Remote Homology Detection: Beyond Hidden Markov Models

Profile Hidden Markov Models remain one of the best general tools to recognize proteins that fold into the same structural motif as a solved protein structure. We look at two different approaches to generalize them to do better at predicting remote protein homologs. The first incorporates evolutionary information in a novel way to construct a better artificial training set. The second is a framework for incorporating pairwise preferences in beta-sheet formation together with HMM models into a Markov Random Field. We show how the second framework can be used to better predict beta-propeller folds, and identify an interesting family of bacterial hybrid two-component sensor system proteins whose N-terminal region we predict contains a double beta-propeller.