Protein conformation sampling using conditional random fields

Protein structure prediction without using templates (i.e., ab initio folding) is one of the most challenging problems in structural biology. In particular, conformation sampling poses as a major bottleneck of ab initio folding. This talk presents CRFSampler, an extensible protein conformation sampler, built upon a probabilistic graphical model Conditional Random Fields (CRFs).
Using a discriminative learning method, CRFSampler can automatically learn thousands of parameters quantifying the relationship among primary sequence, secondary structure and (pseudo)backbone angles. Using only compactness and self-avoiding constraints, CRFSampler can efficiently generate protein-like conformations from primary sequence and predicted secondary structure. CRFSampler is also very flexible in that a variety of model topologies and feature sets can be defined to model the sequence-structure relationship without worrying about parameter estimation. Our experimental results demonstrate that using a simple set of features, CRFSampler can generate decoys with much higher quality than the most recent HMM model and Levitt's lattice model.