Machine learning to navigate sequence-function landscapes for protein engineering

Philip A. Romero, Ph.D.
Assistant Professor, Department of Biochemistry
University of Wisconsin-Madison

Artificial intelligence and machine learning are revolutionizing protein science and engineering by decoding the complex inner workings of proteins at a scale and resolution beyond human comprehension. Predictive sequence-function models enable protein engineers to search for new and useful proteins with broad applications in medicine, bioenergy, biocatalysis, and biotechnology. In this talk, I will present my group's work leveraging deep neural networks to understand the relationships between protein sequence, structure, and function. We have developed supervised learning methods that infer the sequence-function mapping from large-scale experimental data. These models can be applied to extrapolate beyond the training data to design highly active protein variants. I will also discuss our recent work developing fully autonomous "self-driving labs" that combine AI-based decision making and robotic automation to engineer proteins without human intervention. Data driven protein engineering will become increasingly powerful with continued advances in artificial intelligence, deep learning, and high-throughput and automated experimentation.