Statistical Computation in Protein Folding

Samuel Kou*

Department of Statistics, Harvard University, kou@stat.harvard.edu

Predicting the native structure of a protein from its amino acid sequence is a long standing problem. A significant bottleneck of computational prediction is the lack of efficient sampling algorithms to explore the configuration space of a protein. In this talk we will introduce a sequential Monte Carlo method to address this challenge: fragment regrowth via energy-guided sequential sampling (FRESS). The FRESS algorithm combines statistical learning (namely, learning from the protein data bank) with sequential sampling to guide the computation, resulting in a fast and effective exploration of the configurations. We will illustrate the FRESS algorithm with both lattice protein model and real proteins.

Key Words: sequential Monte Carlo methods, energy function, 3D structure, statistical learning