

## **Web Services running Protein Sequence Alignment on Grid: Application to Hepatitis C Viruses.**

*Friday 11 May 2007 14:40 (20 minutes)*

**Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).**

Bioinformatics analysis of data produced by high-throughput biology, for instance genome projects, is one of the major challenges for the next years. Some of the requirements of this analysis are to access up-to-date databanks and relevant algorithms. Since 1998, we are developing the Network Protein Sequence Analysis (NPS@) Web server, that provides the biologist with some most common resources for protein sequence analysis, integrated into several pre-defined and connected workflows.

**Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.**

We have put online two multiple alignments Web Services on the CNRS IBCP servers. One is accessible through a classical Web interface, the other one can be used through a SOAP client such as Taverna or Triana, but also a user one built with gSOAP, perl SOAP::Lite or Java. These Web services can process the submitted alignment on two different computing environments: a local cluster, and the grid platform of the EU-EGEE project. The GPS@ sub-process that have submitted the job, is monitoring the job with the EGEE commands. When achieved, the GPS@ automat downloads the result file containing the multiple alignment computed by ClustalW, and processed it in a HTML page showing, in a colored and graphical way, the list of aligned protein sequences. This is directly inherited from the original NPS@ portal, providing biologists with a well-known interface and way of displaying results.

**With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possible, point out the experience limitations (both in terms of existing services or missing functionality)**

Future works will be done about applying this Grid WebServices interface to other programs. We will, for example, work to put online, as Web Services, a selected panel of other protein alignment methods, but also similarity searching programs, like BLAST or SSEARCH, raising the issues of large and numerous databases management in Grid environment.

**Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications**

GPS@ Web portal and gBIO-WS make the remote access and bioinformatics job submission easier on the grid. We have used, as testcase, the ClustalW multiple alignment tool run on a remote Grid platform, to analyze the variability of a subset of sequences.

Biologists can submit bioinformatics jobs on the Grid by using their usual Web client, but also integrate these grid services within complex workflow combining different databases and tools. They, then, benefit from the large-scale computing resources of the Grid, from their usual and local working environment. Grid computing and storage facilities will also permit GPS@ and gBIO services to scale to thousands of daily user as much as aligning complete genomes or proteomes. In this testcase, we consider a common task for bioinformaticians working on Hepatitis C Virus: doing a multiple alignment of sequences issued from different strains, where user is uploading its own sequence databank of HCV sequences.

**Author:** Dr BLANCHET, Christophe (CNRS IBCP)

**Co-authors:** Dr COMBET, Christophe (CNRS IBCP); Prof. DELEAGE, Gilbert (CNRS IBCP); Mr DARIC, Vladimir (CNRS IBCP)

**Presenter:** Dr BLANCHET, Christophe (CNRS IBCP)

**Session Classification:** Interactivity and Portals

**Track Classification:** Interactivity and Portals