

## **Protein Classification Based on Feature Extraction from Hydrophobicity Scales**

**Sumeet Dua, Ph.D.**

Upchurch Associate Professor of Computer Science,  
Louisiana Tech University, Ruston, LA

### **Abstract**

Proteins contain a large but limited number of features. It has long been recognized that the regular, organized structure of a protein embedded in a non-isotropic environment is reflected in the sequence of chemical residue properties in the protein. Protein folding is frequently guided by local residue interactions that form clusters in the protein core. The interactions between residue clusters serve as potential nucleation sites in the folding process. Evidence postulates that the residue interactions are governed by the hydrophobic propensities that the residues possess. The computational challenge to mine and characterize the variability in protein sequence based on hydrophobicity scales, and to identify and exploit the corresponding secondary structural features is a difficult, but an interesting one. We will present a graph-theory-based data mining framework to extract and isolate protein structural features that sustain invariance in evolutionary-related proteins, through the integrated analysis of five well-known hydrophobicity scales over the 3D structure of proteins. We hypothesize that proteins of the same homology contain conserved hydrophobic residues and exhibit analogous residue interaction patterns in the folded state. The results demonstrate that discriminatory residue interaction patterns shared among proteins of the same family can be employed for both supervised structural classification and the functional annotation of proteins. Extensive experimentation demonstrates enhanced results with boosted specificity and sensitivity of structural classification compared to the previous results in the area.