

# Prediction of Protein Secondary Structures

Kuen-Pin Wu, Shin-Nan Lin, Jia-Ming Chang, Ting-Yi Sung,  
and Wen-Lian Hsu

Institute of Information Science, Academia Sinica

## Abstract

We develop a knowledge-based approach for protein secondary structure prediction. The knowledge base contains small peptide fragments along with their structure information. We define a quantitative measure  $M$ , called match rate, which represents the amount of structure information that a target protein can extract from the knowledge base. Our experimental results show that proteins with a higher match rate, namely those extracting more information from the knowledge base, can be predicted more accurately based on our knowledge-based method. To fully utilize the strength of our knowledge base, we propose a hybrid prediction method as follows: predetermine a cutoff threshold value  $M^*$  for match rate; if the match rate of a target protein is greater than  $M^*$ , we use the extracted information to make the prediction; otherwise, we adopt a popular machine learning approach, such as PHD or PSIPRED. This comprises our hybrid protein structure prediction (PROSP) approach.

We use the DSSP database as our dataset and PSIPRED as our underlying machine learning algorithm. If  $M^*$  is set at 80%, the average prediction accuracy of PROSP is 79.3%, which is 1.03% better than that of PSIPRED on the same dataset.