

Designing and Benchmarking the MULTICOM Protein Structure Prediction System

Jianlin Cheng^{1,2,3*}, Zheng Wang¹, Jesse Eickholt¹

¹Computer Science Department, ²Informatics Institute, ³C. Bond Life Science Center, University of Missouri, Columbia, MO 65211, USA

*corresponding author (chengji@missouri.edu)

Abstract

Background:

Predicting protein structure from sequence is one of the most significant and challenging problems in bioinformatics. Numerous techniques and methods have been developed to tackle every aspect of protein structure prediction ranging from structural feature prediction, template identification, query-template alignment, conformation sampling, model quality assessment, and model refinement. Therefore, how to synergistically integrate and improve these complementary techniques to build a high-performance system is becoming a critical issue for developing a successful, competitive protein structure predictor.

Results:

Here we designed and implemented an accurate, standalone protein structure prediction system MULTICOM that combines multiple sources of information and complementary methods. The components of the system were integrated through a five-layer architecture consisting of template identification, template combination, model generation, model assessment, and model refinement. The system was blindly tested during the ninth Critical Assessment of Techniques for Protein Structure Prediction (CASP9) in 2010 and yielded an outstanding performance. Here we thoroughly investigated both the overall performance of the system and the contributions of its components in each layer on the CASP9 benchmark. Our comprehensive and comparative study provides the valuable and practical insights about how to integrate protein structure prediction components altogether to develop a cutting-edge protein structure prediction system. Our experiment demonstrated that a high-performance protein structure prediction system running on a PC can be built by integrating existing techniques under an open component-based architecture.

Availability:

The MULTICOM web service is available at: <http://caspr.net.missouri.edu/multicom/>. Some software components used in the MULTICOM system can be downloaded at: <http://caspr.net.missouri.edu/download>.