GPS@, Web interface for Protein Sequence Analysis on Grid

Tuesday, 26 September 2006 14:25 (10 minutes)

Bioinformatics analysis of data produced by high-throughput biology, for instance genome projects [1], is one of the major challenges for the next years. Some requirements of such analysis are to access up-to-date databanks (of sequences, patterns, 3D structures, etc.) and relevant algorithms (sequence similarity, multiple alignment, pattern scanning, etc.) [2]. Since 1998, we are developing the NPS@ Web server ([3], Network Protein Sequence Analysis), that provides the biologist with many of the most common resources for protein sequence analysis, integrated into a common workflow. These methods and data can be accessed through simple web browsing and HTTP connection, or through high-level bioinformatics interface like MPSA or AntheProt programs.

GPS@ Web portal (Grid Protein Sequence Analysis, http://gpsa-pbil.ibcp.fr) is the grid-enabled release of the NPS@ bioinformatics portal. GPS@ hides mechanisms required for submitting bioinformatics analyses on the grid infrastructure. By simply selecting the "EGEE"check-box, GPS@ will schedule the submission of the BLAST computation on the EGEE grid when clicking on the "submit"button. The bioinformatics algorithms and databases available on GPS@ have been distributed and registered on the grid and GPS@ runs its own EGEE interface to the grid.

GPS@ portal makes the Bioinformatics job submission easier on the grid, and provide biologists with the benefit of the EGEE grid infrastructure to analyze large biological dataset: for example including several protein secondary structure predictions into a multiple alignment, or clustering a sequence set by analyzing, with BLAST or SSEARCH, each sequence against the others.

[1] Bernal, A., Ear, U., Kyrpides, N. : Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. NAR 29 (2001) 126-127
[2] G. Perrière, C. Combet, S. Penel, C. Blanchet, J. Thioulouse, C. Geourjon, J. Grassot, C. Charavay, M. Gouy, L. Duret and G. Deléage, Integrated databanks access and sequence/structure analysis services at the PBIL. Nucleic Acids Res., 31:3393-3399, 2003.
[3] Combet, C., Blanchet, C., Geourjon, C. et Deléage, G. : NPS@: Network Protein Sequence Analysis. Tibs, 25 (2000) 147-150.

Summary

Bioinformatics, Grid computing, Tool integration, Web portal

Author: Dr BLANCHET, Christophe (CNRS IBCP)

Co-authors: Dr COMBET, Christophe (Institut de Biologie et Chimie des Protéines (IBCP UMR 5086); CNRS; Univ. Lyon 1;); Prof. DELEAGE, Gilbert (CNRS IBCP); Mr MOLLON, Rémi (Institut de Biologie et Chimie des Protéines (IBCP UMR 5086); CNRS; Univ. Lyon 1;)

Presenter: Dr BLANCHET, Christophe (CNRS IBCP)

Session Classification: Life Sciences (NA4)