Abhi K. Adduri, aadduri@cs.cmu.edu

About me

I am a 5th year Ph.D. student at CMU advised by Dr. Hosein Mohimani for the past 2 years. My work uses protein language foundation models (FMs) to identify naturally synthesized drug-like molecules encoded in microbial genomes. My research identified Edaphochelin, an antifungal with targeted activity against *Candida Auris*, a multi-drug-resistant pathogen that the CDC recently classified as an urgent health threat. Downstream of drug discovery, my work also focuses on representation learning and aligning modalities (e.g., small molecule drug and protein target) to predict drug bioactivity and mechanism of action. I am deeply interested in improving generative capabilities for large biological FMs, understanding what the models "understand", and using structured data to reduce model hallucinations, to help discover therapeutics with novel mechanisms of action.

Education

Carnegie Mellon University **Expected Fall 2024**

Ph.D. Computer Science / Computational Biology, GPA: 4.0

Ph.D. Minor Machine Learning, GPA: 3.94

Teaching Assistant: 10-701 Machine Learning, 02-613 Algorithms and Advanced Data Structures

University of California, Berkeley **May 2019**

B.A. Computer Science, GPA 3.8

B.A. Biochemistry and Molecular Biology, GPA 3.7

Programming

Rust, Python, C++, C, Docker, Bash, Typescript, HTML, CSS, Java

Frameworks

PyTorch, Transformers, Git, LaTex, PyG, React

Fellowships

Felicis Fellowship in AI **March 2024 - 2025**

Manuscripts Under Review

Adduri, A.*, Ellington, C.*, ..., Xing, E. Ultra-High-Throughput Virtual Screening for Antimicrobials. *Under review at **Machine Learning in Computational Biology.***

Ellington, C., Lengerich, B., Watkins, T., Yang, J., **Adduri, A.**, Xiao, H., Kellis, M., Xing, E. Contextualized networks reveal heterogeneous transcriptomic regulation in tumors at sample-specific resolution. *Manuscript under review at **PNAS**.*

Adduri, A., ..., Xing, E., Koes, D., Mohimani, H. Interpretable adenylation domain specificity prediction using protein language models. *Manuscript submitted.*

Behsaz, B.*, **Adduri, A.***, Guler, M.*, ..., Mohimani, H. Pathogen-oriented platform for large-scale natural product discovery identifies novel antifungal targeting drug-resistant candidiasis. *Manuscript under review at **Nature Chemical Biology**.*

Publications

Adduri, A., Kim, S. (2024). Ornaments for accurate and efficient allele-specific expression estimation with bias correction. *Manuscript accepted in **The American Journal of Human Genetics.***

Yan, D.*, Zhou, M.*, **Adduri, A.***, ..., Mohimani, H. (2024). Seq2PKS: Discovering modular type I cis-AT polyketide natural products by integrating computational mass spectrometry and genome mining. ***Nature Communications** 15.1 (2024): 5356.*

Mongia, M., Baral, R., **Adduri, A.**, ..., Mohimani, H. (2023). AdenPredictor: Accurate prediction of the adenylation domain specificity of nonribosomal peptide biosynthetic gene clusters in microbial genomes. Bioinformatics, 39(Supplement 1), i40-i46.

Adduri, A. (2019). IoT integration in manufacturing processes. In Simulation for Industry 4.0: Past, Present, and Future (pp. 129-139).

Adduri, A., Schruben, L. (2017). The Tao of Simulation. In 2017 Winter Simulation Conference (WSC) (pp. 607-616). IEEE.

* Equal contribution.

Ongoing Research

NPDiscover and MASPR **January 2022 – Present**

- Developed MASPR, which uses multi-task training and protein language models to improve small molecule binding prediction in drug-like natural product biosynthesis by over 20%
- Used MASPR to develop NPDiscover, a multi-omics algorithm to identify putative natural product drug structures encoded in microbial gene sequences - available at https://run.npanalysis.org/
- Discovered a novel antifungal, Edaphochelin, with promising results in mouse models against the multi-drug-resistant fungal pathogen *Candida Auris*, deemed an urgent threat by the CDC

Ultra-High-Throughput Virtual Screening **March 2024 – Present**

- Used attention pooling to aggregate per-residue embeddings from protein language models to improve protein target representation, and subsequently, drug target interaction (DTI) prediction
- Setup chroma store to enable fast pan-species DTI queries - approach can screen a drug against 4.3 million potential protein targets in 5ms
- Exploring contrastive learning improvements using by protein-protein interactions and drug-drug interactions to improve drug-protein co-interaction predictions

Generative Modeling of Microbial Biosynthetic Genes **May 2024 – Present**

- Developing dataset to fine-tune DNA foundation models on biosynthetic gene clusters in microbial genomes
- Investigating generation of biosynthetic gene sequences to explore naturally synthesizable drug-like molecules

Work Experience

Mozilla, Portland, OR

Software Developer Intern **May 2017 – August 2017**

- Designed and implemented Firefox Send, a tool for private, end-to-end encrypted file transfer with 13k stars on GitHub
- Firefox Send was used globally and was covered by news outlets such as Forbes and TechRepublic

Rubrik, Palo Alto, CA

Software Developer Intern **May 2018 – August 2018**

- Engineered a distributed cache system to enhance system performance and scalability, improving file access times and overall network speed by 40%
- Designed RPCs for inter-node communication to facilitate node cache management

Square / Caviar, San Francisco, CA

Data Science Engineer Intern **May 2019 – August 2019**

- Improved recall/precision of chargeback fraud-detection models by 3.5x / 2.5x respectively
- Improved pre-checkout food delivery ETA predictions for restaurants

Patents

Automated Review of Source Code for Style Issues, Rubrik 2018

Honors and Awards

Felicis Fellowship in AI Pitch Competition Winner

CMU Ph.D. in Computational Biology: Diversity, Equity and Inclusion Award

CMU SCS Graduate Student Service Award Winner

Coursework

Relevant Graduate Coursework

10-701 Machine Learning
10-708 Probabilistic Graphical Models
02-730 Cellular & Systems Modeling
10-725 Convex Optimization
10-707 Advanced Deep Learning
36-705 Intermediate Statistics
02-710 Computational Genomics
PT-2030 Intro to Computational Structural Biology

Relevant Undergraduate Coursework

CS 170 Efficient Algorithms
CS 186 Database Systems
EE 120 Signals and Systems
CS 189 Machine Learning
CS 294 Machine Learning in Biology
CS 162 Operating Systems
Math 110 Linear Algebra
EE 123 Digital Signal Processing
EE 126 Probability and Random Processes
CS 176 Algorithms in Computational Biology