Data-driven protein engineering

Philip Romero
Assistant Professor
Department of Chemical and Biomolecular Engineering
University of California, Los Angeles

Abstract: A protein’s amino acid sequence encodes its chemical properties and biological functions. A detailed and quantitative understanding of this encoding would have a profound impact across all areas of biology, medicine, and biotechnology. I am interested in data-driven approaches for systematically exploring the relationships between protein sequence, structure, and function. In this talk, I’ll present how statistical learning algorithms can infer sequence-function relationships from experimental data. These statistical models are exceptionally accurate because they capture a protein’s overall global behavior, and their predictions generalize well to new, previously unseen sequences. I will also describe how recent advances in high-throughput experimentation are enabling us to map protein sequence space on an unprecedented scale. We combine microfluidic screening technologies with next-generation DNA sequencing to measure sequence-function relationships for millions of protein variants. The resulting data serves as a rich resource for dissecting the molecular basis of protein function in a comprehensive and unbiased manner.

Biography: Phil Romero is an assistant professor in the Department of Chemical and Biomolecular Engineering at UCLA. He performed postdoctoral research in Adam Abate’s lab at UCSF where he developed microfluidic technologies for high-throughput biochemistry. As a graduate student at Caltech, he worked in Frances Arnold's laboratory where he engineered proteins for a variety of applications including medical imaging, cancer therapeutics, and biofuel production. His thesis research focused on developing new statistical methods to learn sequence-function relationships from experimental data.