Nikša Praljak, PhD Candidate

niksapraljak1@uchicago.edu | niksapraljak1@gmail.com | +1 (440) 532-3157| (**GitHub**) https://github.com/PraljakReps |(**LinkedIn**) https://www.linkedin.com/in/npraljak

SUMMARY

- Computational biologist and AI/ML expert specializing in protein design with deep generative models and multimodal language models
- Inventor on 3 AI-based patents with 2 licensed to biotech companies (Evozyne, BioChip Labs)
- Author of AI+Bio publications including 3 first-author papers in NeurIPS, Cell Systems and ACS Synthetic Biology
- Interdisciplinary researcher bridging cutting-edge AI (protein language models, diffusion models) with wet-lab experiments using high-throughput assays and next-gen sequencing technologies

EDUCATION

University of Chicago

Ph.D. Biophysics Graduate Program: Physics & Biological Sciences Division **August 2020 - Present**

Cleveland State University

B.S. Physics (Honors): Department of Physics **2020**

B.S. Mathematics (Honors): Department of Mathematics and Statistics **2020**

Summa Cum Laude (COSHP Valedictorian)

RESEARCH EXPERIENCE

The University of Chicago, *PhD Candidate in Biophysics* (NSF Graduate Fellow) **August 2020-Present**

Physical and Biological Sciences Divisions, Laboratory of Rama Ranganathan and Andrew L. Ferguson

- Developed deep generative models for molecular design, specifically variational autoencoders and autoregressive language models, and their combination to generate *de novo* proteins with PyTorch.
- Developed active learning and Bayesian optimization workflows with GPyTorch and BoTorch, leveraging NVIDIA GPUs, for property-guided design to evaluate protein functionality in the wet lab.
- Developed multimodal language and diffusion models by bridging biomedical language, protein language models, diffusion decoders, and contrastive learning to enable natural language-prompted design of proteins that function *in vivo* and *in vitro* given their functional description (BioM3).
- Scaled self-supervised (e.g., masked language models) and generative (e.g., diffusion) architectures with PyTorch, PyTorch Lightning, and DeepSpeed over hundreds of GPUs using Argonne National Lab's supercomputer clusters—Polaris and Aurora Exascale cluster.
- Developed reinforcement learning and post-training algorithms with direct preference optimization for protein language models on large-scale mutational effect data.
- Developed LLM agent for optimizing prompts to improve text-to-protein generation of BioM3.
- Synthesized and assembled large-scale DNA libraries that encode protein designs based on generative models using PCR and oligo pools.
- Conducted high-throughput experimental validation of *in vivo* function using next-generation sequencing (Illumina) and selection assays.
- Purified *de novo* protein sequences and conducted biochemical measurements for binding with fluorescence-based assays and thermal stability with isothermal calorimetry.

Case Western Reserve University, *Visiting Undergraduate Researcher* **August 2018-2020**

- Processed large amounts of brightfield images and video of sickle cell dynamics within microfluidic devices using OpenCV, Pillow, scikit-image, and torchvision in PyTorch.
- Developed a deep learning-based model with U-Net architecture and trained with distributed GPU parallelism for tracking, segmenting, and classifying millions of adhered sickle and normal red blood cells with various morphologies for pathological detection.

Cleveland State University, *Undergraduate Researcher / NSF REU Researcher* **August 2016-2020**

- Developed numerical simulations with fluid-structure interaction algorithms and finite element analysis to study renal tubule fluid flow in various disease states.

TECHNICAL SKILLS

Programming & Computing: Python, C++, R, MATLAB, bash, Linux, HPC environments, GPU clusters

Machine Learning & AI: PyTorch, DeepSpeed, HuggingFace, GPyTorch, BoTorch, TensorFlow, distributed training, protein language models (ESM2), large language models (Llama3), AlphaFold, ESMFold

Data Analysis & Visualization: NumPy, Pandas, Matplotlib, SciPy, OpenCV, scikit-image

Bioinformatics & Molecular Biology: Biopython, BLAST, MMSeqs2, Foldseek, next-generation sequencing (Illumina), PCR, protein purification, fluorescence assays

SELECT PUBLICATIONS (link: Google Scholar)

Praljak N, Yeh H, Moore M, Socolich M, Ranganathan R, Ferguson AL. Natural Language Prompts Guide the Design of Novel Functional Protein Sequences. *AIDrugX Workshop @* ***NeurIPS***, doi:10.1101/2024.11.11.622734 (2024).

Lian X*, **Praljak N***, Subramanian SK*, Wasinger S, Ranganathan R, Ferguson AL. Deep-learning-based design of synthetic orthologs of SH3 signaling domains. ***Cell Systems*** 15, 725-737 (2024).

- Featured in: Fu X. How deep can we decipher protein evolution with deep learning models. ***Patterns*** 5, 101043 (2024).

Praljak N*, Lian X, Ranganathan R, Ferguson AL. ProtWave-VAE: Integrating autoregressive sampling with latent-based inference for data-driven protein design. ***ACS Synthetic Biology*** 12, 3544-3561 (2023).

- Featured in: Martín García H, Mazurenko S, Zhao H. Special Issue on Artificial Intelligence for Synthetic Biology. ***ACS Synth. Biol***. 13, 408-410 (2024).

MANUSCRIPTS IN SUBMISSION/REVIEW

Praljak N, Yeh H, Hwang SW, Berlaga A, Liu A, Ferguson AL. Aligning Protein Language Models with Large-Scale Mutational Preferences. *Submitted* 2025.

Praljak N. Ferguson AL. RiP^2: Reinforcement-Informed Prompting for Proteins. *Submitted* 2025.

PATENTS

Ferguson AL, **Praljak N**, Ranganathan R. Techniques for Artificial Intelligence (AI) Based Protein Engineering Using Natural Language Prompting. U.S. Patent Application Attorney Docket Number: 27373/70285P. Application Submitted.

Ferguson AL, **Praljak N.** System, Method, and Computer Readable Storage Medium for Auto-Regressive Wavenet Variational Autoencoders for Alignment-Free Generative Protein Design and Fitness Prediction. U.S. Patent Application 18/176,375 (2023).

- Licensed to Evozyne.

Praljak N, Shamreen I, Goreke U, Hinczewski M, Hill A, Gurkan U, Singh G. Classification of Blood Cells. U.S. Patent Application 17/928,976.

- Licensed to BioChip Labs.

SELECT CONFERENCE PRESENTATIONS

(Oral) Praljak N. "Natural language prompts guide the design of novel functional protein sequences" Invited talk for PhD Student Research Day at the Data Science Institute, University of Chicago, December 2024

(Oral) Praljak N., "A multimodal generative model with natural language for protein design" Multimodal AI Workshop, Toyota Tech Institute at Chicago (TTIC), January 2024.

(Oral) Praljak N., "Data-driven protein design" Nominated candidate at the Grier Prize Symposium, University of Chicago, December 7th, 2021