Public Abstract

First Name:Yan

Middle Name:

Last Name:Chen

Adviser's First Name:Yi

Adviser's Last Name:Shang

Co-Adviser's First Name:

Co-Adviser's Last Name:

Graduation Term:FS 2013

Department:Computer Science

Degree:MS

Title:Multi-Dimensional Scaling and MODELLER based Evolutionary Algorithms for Protein Model Refinement

To computationally obtain an accurate prediction of the three-dimensional structure of a protein from its primary sequence is one of the most important problems in bioinformatics and has been actively researched for many years. Although a number of software packages have been developed and they sometimes perform well on template-based modeling, further improvement is needed for practical use. Model refinement is a step in the prediction process, in which improved structures are constructed based on a pool of initially generated models. Since the refinement category being added to the Critical Assessment of Structure Prediction (CASP) competition in 2008, CASP results show that it is a challenge for existing model refinement methods to improve model quality consistently.

This project focuses on evolutionary algorithms for protein model refinement. Three new algorithms have been developed, in which multidimensional scaling (MDS), MODELLER, and a hybrid of both are used as crossover operators, respectively. The MDS-based method takes a purely geometrical approach and generates a child model by combining the contact maps of multiple parents. The MODELLER-based method takes a statistical and energy minimization approach and uses the remodeling module in MODELLER program to generate new models from multiple parents. The hybrid method first generates models using the MDS-based method and then run them through the MODELLER-based method, aiming at combining the strength of both. Promising results have been obtained in experiments using CASP datasets. The MDS-based method improved the best of a pool of predicted models in terms of the global distance test score (GDT-TS) in 9 out of 16 test targets. For instance, for target T0680, the GDT-TS of a refined model is 0.833, much better than 0.763, the value of the best model in the initial pool.