

Develop Generative AI for Functional Protein Design

Research Opportunities for CS Undergraduate and Graduate Students

The [Linna An Lab at Rice University](#) is recruiting motivated Computer Science undergraduate and graduate students to develop the next generation of **generative machine-learning models for biomolecular design**.

Modern generative models can create images, text, and software. Our goal is to build models that generate **functional proteins, enzymes, molecular sensors, and drug-binding proteins**^{1–3}. These systems must reason across protein sequence, three-dimensional structure, molecular interactions, conformational flexibility, and experimental measurements—creating fundamental and highly challenging problems in machine learning.

Potential Research Projects

Students may work on projects involving:

- Generative models for protein sequence³ and three-dimensional structure design⁴
- Diffusion, flow-matching, autoregressive, and multimodal architectures
- Joint generation of proteins and their molecular binding partners
- Geometric deep learning and equivariant neural networks
- Protein and molecular representation learning
- Model confidence, uncertainty estimation, and active learning
- Reinforcement learning and experimental-feedback-guided generation
- Efficient training, inference, and model scaling on GPUs
- Benchmarking generative models for physical accuracy and biological function
- Building high-quality datasets and computational pipelines for model development

Students will have opportunities to develop **new machine-learning methods**. Model predictions may be evaluated experimentally by researchers in our interdisciplinary team, allowing computational ideas to be tested in real biological systems.

Who Should Apply?

We welcome:

- Rice undergraduate students seeking research experience or a long-term project
- MS and PhD students interested in generative AI, scientific machine learning, or computational biology
- Students from Computer Science, Data Science, Applied Mathematics, Statistics, or related fields

Useful experience includes Python, PyTorch or JAX, deep learning, generative modeling, geometric learning, high-performance computing, or scientific software development. Prior biology experience is **helpful but not**

required. We are especially interested in students who enjoy developing new algorithms, reading recent ML literature, and working on open-ended research problems.

What You Will Gain

- Experience developing state-of-the-art generative models
- Mentorship at the intersection of AI, protein design, and experimental science
- Opportunities to contribute to publications and open-source research software
- Access to interdisciplinary collaborations across BioSciences, Bioengineering, Chemistry, and Computer Science
- Training in translating fundamental ML advances into real scientific applications

How to Apply

Please email **Prof. Linna An** at la72@rice.edu with:

1. A short description of your research interests
2. Your résumé or CV
3. Relevant coursework, projects, publications, or GitHub repositories
4. Your anticipated availability and desired research timeline

Join us in building generative models that design new molecular functions.

Reference and selective readings

1. An, L. *et al.* Binding and sensing diverse small molecules using shape-complementary pseudocycles. *Science* 385, 276–282 (2024).
2. Tran, L. *et al.* De novo design of orthogonal far-red, orange, and green fluorophore-binding proteins for multiplexed imaging. *Science* eaeb0822 (2026) doi:10.1126/science.aeb0822.
3. Dauparas, J. *et al.* Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* 22, 717–723 (2025).
4. Butcher, J. K. V. *et al.* De novo Design of All-atom Biomolecular Interactions with RFdiffusion3. Preprint at <https://doi.org/10.1101/2025.09.18.676967> (2025).