

YIJIA XIAO

[Homepage](#) [Linkedin](#) [Google Scholar](#) [Email](#)

EDUCATION

- Bachelor in Computer Science and Technology, Tsinghua University** 2018 – 2022
- Ph.D. Student in Computer Science, University of California, Los Angeles** 2022 – Now
Ph.D. advisor: [Professor Wei Wang](#).

RESEARCH EXPERIENCE

- Unsupervised Contact Prediction with Pretrained MSA Language Model** June 2021 - June 2022
Research Assistant Advised by *Prof. Jie Tang* and *Prof. Jinbo Xu*
- Pre-trained one MSA language model with 1 billion parameters, exceeding *SOTA* results by 3.5%.
 - Proposed *fragment training* to trade-off between the width and depth of MSA, using only 10% of *SOTA*'s training data.
 - Demonstrated the growth in model scale contributes to capturing long-range dependencies in sequences.
- Large-scale Protein Language Model Pretraining** Dec. 2020 - Aug. 2021
Research Assistant Advised by *Prof. Jie Tang*
- Pre-trained the world's 2nd largest protein language model (*Wen Su*), improving contact prediction from 36% to 75% ([Wikipedia Page](#)).
 - Model file used by researchers from over 20 institutions, including *Harvard, UofT, NVIDIA* (statistics from BAAI).
 - Published a first-author paper, see *Paper 1*.
- Protein Mutation Effect on Protein-Ligand Binding Affinity** July 2020 - Dec. 2020
Research Intern Mentored by *Dr. Chang-Yu Hsieh*
- Predicted ligand binding affinity ($\Delta\Delta G$) changes upon protein mutations for cancer target Abl kinase.
 - Improved the Pearson and AUPAC by 8.9% and 5.6% respectively, compared with *Rosetta* baseline.
 - Combined self-paced learning (SPL) and protein family diversity loss penalty (D) as SPLD.
 - Adopted *extremely randomized regression trees* (ExtraTrees) to fit high dimension features of the compound.
 - Paper published at *BIB* (*impact factor* = 11.62), see *Paper 2*.

PUBLICATIONS

1. *Modeling Protein Using Large-scale Pretrain Language Model*.
Yijia Xiao, Jiezhong Qiu, Ziang Li, Chang-Yu Hsieh and *Jie Tang.
Pretrain Workshop, KDD 2021. [Paper](#) [Github](#) (70+ stars)
2. *SPLDExtraTrees: Robust Machine Learning Approach for Predicting Kinase Inhibitor Resistance*.
Ziyi Yang, Zhaofeng Ye, **Yijia Xiao**, Chang-Yu Hsieh, and *Shengyu Zhang.
Briefings in Bioinformatics. [Paper](#)
3. *A Deep Learning Inference Scheme Based on Pipelined Matrix Multiplication Acceleration Design and Non-uniform Quantization*.
Yuyang Zhang, Dik Hin Leung, *Min Guo, **Yijia Xiao**, Haoyue Liu, Yunfei Li, Jiyuan Zhang, Guan Wang and Zhen Chen.
IEEE CCIS 2021 (Cloud Computing and Intelligence Systems). [Paper](#)

INTERNSHIPS

- Analyst Intern, Sequoia Capital China** July 2022 - Sep. 2022
Investment field: computer infrastructure, renewable energy (e.g. photovoltaic), and AI4Science.
- Research Intern, Tencent Inc.** Sep. 2021 - May 2022
Project: modeling protein multiple sequence alignments, localized services for protein databases.
- Research Intern, Tencent Inc.** July 2020 - Dec. 2020
Project: deep learning assisted binding affinity prediction between protein receptors and drug molecules.

SELECTED AWARDS AND SCHOLARSHIPS

Awards in Tsinghua University

Research Excellence Scholarship (<i>Top 2%, 3 / 230</i>)	2021
Science and Technology Innovation Scholarship (<i>Top 2%, 3 / 230</i>)	2021
Technology Innovation Scholarship (<i>Top 5%, 6 / 230</i>)	2020
Community Contribution Award	2019

International Collegiate Programming Contests (ICPC)

<i>Silver Medal, ICPC Asia East Continent Final</i>	2020
<i>Gold Medal, ICPC Asia Regional Contest</i>	2020

Olympiad

<i>National Bronze, Chinese Physics Olympiad</i>	2017
<i>National Bronze, China Western Mathematics Invitational</i>	2017
<i>1st Prize, Chinese Collegiate Physics Olympiad</i>	2019
<i>1st Prize, China High School Biology Olympiad</i>	2017
<i>2nd Prize, China High School Chemistry Olympiad</i>	2017

LEADERSHIP AND EXTRACURRICULAR

Chinese Entrepreneurs Organization *Business Development Department* 2022 – Now

Tsinghua Future Internet Computing Club *Core member* 2020 – 2022

Tsinghua Student Entrepreneurial Talent Program *1 admission per department.* 2019 – 2021

The 46th International Conference Nov. 2020

Attendee Business Today and Princeton University

- *Citigroup* CEO and *Y Combinator* CEO delivered keynotes. ~ 200 students admitted worldwide.
- Conference topic: *Lead the Next Generation*. Track: drug innovation, smart healthcare, and food security.

The 7th China International College Students' Innovation and Entrepreneurship Competition 2020

Team Leader Prototype design and presentation

- *Purple Bamboo* (Chinese teaching chat-bot for foreigners) rated as a *national key project*.
- Won 3rd prize in the final, ranked 11th in Tsinghua.
- Project was funded \$8k; total funds: \$20k.

Tsinghua Deep Learning Summer Camp June. 2019

Volunteer

- Provided translation and cultural help to students from 10+ countries: U.S., U.K., Germany, Russia, India, Israel, Malaysia, Norway, New Zealand, Brazil, etc.

INVITED TALKS

Invited Speaker at BAAI (Beijing Academy of Artificial Intelligence) Aug. 2021

Delivered a talk on the progress and applications of pre-trained protein models to 12 AI start-ups.

Invited Speaker at BioMap Aug. 2022

Delivered a talk on efficient training of large scale pLMs (pretraining protein language models) to researchers.

SKILLS

Programming: Python, C++, Java, HTML, MySQL, Latex.

Operations and Maintenance: Manage a cluster of over 500 GPUs for large-scale pre-training.

Language: English (fluent), Mandarin (native).