Public Abstract
First Name:zhiquan
Middle Name:
Last Name:he
Adviser's First Name:Dong
Adviser's Last Name:Xu
Co-Adviser's First Name:
Co-Adviser's Last Name:
Graduation Term:SP 2014
Department:Computer Science
Degree:MA
Title:METHODS FOR PROTEIN STRUCTURE PREDICTION

With large amount of protein sequences generated by genome-sequencing projects, the lack of tertiary structures is a main obstacle to fully understanding the functions of these proteins. Traditionally, experimental determination of protein structures has utilized both X-ray crystallography and nuclear magnetic resonance (NMR), which are time consuming and costly. Computational structure prediction from amino acid sequence is a viable solution. Recent reviews showed that predicted models of different qualities can be used in various applications from drug design to helping predict protein functions. Although several decades of efforts have been made to push protein structure prediction forward, it is still challenging nowadays. The major reason for this is the difficulty to capture the fundamental relationship between protein sequences and structures, especially when the sequence similarities among proteins are relatively low.
The widely used method for protein structure prediction is comparative protein modeling, which heavily relies on fold recognition performance and alignment accuracy. Another step in protein structure prediction is the structural assessment for predicted protein structures, which obviously plays a critical role.
In this thesis, we discussed several methods for protein structure prediction to address the two important issues. The corresponding tools have been applied in our in-house protein structure prediction platform (MUFOLD).
More specifically, we implemented a protein sequence alignment tool which is based on Conditional Random Field and improved its alignment quality by incorporating more complex scoring models. After deeper study of fold recognition and alignment problem, we proposed a new protocol to improve the quality of sequence profiles, which intrinsically affects the performance of fold recognition and alignment accuracy. Besides this, several machine learning methods have been proposed to combine knowledge scoring functions and consensus methods from different perspectives for structural quality assessment purpose. For example, graphical probability models such as Hidden Markov Model and Conditional Random Field have been used to combine sequence and structural features to predict the structural quality of predicted protein models.
These tools have demonstrated good performance in discriminating protein models of different qualities.