Rishal Aggarwal

@AggarwalRishal | rishal.aggarwal@gmail.com | Rishal Aggarwal | RishalAggarwal

Research Interests: Machine learning methods for accelerated molecular screening, sampling, and design.

Education

CMU-Pitt Computational Biology — Pittsburgh, USA
PhD Student — *Aug. 2022 – Present*

International Institute of Information Technology — Hyderabad, India
MS by Research in Bioinformatics — *Jan. 2021 – July 2022*

Birla Institute of Technology and Sciences, Pilani — Rajasthan, India
Bachelor of Pharmacy — *Aug. 2016 – May 2020*

Academic and Research Achievements

- Authored multiple publications in top venues including submissions to **NeurIPS**, **ICLR**-Blogposts, **ICML**-Genbio, **ICML**-SPIGM, **ICML**-CompBio, ACS JCIM, Wiley, BMC Biology etc; recognized with **1400+ citations** and an **h-index of 9**. Google Scholar link.
- **Contributions to Machine Learning:** Developed novel, scalable simulation-free methods for training energy-based models. Achieved amortized Boltzmann sampling using flow matching and EBMs, and conducted one of the first comprehensive evaluations of leading energy-trained Boltzmann samplers.
- **Contributions to AI in Chemistry:** Contributed to several projects significantly including the widely used molecular docking software **GNINA**. Co-developed two novel pharmacophore elucidation methods: PharmRL (reinforcement learning–based) and Pharmacoforge (diffusion/flow matching–based). **PharmRL** was recognized with a **Spotlight Poster** at the MLDD Conference, MIT - Broad Institute 2023.

Selected Publications

- **Rishal Aggarwal**, Jacky Chen, Nicholas M. Boffi, and David R. Koes. BoltzNCE: Learning Likelihoods for Boltzmann Generation with Stochastic Interpolants and Noise Contrastive Estimation. *Sumbitted NeurIPS 2025, ICML - GenBio 2025*
- **Rishal Aggarwal**, Daniel Peñaherrera, Justin Shao, Minhyek Jeon, David R. Koes. Models trained with unnormalized density functions: A need for a course correction. *ICLR - Blogpost Track 2025*
- Anish Sevekari*, **Rishal Aggarwal***, David R. Koes, Maria Chikina. Accelerating NCE Convergence with Adaptive Normalizing Constant Computation. *ICML - Structured Probabilistic Inference & Generative Modeling 2024*
- **Rishal Aggarwal**, and David R. Koes. PharmRL: Pharmacophore elucidation with deep geometric reinforcement learning. *Springer BMC Biology, 2024*
- **Rishal Aggarwal**, Akash Gupta, and U. Deva Priyakumar. Apobind: a dataset of ligand unbound protein conformations for machine learning applications in de novo drug design. *ICML - Computational Biology 2021*
- **Rishal Aggarwal**, Akash Gupta, Vineeth Chelur, C. V. Jawahar, and U. Deva Priyakumar. DeepPocket: Ligand Binding Site Detection and Segmentation using 3D Convolutional Neural Networks. *Journal of Chemical Information and Modeling,* 2021.
- Viraj Bagal, **Rishal Aggarwal**, P. K. Vinod, and U. Deva Priyakumar. MolGPT: Molecular Generation Using a Transformer-Decoder Model. *Journal of Chemical Information and Modeling,* 2021.
- Andrew T McNutt, Paul Francoeur, **Rishal Aggarwal**, Tomohide Masuda, Rocco Meli, Matthew Ragoza, Jocelyn Sunseri, and David Ryan Koes. GNINA 1.0: molecular docking with deep learning. *Journal of cheminformatics,* 13(1):1–20, 2021.

Research Experience

Graduate RA / PhD student - Koes Group — August 2022 – Present
CMU - Pitt Computational Biology — *Pittsburgh, USA*

- **Amortized Boltzmann Sampling (BoltzNCE):** Developed a flow-matching and EBM-based method for efficient Boltzmann sampling of molecular systems. Achieved accurate estimation of free energy surfaces and energy histograms with orders-of-magnitude speedup over Boltzmann Generators and MD simulations. *Published at ICML-GenBio 2025; submitted to NeurIPS 2025.* **Link**

- **Boltzmann Sampling — Generative Models vs. MCMC:** Performed an in-depth evaluation of generative flow matching and diffusion models trained with energy functions, benchmarking them against MCMC approaches for Boltzmann sampling. Identified key insights on **transferability** and **domain adaptation**, outlining promising future research directions. *Published in ICLR Blogpost Track 2025*. **Link**
- **Accelerating Noise Contrastive Estimation (NCE):** Analyzed the impact of the NCE objective on the partition function parameter and developed a novel optimization routine leveraging **Bennett's Acceptance Ratio**, achieving significantly faster convergence. *Published at ICML — Structured Probabilistic Inference & Generative Modeling 2024*. **Link**
- **Pharmacophore elucidation with Generative models:** Contributed significantly to the development of pharmacophore generation on protein binding sites using diffusion/flow matching. *Published Frontiers in Bioinformatics*. **Link**
- **Pharmacophore elucidation with reinforcement learning:** Developed a framework for pharmacophore elucidation on protein binding pockets using geometric SO(3)-equivariant neural networks and Deep-Q Learning. *Published Spring BMC Biology 2024*. **Link**
- **Molecular Docking with Deep Learning (GNINA):** Contributed extensively to the development and validation of a widely adopted deep learning–based docking framework, now applied across numerous drug discovery projects. *Published in Journal of Cheminformatics, 2021 and 2024*. **Link**

Graduate Research Assistant - Deva Lab May 2020 – August 2022

Center for Computational Natural Sciences and Bioinformatics *IIITH, India*

- **Binding Pocket Identification (DeepPocket):** Trained Convolutional Neural Networks to detect and segment ligand binding sites on 3D protein structures, achieving state-of-the-art performance. *Published in ACS JCIM*. **Link**
- **Molecular Generation with Transformers (MolGPT):** Demonstrated the effectiveness of attention-based models for molecular generation, enabling inverse design of molecules conditioned on user-defined scaffolds and property values. *Published in ACS JCIM*. **Link**
- **APOBind — A Dataset of Apo–Holo Structures:** Created a benchmark dataset of unbound (apo) protein structures corresponding to protein–ligand complexes in PDBbind, enabling improved evaluation of binding affinity prediction and binding site detection methods under realistic apo conditions. Demonstrated that popular docking algorithms perform poorly on this dataset, underscoring its value for advancing method development. *Published in ICML -Computational Biology Workshop 2021*. **Link**
- **Hidden Biases in Binding Affinity Datasets:** Investigated systematic biases in protein–ligand datasets that influence the performance of ML models, providing a more holistic assessment of deep learning approaches for virtual screening. *Published in ACS Omega*. **Link**

Relevant Coursework

Machine Learning: PhD level grounding in methods, mathematics and algorithms used in Machine Learning
Probabilistic Graphical Models: Undirected, directed graphical models, sampling methods and deep generative models
Statistical Mechanics: Stastical theory behind equilibrium thermodynamics, entropy, free energy, phase transitions etc.
Data Structures and Algorithms: Graduate level course on data structures and common algorithms used for problem solving

Technical Skills

Languages: Java, Python, C/C++, R, Matlab
Deep Learning tools: Tensorflow, Keras, Pytorch
Bioinformatics tools : RDkit, OpenBabel, Biopython, Prody, Pymol
Utility Tools: Git, Linux, MS office, Photoshop