Public Abstract

First Name:Xin

Middle Name:

Last Name:Deng

Adviser's First Name:Jianlin

Adviser's Last Name:Cheng

Co-Adviser's First Name:

Co-Adviser's Last Name:

Graduation Term:FS 2013

Department:Computer Science

Degree:PhD

Title:Improved Computational Methods Of Protein Sequence Alignment, Model Selection And Tertiary Structure Prediction

Protein sequence and profile alignment is a basic tool for bioinformatics research and analysis. It has been used essentially in almost all bioinformatics tasks such as protein structure modeling, gene and protein function prediction, DNA motif recognition, and phylogenetic analysis.

We designed and developed a new method, MSACompro, to synergistically incorporate some predicted structural information to improve the accuracy of multiple sequence alignments.The evaluation results demonstrated that our method improves the multiple sequence alignment accuracy over the leading multiple protein sequence alignment tools without using the predicted structural information.And the performance of the method is comparable to the state-of-the-art method PROMALS of using structural features and additional homologous sequences by slightly lower scores. We also developed a novel profileprofile pairwise protein sequence alignment method based on pair HMM (Hidden Markov Model) by integrating some novel information. The evaluation showed that our method significantly improved the alignment accuracy in comparison with the state of the art methods HHsearch and HHsuite.

Protein Model selection is also a key step in protein tertiary structure prediction. We developed two SVM model quality assessment methods, taking either a query-single template pairwise alignment or a query-multi template alignment as input. The assessment results illustrated that such a novel, effective method may help improve the model selection, protein structure prediction and many other bioinformatics problems.

Protein tertiary structure prediction has been playing a profound role in drug design, pharmacy and medical field. Based on the above methods, we developed a protein tertiary structure pipeline, and many components were used in our group's MULTICOM system of protein tertiary structure prediction. The MULTICOM system performed well in the CASP10 (Critical Assessment of Techniques for Protein Structure Prediction) competition.