Shrishti Barethiya

Amherst, Massachusetts, USA

sbarethiya@umass.edu www.linkedin.com/in/shrishti-barethiya

Education

Ph.D. in Computational Chemistry 2022 – Expected 2027

Department of Chemistry, University of Massachusetts Amherst

Advisor: Dr. Jianhan Chen

Focus: Physics-informed ML for sequence-to-function prediction, Generative modeling of dynamic protein conformations, multi-scale simulation of intrinsically disordered proteins.

Integrated M.Sc. degree 2017-2022

School of Chemical Sciences, National Institute of Science Education and Research Bhubaneswar, Odisha, India

Recipient of DISHA Scholarship, Department of Atomic Energy (Govt. of India)

Advisor: Dr. Upakarasamy Lourderaj

Focus: Machine learning for potential energy surface representation (potential energy surface) of small molecules.

Technical Skills

- **Machine Learning**: Deep learning (CNN, GNN, VAE, latent DDPM, Transformers, DiT), dimensionality reduction, clustering, traditional ML, PyTorch, JAX, TensorFlow, scikit-learn
- **Molecular Simulation**: All-atom (CHARMM, GROMACS, OpenMM) and coarse-grained modeling and simulation
- **Computational Modeling Tools**: Rosetta, PyRosetta, MDAnalysis, Modeller, RDkit, VMD, PyMOL
- **Programming**: Python, Bash, Git
- **Data Analysis** & **Visualization**: NumPy, Pandas, Matplotlib, Seaborn, Origin
- **Computing Environments**: Linux (Ubuntu), Windows, MacOS, HPC clusters (SLURM)

Research Experience

Ph.D. Researcher, Computational Biophysics Lab, University of Massachusetts Amherst, MA

- Physics-informed Machine learning for Protein Sequence-to-Function Prediction
  - Developed CNNs and GCNs incorporating biophysics-based features for a robust protein variant effect prediction. Achieved significantly high performance for positional extrapolation.
  - Led collaborative project to integrate sequence, evolutionary, structural, and biophysical features to identify resistance-conferring mutations in Mycobacterium tuberculosis, resulting in a resistance forecast model.
- Generative Ensemble Modeling of Intrinsically Disordered Protein
  - Trained latent diffusion transformer models on a coarse-grained (CG) molecular dynamics dataset (∼2.6M disordered protein conformations) to generate conformational ensembles for IDPs. Achieved a generative model that provides accurate secondary and global structure predictions.
  - Designed and trained new latent diffusion transformer-based generative model using ESM embedding and a folded protein structure database, aimed at improving accuracy beyond the CG disordered protein conformation dataset.

Master's thesis, Theoretical and Computational Chemistry Lab, National Institute of Science Education and Research Bhubaneswar, India

- Machine Learning for Chemical Dynamics Simulations
  - Applied neural networks and Gaussian process to represent potential energy surface of a three-dimensional system, enabling scalable modeling of molecular dynamics.
  - Compared neural network approaches with traditional fitting methods (non-linear least squares, cubic splines) for small-molecule potential energy surface representation.

- Applied ANN models to predict outcomes of classical trajectory simulations, enabling sampling of states across higher energy barriers.

Publication

*: co-first author

5 M. Tasmin, **S. Barethiya**, Y. Wang, L. Kang, A. Green, and J. Chen, Forecasting Antibiotic Resistance in *Mycobacterium tuberculosis* using Biophysics and Machine Learning (in preparation).

4 **S. Barethiya**, J. Huang, H. Guan, and J. Chen, Overcoming Extrapolation Challenges of Deep Learning by Incorporating Physics in Protein Sequence-Function Modeling (submitted).

3 **S. Barethiya***, S.Schultz*, Y. Zhang, and J. Chen, Coarse-Grained Simulations of Phosphorylation Regulation of p53 Autoinhibition. *Biochemistry* (2025).

2 E. B. Nordquist*, G. Zhang*, **S. Barethiya**, N. Ji, K. M. White, L. Han, Z. Jia, J. Shi, J. Cui, and J. Chen, Incorporating physics to overcome data scarcity in predictive modeling of protein function: a case study of BK channels. *PLoS Comput. Biol.* (2023)

1 L. Zheng*, **S. Barethiya***, E. B. Nordquist, and J. Chen, Machine learning generation of dynamic protein conformational ensembles. *Molecules* (2023)

Selected Talks and Posters

- Presented poster in University of Massachusetts Amherst, ResearchFest 2025, *Generative Ensemble Modeling of Intrinsically Disordered Protein*.
- Presented poster in Biophysical Society Annual Meeting 2025, *Overcoming Extrapolation Challenges of Deep Learning by Incorporating Physics in Protein Sequence-Function Modeling*.
- Presented a talk in University of Massachusetts Amherst, Chemistry-Biology Interface 2025, *Physics-informed machine learning for protein structure and function*.
- Presented poster University of Massachusetts Amherst, ResearchFest 2024, *Combining Physics and Machine Learning for Antibiotic Resistance Forecast*.