Swarnadeep Seth

Research Associate, Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia
Ph.D. in Physics, University of Central Florida

(509) 913-3602 swarnadeepseth@vt.edu Website LinkedIn Github google-scholar

Professional Profile

• A computational researcher with 7+ years of expertise in large-scale atomistic **Molecular Dynamics (MD)** simulation, coarse-grained modeling of proteins, glycans, biomaterials, and **force-field development**. Utilizing **ML**, **DL**, and **NLP** techniques to generate proteins with binding specificity and **design functional glycomaterials** incorporating **Generative AI** models.
• Developer and maintainer of **GlycoData web server** facilities under GlycoMIP, an **NSF Material Initiative** platform (22.9 million USD) for **novel glycomaterial** discovery with **500+ users** across the globe.

Work Experience

Virginia Tech, VA **Dec 2023 - Present:** *Research Associate*
• Utilizing auto-encoders and natural language processing techniques to enhance all-atom MD simulations, facilitating the design of functional glycomaterials with customizable rigidity and lubrication properties.
• Developed optimized force field parameters for modeling folded and disordered proteins within a coarse-grained framework. Achieved statistical properties comparable to all-atom simulations while significantly reducing the computational time.
• Created GlycoGPT, a chatbot powered by fine-tuned LLMs, adept in digesting glycomaterial-related articles via GlycoFetch, an automated software for retrieving articles from journal websites found in Google Scholar searches, thus broadening access to organized research data worldwide and fortifying NSF's Material Discovery initiative.

University of Central Florida, FL **Aug 2018 - Dec 2023:** *Graduate Research Assistant*
• Validated and optimized a coarse-grained model for intrinsically disordered proteins, matching the experimental radius of gyrations. Developed a neural network using CG model training data to predict the IDP radius of gyration with 95% accuracy, enabling analysis of point mutation impacts and reducing search space by 1000-fold. Identified crucial regions and amino acid mutations affecting disease progression.
• Developed in-house MD code in C capable of electrophoretic modeling of DNA capture and translocation through solid-state nanopores, aiding in sub-nanometer-scale insights for cost-effective, chemical-free nanopore sequencing device design.
• Proposed patented numerical methods to construct an accurate genomic map of oligo-flap markers utilizing the double nanopore setup and verified using the experimental dataset from Nooma-Bio.

Los Alamos National Laboratory, NM **May 2020 - August 2020:** *Research Intern*
• Utilized *ab-initio* calculations to analyze the effect of dipole moment on the growth of W-Ge on hBN-Graphene heterostructure.

Education

Ph.D in Physics, University of Central Florida **2018-2023:** *Best Dissertation Award*
Thesis: "DNA Capture and Translocation Through Nanopore".

BS-MS dual degree, IISER Kolkata, India **2013-2018:** *DST Inspire Fellowship*

Honors and Awards

• **Best Ph.D. Dissertation Award** from College of Science (COS), UCF in Fall 2023.
• UCF **Research Symposium Award** in Spring 2022.
• Recipient of the UCF **Graduate Research Week Best Poster** Presentation award in 2021.
• **Center for Nonlinear Studies (CNLS) Fellowship** from Los Alamos National Laboratory in 2019.

Technical and Leadership Skills

Generative AI and Machine Learning Methods with MD Simulation (5+ years):

- Coupled language models, VAE, and GAN with MD simulations to design functional glycomaterials with specific lubricant and rigidity properties.
- Conducted large-scale, all-atom biomolecular simulations of proteins, lipids, and hybrid glycomaterials utilizing Charmm and Amber force fields within NAMD, GROMACS, and OpenMM.
- Optimized coarse-grained potentials for amino acids and carbohydrates to facilitate microsecond MD simulations and accelerated glycomaterial discovery.
- Developed parallel C/C++ code for long biopolymers (DNA, RNA, Proteins) using Langevin/Brownian dynamics. Implemented non-local electric field in nanopore geometry using Finite Element Analysis.

Webserver Development and Database Management (4+ years)

- Using Django and AWS S3 cloud storage to maintain the GlycoData website and server, and created a user data sharing platform.
- Utilizing SQL and MongoDB databases to store and retrieve simulation and experimental datasets for GlycoMIP, NSF Material Initiative Platform.

Leadership and Interpersonal Skills:

- Showcased exemplary communication skills, adeptly engaging diverse audiences. Collaborated seamlessly with biostatisticians, experimental scientists, and stakeholders from Nooma Bio, McGill University, Virginia Tech, Brandeis University, and various national labs in NIH/NSF-funded projects.
- Demonstrated robust problem-solving abilities within a results-driven team, mentoring and guiding four master's students to publication success, emphasizing effective mentorship. Embraces risk, accountability, and a continuous learning mindset.

Coding and Software

Machine Learning: Efficient in implementing scalable machine learning and deep learning codes using scikit-learn, PyTorch, and TensorFlow.

- Supervised Learning: Linear and Logistic Regression, SVM, Random Forest, Gradient-Boosting Algorithms, and Neural Networks (ANN, CNN, RNN, GNN, LSTM).
- Unsupervised Learning: PCA, and Clustering Algorithms.
- Statistics: Parameter Estimation, Hypothesis Testing, Model Validation.

Generative AI: Variational auto-encoder (VAE), GAN, NLP, and LLMs.
Programming Languages: Proficient in C, C++, Python (NumPy, SciPy, Pandas, Modin, Selenium, Plotly, Streamlit, Dash), Javascript, and Shell script.
Molecular Dynamics Software: LAMMPS, HOOMD, ESPResSo, NAMD, GROMACS, OpenMM.
Docking Simulation: AutoDock, RosettaDock.
Density Functional Theory: VASP, Material Studio, VESTA.
Parallel Programming: OpenMP in HPC and avid user of INTEL's oneAPI Toolkit.
Web Development: Django, Flask, Angular JS.
Databases and Cloud: MySQL, MariaDB, MongoDB, Google Cloud, AWS S3.
Other Scientific Software: MATLAB, Mathematica, Gmsh, FEniCS.
Familiar with Git/GitHub, Microsoft Excel, Power BI, Android Studio, and Inkscape.

Recent Publications, Book Chapters, and Patents

1. **Seth S.**, Stein B., and Bhattacharya A. "Fine Structures of Intrinsically Disordered Proteins." *J. Chem. Phys.*, 2024, 160, 014902.
2. Bair J., **Seth S.**, and Bhattacharya A. "Universality in conformations and transverse fluctuations of a semi-flexible polymer in a crowded environment." *J. Chem. Phys.*, 2023, 158, 204902.
3. **Seth S.**, Rand A., Reisner W., Dunbar W. B., Sladek R., and Bhattacharya A. "Discriminating protein tags on a dsDNA construct using a Dual Nanopore Device." *Sci Rep*, 2022, 12 (1), 11305.
4. Bhattacharya A. and **Seth S.** "Methods of determining dna barcodes for efficient species categorization using nanopore translocation." – *US Patent Application* 17/649,577, 2022.

• Full publication list: google-scholar