

# Statistical pattern recognition methods for protein secondary structure

Peter Hooper

hooper@stat.ualberta.ca

University of Alberta

A protein is a polymer made of amino acids. The linear sequence of amino acids is called the primary structure of the protein. A fundamental problem in biology is to predict the three-dimensional structure of a protein given its primary structure. As a first step, researchers have investigated the prediction of secondary structure; i.e., the sequence of recurring three-dimensional patterns usually characterized as alpha helix, beta strand, and coil. I will briefly survey work in this area and describe my own efforts. My approach is based on a recently developed extension of logistic discriminant analysis called reference point logistic classification.