

# UW CENTER FOR PATTERN ANALYSIS AND MACHINE INTELLIGENCE

## GRADUATE SEMINAR SERIES

### Synthesizing Aligned Random Pattern Digraphs from Protein Sequence Patterns

**Speaker:** En-Shiun Annie Lee

**Date:** October 19, 2011

**Time:** 4:30 pm - 5 pm

**Place:** E5 (5128) Refreshments will be served

#### **Abstract :**

An essential step of protein function analysis is discovering patterns that represent functional regions in a set of protein family sequences; however, functional segments may contain biological variations. To capture these associations, we used high order patterns from an effective sequence pattern discovery algorithm as input. Our pattern synthesis system groups and aligns these similar discovered patterns into Aligned Random Pattern Clusters (ARPCs) where each ARPC is transformed into a probabilistic structural pattern, the Aligned Random Pattern Digraph (ARPD). The advantages of our system are a) the ARPCs captures the patterns' identical and variable sites, b) the search space for searching patterns is smaller than the search space for searching sequences, and c) the synthesized ARPD has rich graphical analysis capabilities. Our method successfully discovers two functional protein regions of the Cytochrome Complex protein family: the proximal and distal binding segment that binds the iron molecule of the heme ligand on each side of the plane.

**WATERLOO**  
**ENGINEERING**

**CPAMI**