C-5

Genome Scale Protein Secondary Structure Prediction Using a Data Distribution on a Grid Computing

Min-Kyu Cho[1], Soojin Lee[2], Jin-Won Jung[1], Jai-Hoon Kim[2], and Weontae Lee[1]

[1] HTSD-NMR Laboratory, Department of Biochemistry, Yonsei University, Seoul, 120-749, Korea [2]Distributed and Mobile Computing Laboratory, Graduate School of Information and Communication, Ajou University, Suwon, 442-749, Korea

After many genome projects, algorithms and software to process explosively growing biological information have been developed. To process huge amount of biological information, high performance computing equipments are essential. If we use the remote resources such as computing power, storages etc., through a Grid to share the resources in the Internet environment, we will be able to obtain great efficiency to process data at a low cost. Here we present the performance improvement of the protein secondary structure prediction (PSIPred) by using the Grid platform, distributing protein sequence data on the Grid where each computer node analyzes its own part of protein sequence data to speed up the structure prediction. On the Grid, genome scale secondary structure prediction for Mycoplasma genitalium, Escherichia coli, Helicobacter pylori, Saccharomyces cerevisiae and Caenorhabditis elegans were performed and analyzed by a statistical way to show the protein structural deviation and comparison between the genomes. Experimental results show that the Grid is a viable platform to speed up the protein structure prediction and from the predicted structures.