Yumeng Zhang

Yumeng.Zhang1@monash.edu | Github | Google Scholar | LinkedIn

Summary

- Computational biologist with expertise in machine learning for immune recognition and protein design.
- First/co-first/second author of three *Nature Machine Intelligence* papers on TCR-antigen binding specificity, mutation effects in immune recognition, as well as antigen presentation and immunogenicity.
- Proven leadership, project management skills, and contributions in interdisciplinary collaboration with a wet-lab team.

Education

Monash University 2024/09 – 2027/09 (Expected)
Ph.D. candidate in Computational Biology Melbourne, VIC, Australia
- Director's Highly Commended Publication Awards by Monash BDI (2025)
- Visiting student, Peking University | 2026/07 – Present

Shanghai Jiao Tong University 2021/09 – 2024/06
M.Sc. in Biology Shanghai, China
- GPA 3.81/4.0, National Scholarship (2022), Excellent Graduate of Shanghai (2024)
- Courses: Structural Bioinformatics, Omics Big Data, Statistical Methods in Data Science, etc.

Shanghai Jiao Tong University 2017/09 – 2021/06
B.Sc. in Bioinformatics Shanghai, China
- GPA 3.58/4.0 (86.90/100), the iGEM Competition Silver Medal (2020)
- Courses: Data Structures & Algorithms, Bioinformatics, Biostatistical Modeling, Molecular Biology, etc.

Skills

Programming Languages: Python, R, Bash, SQL
Machine Learning: PyTorch, Transformer, conditional diffusion model, graph neural network, knowledge distillation, curriculum learning, contrastive learning
Bioinformatics: protein design (TCR/VHH), protein language model, protein structure prediction, TCR/BCR repertoire analysis, TCR–pMHC binding prediction, protein stability prediction

Experience

Antigen-specific TCR Design github.com/zhangyumeng1sjtu/TCRDiff 2024/11 – 2026/04
Biomedicine Discovery Institute, Monash University Melbourne, VIC, Australia
- Develop a pMHC-conditioned discrete diffusion model for iterative generation of paired TCR CDR3αβ sequences.
- Incorporate backbone-conditioned priors into the diffusion process to improve predicted structural plausibility.
- Establish a TCR design pipeline and validate the designs targeting the MAGE-A3 epitope using luciferase and activation assays with our collaborator.
- Poster presentations at *NeurIPS Workshop AI for Drug Discovery: Bridging the Translation Gap* (2026) and *Unravelling T-cell Recognition: Insights from Immunology and AI* (2025). (see Pub. 1)

TCR-antigen Recognition Prediction github.com/zhangyumeng1sjtu/EPACT 2023/07 – 2024/09
Biomedicine Discovery Institute, Monash University Melbourne, VIC, Australia
- Predict binding specificity and interactions between T-cell receptors (TCRs) and epitopes presented by MHC-I molecules.
- Interpret spike-specific T-cell responses upon SARS-CoV-2 exposure using predicted TCR specificity and binding ranks.
- Elucidate T-cell cross-reactivity via predicted residue-level contacts and simulated peptide library screening.
- Poster presentation at *Lorne Proteins* (2025). (see Pub. 4)

Bacterial Secreted Protein Prediction github.com/zhangyumeng1sjtu/DeepSecE 2022/03 – 2023/10
State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University Shanghai, China
- Predict bacterial secreted proteins of five major types in Gram-negative bacteria.
- Explore the distribution of bacterial secretion systems and secreted proteins and provide an analytic platform.
- Oral presentation at *National Conference on Computational Biology and Bioinformatics* (2022). (see Pub. 5)

Protein Thermostability Prediction github.com/zhangyumeng1sjtu/ThermoGNN 2021/06 – 2021/09
Shanghai Artificial Intelligence Laboratory Shanghai, China

- Develop a Siamese graph attention network integrating coevolutionary and energy features to predict variations in protein thermostability upon point mutations. (see Pub. 6)
- Identify potential variants of *Streptomyces mobaraenesis* transglutaminase (smTG) with improved thermostability.

Selected Publications

1. **Y Zhang**†[1], W Liang†, S Xu, et al. (2026) Generative design of antigen-specific T-cell receptor sequences with a conditional diffusion model. *bioRxiv*, 2026.06.10.730756. [paper] *Manuscript in revision at Nat Mach Intell*
2. R Han†, **Y Zhang**†, X Liu†, L Fu†, et al. (2026) Generalizable mutation-effect prediction across adaptive immune recognition via unified multimodal framework. *Nat Mach Intell*, 8, 913–929. [paper]
3. LC Shen, **Y Zhang**, Z Wang, et al. (2025) Self-iterative multiple instance learning enables the prediction of CD4+ T cell immunogenic epitopes. *Nat Mach Intell*, 7, 1250–1265. [paper]
4. **Y Zhang**, Z Wang, Y Jiang, et al. (2024) Epitope-anchored contrastive transfer learning for paired CD8+ T cell receptor-antigen recognition. *Nat Mach Intell*, 6, 1344–1358. [paper]
5. **Y Zhang**, J Guan, C Li, et al. (2023) DeepSecE: a deep learning-based framework for multi-class prediction of secreted proteins in Gram-negative bacteria. *Research*, 6, 0258. [paper]
6. H Gong†, **Y Zhang**†, C Dong, et al. (2023). Unbiased curriculum learning enhanced global-local graph neural network for protein thermodynamic stability prediction. *Bioinformatics*, 39(10), btad589. [paper]
7. **Y Zhang**†, Y Zhang†, Y Xiong, et al. (2022) T4SEfinder: a bioinformatics tool for genome-scale prediction of bacterial type IV secreted effectors using pre-trained protein language model. *Brief Bioinform*, 23(1), bbab420. [paper]

[1] †Co-first author