors

· Academic Excellence Scholarship | 2021 - 2024
· Certificate of Honor for volunteer in COVID-19 Prevention | 2022