# UW CENTER FOR
# PATTERN ANALYSIS AND MACHINE INTELLIGENCE

## GRADUATE SEMINAR SERIES

## Identifying Protein Binding Functionality of Protein Family Sequences by Aligned Pattern Clusters

**Speaker:** En-Shiun Annie Lee

**Date:** September 12, 2012

**Time:** 4:00 pm – 4:30 pm

**Place:** E5 (4128) Refreshments will be served

**Abstract :**

A basic task in protein analysis is to discover a set of sequence patterns that reflect the function of a protein family. This set of sequence patterns contains non-exact significant residue associations. Currently, the existing combinatorial methods are computationally expensive and probabilistic methods require richer representation of the amino acid associations. To undertake this task, we create a synthesized pattern representation called an Aligned Pattern (AP) Cluster that identifies the residue associations in the binding segment and the site variations in the aligned residues. In this paper, our algorithm identifies the binding segments for two protein families: the Cytochrome Complex and the Ubiquitin protein families. For each of the experiments, the AP Clusters obtained correspond to protein binding segments including a few beyond those identified by the other protein databases, PROSITE and pFam. Furthermore, the columns of aligned sites that exist only as a single value in the AP Clusters also correspond to the binding residues. Additional information retained by the AP Clusters can reveal the amino acid residues of interest, thus averting time-consuming simulations and experimentation.

**WATERLOO ENGINEERING**

**CPAMI**