BIOINFORMATICS 2011 FALL SEMINAR SERIES
Hosted by: Department of Computer and Information Sciences, Department of Electrical and Computer Engineering & Center for Bioinformatics and Computational Biology
http://bioinformatics.udel.edu/seminars

Improving Interacting Residue Prediction Using Long-Distance Information in Hidden Markov Models

Colin Kern
Ph.D. Student, Department of Computer & Information Sciences University of Delaware
Monday, November 14, 2011 4:00pm DBI – Room 102

ABSTRACT: Identification of interacting residues involved in protein protein and protein-ligand interaction is critical for the prediction and understanding of the interaction and has practical impact on mutagenesis and drug design. Incorporating structural information, a recent work has been developed to identify the interacting domains and residues in protein sequences using the Interacting profile hidden Markov models (ipHMMs). Despite the improved performance, the ipHMMs by design are not capable of capturing the long-distance correlations shown to be existing among interacting residues, sometimes separated by dozens of amino acids in the primary structure.

In this work we introduce a new decoding algorithm with early trace back mechanism in ipHMMs to incorporate the long-distance correlations between interacting residues to improve prediction accuracy. It is shown that an improvement of 3.68% in accuracy, measured by area under curve (AUC), was achieved when applied to a dataset from the 3DID database. To gauge and assess the methods effectiveness in capturing the correlation signals, we devised a scheme to create sets of simulated data based on the 3DID dataset with controllable correlation between interacting residues, and demonstrated that the prediction consistently improves as the correlations increase, achieving an improvement of 12.80% on the simulated data where the nearest neighboring interacting residues are replaced the most likely residues collected from the domain-domain interaction family.