

# Jordan Chen, Ph.D.

jordan.chen@email.com | +1 (555) 123-4567

Richland, WA (Open to relocation to Westbrook, MA)

Computational Protein Engineering | ML-Guided Enzyme Design | Biomolecular Simulation

## Summary

---

Computational biologist with 8+ years combining **protein language models** and **molecular dynamics simulations** for enzyme engineering and drug discovery. Fine-tuned ESM-2 on 45K experimental stability measurements to screen 8,500 enzyme variants at 3,000× experimental throughput, with 5 hits confirmed by collaborators. Co-developed open-source transfer learning framework adopted by 4 external groups. 15 publications (7 first-author) in ACS Catalysis, J. Chem. Theory Comput., and J. Med. Chem.

## Technical Skills

---

### Molecular Simulation & Modeling

- **GROMACS**, OpenMM, AMBER – metadynamics, replica exchange MD, free energy perturbation, umbrella sampling
- AlphaFold2, Rosetta, AutoDock Vina – protein structure prediction, homology modeling, and molecular docking
- CHARMM36m, AMBER ff19SB, OPLS-AA/M – force field selection and benchmarking for disordered proteins
- Collective variable design, enhanced sampling protocol development, convergence analysis, MM/PBSA free energy

### Machine Learning & Data Science

- **Protein language models** (ESM-2, 650M params), graph neural networks, transfer learning, active learning loops
- **PyTorch**, scikit-learn, BioPython – model fine-tuning, embedding extraction, feature engineering, sequence analysis
- Regression, classification, cross-validation, Spearman/RMSE benchmarking, dataset curation from public databases

### Programming & HPC

- **Python**, Bash, SQL – scientific computing, data pipelines, automated analysis workflows, database management
- **SLURM**, Snakemake, Git, DVC – HPC job scheduling, workflow automation, version control, reproducible research

### Analysis & Visualization

- MDAAnalysis, ProDy, PyMOL, matplotlib, seaborn – trajectory analysis, structural visualization, publication figures
- PostgreSQL, pandas, NumPy – curated stability databases with automated quality filters for ML pipelines

### Domain Expertise

- Protein engineering, enzyme thermostability, folding thermodynamics, drug discovery, virtual screening workflows
- Intrinsically disordered proteins, ligand binding free energy, biocatalysis, directed evolution, rational design

## Research Experience

---

### ML-Accelerated Protein Engineering and Computational Enzyme Design

Aug 2023 – Present

*Postdoctoral Research Associate, Lakewood University*

- Fine-tuned ESM-2 protein language model on 45K experimental melting temperatures, achieving 0.82 Spearman correlation and enabling 3,000× throughput screening of 8,500 enzyme variants for industrial thermostability.
- Co-developed transfer learning framework from protein language models reducing labeled training data by 60% across 5 enzyme families, released as open-source tool with 200+ GitHub stars.
- Extended protein language model to predict enzyme solvent tolerance across 8 organic co-solvent systems, validating against 50-ns explicit-solvent MD for 80 enzyme variants and identifying 4 candidates for green chemistry.
- Automated sequence-to-simulation pipeline using Snakemake workflow manager, reducing per-variant setup from 4 hours to 10 minutes and supporting 6 researchers across 3 active projects.
- Revealed sequence-dependent enzyme unfolding pathway divergence at 340 K through 200-ns replica exchange MD simulations, identifying stabilizing salt bridge networks that informed rational design criteria.
- Mentored 3 graduate students on protein ML pipelines and MD simulation workflows, with 1 student co-authoring a peer-reviewed publication within 8 months of joining.

### Enhanced Sampling Methods for Protein Folding and Ligand Binding

Aug 2018 – Jul 2023

*Ph.D. Researcher, Westfield Institute of Technology*

- Developed metadynamics-based enhanced sampling protocol for protein folding free energy landscapes, predicting folding temperatures within 8 K of experiment across 6 small proteins.
- Calculated relative binding free energies for 40 congeneric ligand pairs via free energy perturbation, achieving 0.9 kcal/mol RMSE against experimental IC<sub>50</sub> data across 3 drug target families.
- Built curated protein thermostability database integrating 12,000 experimental melting temperatures from 3 public sources, with automated quality filters adopted by 8 lab members for ML training set construction.

- Benchmarked 4 protein force fields on 15 intrinsically disordered protein sequences, establishing CHARMM36m as the optimal choice for IDP conformational ensemble prediction with 40% better agreement with SAXS data.

## Computational Biophysics and Structural Analysis

May 2016 – Jul 2018

*Undergraduate Research Assistant, Eastgate University*

- Performed homology modeling and 100-ns MD simulations of 4 mutant lysozyme variants, identifying destabilizing cavity mutations consistent with published experimental unfolding data.
- Built Python analysis scripts for automated hydrogen bond occupancy tracking across 500-ns aggregate trajectories, adopted by 3 lab members for ongoing protein stability projects.

## Education

Ph.D., Biomedical Engineering Westfield Institute of Technology, Westfield, MA	Aug 2018 – Jul 2023 GPA: <b>3.92/4.00</b>
B.S., Biochemistry (Honors) Eastgate University, Portland, OR	Aug 2014 – May 2018 GPA: <b>3.87/4.00</b>

## Selected Publications (15 papers | 280+ citations)

- **J. Chen**, R. Nakamura, S. Patel, K. Holmberg, M. Rivera. “Deep Learning-Guided Screening of Thermostable Enzyme Variants for Industrial Biocatalysis.” *ACS Catalysis*, 2025.
- **J. Chen**, M. Rivera, K. Holmberg. “Transfer Learning from Protein Language Models for Low-Data Enzyme Property Prediction.” *Bioinformatics*, 2024.
- **J. Chen**, L. Alvarez. “Ligand Binding Free Energies via Enhanced-Sampling FEP for Three Drug Target Families.” *J. Med. Chem.*, 2023.
- **J. Chen**, L. Alvarez. “Metadynamics Protocol for Protein Folding Free Energy Landscapes.” *J. Chem. Theory Comput.*, 2022.
- **J. Chen**, P. Kowalski, L. Alvarez. “Force Field Benchmarking for Intrinsically Disordered Protein Ensembles.” *J. Chem. Theory Comput.*, 2021.
- **J. Chen**, P. Kowalski, L. Alvarez. “Curated Thermostability Database for ML-Ready Protein Engineering Benchmarks.” *Bioinformatics*, 2021.
- **J. Chen**, T. Yamamoto, K. Holmberg. “Predicting Enzyme Solvent Tolerance with Fine-Tuned Protein Language Models.” *Proteins: Struct., Funct., Bioinf.*, 2025 (under review).

## Honors & Awards

- **NSF Graduate Research Fellowship**, National Science Foundation (2019). Three-year fellowship supporting doctoral research in computational protein engineering.
- **Best Oral Presentation**, Westfield Biophysics Symposium (2022). Enhanced sampling methods for protein folding thermodynamics.
- **Dean’s Teaching Award**, Westfield Institute of Technology (2021). Outstanding TA in computational biology.
- **Outstanding Poster Award**, Gordon Research Conference on Proteins (2023). ML-guided enzyme thermostability screening.

## Selected Presentations

- “ML-Guided Screening of Thermostable Enzyme Variants.” *Gordon Research Conference on Proteins*, Ventura, CA (2023). Poster.
- “Transfer Learning from Protein Language Models for Low-Data Enzyme Property Prediction.” *ACS National Meeting*, San Francisco, CA (2024). Oral.
- “Enhanced Sampling Protocols for Protein Folding Free Energy Landscapes.” *Biophysical Society Annual Meeting*, San Diego, CA (2022). Oral.
- “Force Field Benchmarking for Intrinsically Disordered Protein Ensembles.” *AIChE Annual Meeting*, Boston, MA (2021). Poster.
- “Automated Screening Pipelines for ML-Guided Enzyme Engineering.” *Lakewood University Computational Biology Seminar Series* (2024). Invited talk.

*Authorized to work in the United States*