

Kaustubh Amritkar
Department of Bacteriology, University of Wisconsin-Madison

NASA Astrobiology Early Career Collaboration Award Report

I traveled to the Seoul National University at Seoul, South Korea to visit Dr. Martin Steinegger's lab. The goal of the trip was to gain hands-on training in developing Machine Learning approaches for correlating protein sequence information with experimental measurements, with the longer-term objective of applying these methods to ancestral to better understand their evolution and to guide protein engineering.

During the visit, I learned practical workflows for building and evaluating sequence-based prediction models, including transformer-based protein language models and strategies for fine-tuning them for regression tasks. This training improved my understanding of modern model architectures, data preprocessing, and benchmarking practices needed to integrate large-scale sequence data with experimental readouts.

In addition to machine learning, I gained experience with structural bioinformatics tools central to modern comparative protein analysis, which were developed in Dr. Steinegger's lab like ColabFold, Foldseek, Foldmason etc. I got to learn how these tools work, and a better understanding of their strengths and limitations providing me with valuable perspective on how structure-aware homology detection and large-scale structural comparisons can refine evolutionary analyses, particularly for deeply diverged proteins relevant to early Earth metabolism.

I am grateful to the NASA Astrobiology Program for supporting me in this opportunity. This visit strengthened my technical skill set and broadened my computational toolkit. It helped me develop connections with other graduate students in my field, and work with researchers at the forefront of structural bioinformatics and machine learning. I believe this collaboration visit will shape the trajectory of my research career, and the learnings will remain very important across different projects.