
Probabilistic methods for designing functional protein structures

Brian Trippe

Postdoctoral Fellow, Department of Statistics, Columbia University,
Visiting Researcher, Institute for Protein Design, University of Washington

February 22, 2024

1:30PM-2:50PM

MSOB X303

Abstract:

The biochemical functions of proteins, such as catalyzing a chemical reaction or binding to a virus, are typically conferred by the geometry of only a handful of atoms. This arrangement of atoms, known as a motif, is structurally supported by the rest of the protein, referred to as a scaffold. A central task in protein design is to identify a diverse set of stabilizing scaffolds to support a motif known or theorized to confer function. This long-standing challenge is known as the motif-scaffolding problem.

In this talk, I describe a statistical approach I have developed to address the motif-scaffolding problem. My approach involves (1) estimating a distribution supported on realizable protein structures and (2) sampling scaffolds from this distribution conditioned on a motif. For step (1) I adapt diffusion generative models to fit example protein structures from nature. For step (2) I develop sequential Monte Carlo algorithms to sample from the conditional distributions of these models. I finally describe how, with experimental and computational collaborators, I have generalized and scaled this approach to generate and experimentally validate hundreds of proteins with various functional specifications.

Readings:

1. Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem (link: arxiv.org/abs/2206.04119)

2. De novo design of protein structure and function with RFdiffusion (link: [nature.com/articles/s41586-023-06415-8](https://www.nature.com/articles/s41586-023-06415-8))

3. Practical and asymptotically exact conditional sampling in diffusion models (link: <https://arxiv.org/abs/2306.17775>)