Computational Proteomics and 'Encyclopedia of Life' project

Cheng-Yan Kao

Computer Science and Information Engineering, National Taiwan University

Abstract

This talk will address our efforts in computational proteomics and the future application in the encyclopedia of life project (EOL). Large-scale genomics and proteomics technologies are crucial for reconstructing molecular circuitry of a living cell. Recent techniques for construction and prediction of large-scale protein interaction networks are surveyed, focusing on computational processing steps in the case of experimental techniques. The computational proteomics plays a crucial role in assigning function to sequenced proteins, defining pathways in which the targets are involved, and understanding structure-function relationships of the protein targets.

In the problems of protein structure prediction, we used Support Vector Machine to effectively predict the protein disulfide connectivity pattern directly from its amino acid sequence at first. The correct prediction of disulfide bridges can reduce the conformational search space and improve the results while predicting protein structures. Then, we conducted family competition evolutionary algorithms to *ab initio* structure prediction with lattice models. After determining the structure of the backbone of a protein, we developed an evolutionary approach to predict protein side-chain conformations. Although we used evolutionary approaches to lighten the enormous load of the complicated computation, it will be much faster by using hardware to speed up. Therefore, we are implementing the Amber Force-field function chip to solve the calculation of scoring function, which costs most time in structure prediction. Moreover, we also explored the problems of protein docking, developing an evolutionary approach and integrating Quantitative Structure Activity Relation for flexible ligand docking.

In the future, we proceed to apply our current technologies into the Encyclopedia of Life (EOL) project. The Encyclopedia of Life is a joint development of the San Diego Supercomputer Center (SDSC) and scientists and biological resources worldwide. EOL involves the largest computation ever attempted to predict putative functional and 3-D structure assignment of a protein. In addition, it aims to integrate the key biological resources in the world and promote future collaborative developments. Our lab will focus on joining the grid computing of EOL and applying our *ab initio* method in 3-D structure prediction. Furthermore, Prof. Wen-Hsiung Li will provide the valuable using experience from the aspect of biologists to help us improve EOL.