Contribution ID: **33**   Type: **oral presentation**

"High throughput" protein structure prediction application in EUChinaGRID

Thursday 10 May 2007 15:00 (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).

In nature there exists only a tiny fraction of all the theoretically possible protein sequences. It is thus of interest for the biologists to study the properties of proteins not present in nature as a way to improve our knowledge on the unique properties of the existing protein sequences. The demo will deal with the use of the protein structure prediction software Rosetta in a grid enviornment as a tool to study the structural properties of large protein sequence libraries.

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.

The demo will show how the application and related services for job submission and retrieval of the output have been integrated within the GENIUS portal. In addition it will be also demonstrated how a non grid trained user can submit jobs, retrieve the output and obtain a graphical representation of the output also through the use of molecular graphics plugins. This is a critical point when dealing with the biology and biomedical community as proper informatics training is still scarce in these disciplines, while the use of informatics tools in general and of the grid infrastructure in particular is of unvaluable help in the research in these fields.

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)

The demo presentation will also deal with aspects that still need development/improvement. In particular, a simplified, web-based access to the grid infrastructure allows easy and fast submission of jobs and further developments along this path are critical for the access of novel communities to the grid resources.

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

The demo will show how, through the use of grid resources, the problem of studying the structural properties of a huge number of protein sequences can be tackled. Through the use of a parametric job submission procedure, the user can submit multiple structure prediction jobs and perform a large number of

predictions in a reasonable time frame. This is essential when dealing with a library made up of 10ˆ5 10ˆ7 protein sequences. In fact, the size of the library is essential to derive general rules that allow us to pinpoint the particular properties of natural proteins that make them unique.

Primary author: Prof. POLTICELLI, Fabio (Department of Biology, University Roma Tre)

Co-authors: Dr MINERVINI, Giovanni (Department of Biology, University Roma Tre, Italy); Dr LA ROCCA, Giuseppe (INFN Catania, Italy); Prof. LUISI, Pier Luigi (Department of Biology, University Roma Tre, Italy)

Presenter: Dr MINERVINI, Giovanni (Department of Biology, University Roma Tre, Italy)

Session Classification: Experience with application domains