Construction of Fully Automated Protein Structure Modeling System

Joo-Hyun Seo and Byung-Gee Kim

School of Chemical and Biological Engineering, Seoul National University

In the post-genomic era, although the number of the verified structure of protein is increasing rapidly, there are still many proteins whose structure is not revealed yet. But protein structure is in the high demand for the study of protein, especially for the investigation of substrate specificity, stability, function and so on. In thisbackground, protein modeling is highly needed because experimental structure verification needs a lot of time and effort. Here, we introduce the fully automated protein modeling system. The system was built using THREADER 3.5 and MODELLER 6.2 and Python 2.4.1 programming language on the Linux platform. If user input the sequence in FASTA format, sequence investigation module and sequence parsing module, which are encoded with Python, examines the submitted sequences. Then submitted sequences are modeled one by one. Querysequence is submitted to THREADER for the fold recognition. The system uses the best PDB structure from the analysis of fold recognition result as template. Then query sequence and template structure is submitted to MODELLER for the homology modeling. We tested the accuracy of the model from our system with the 10 PDB structure. Protein sequence is modeled and modeling result is compared to its experimental PDB structure. Our system gives $0.5\text{Å} \sim 1.4\text{Å}$ RMSD for the test set. We compared our result to those from EsyPred3D (Bioinformatics, 2002,18(9),1250-1256) EsyPred3D also showed $0.5\text{Å} \sim 1.3\text{Å}$ RMSD. In some case our system gives better RMSD and in some case EsyPred3D gives better result. The system is available by web. Anyone can access the system by following the guide: "http://plaza.snu.ac.kr/~byungkim →Virtual Institute → Bioinformatics"