Contribution ID: **32** Type: **poster**

Protein structure prediction of ”never born proteins”. An experience within the EUChinaGRID framework

Wednesday 9 May 2007 17:30 (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 (the “never born proteins”) as a way to improve our knowledge on the fundamental properties that make existing protein sequences so unique. Protein structure prediction tools combined with the use of large computing resources allow to tackle this problem.

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 biological community is increasingly taking advantage of the informatics tools available. However informatics training is still scarce in biology courses. In our experience in porting protein structure prediction applications in a grid environment, the main difficulty we encountered was in being able to formalize job description and submission as well as output retrieval through the use of grid middleware. Essential to our experience was the integration of our applications within the GENIUS portal which allows job submission and management through the use of a user friendly grid interface. This allows non grid-trained users to profit from the advantages provided by the grid infrastructure.

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)

According to our experience, a simplified approach to the use of

grid resources
will be the key point to attract new communities, especially as
far as the
biomedical community is concerned. From this viewpoint, the further
development of web-based, user friendly services will be critical
to allow the
access of non informatics trained scientists to grid resources
and for the
success of grid computing approach.

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 study of never born proteins requires the generation of a
large library of
protein sequences not present in nature and the prediction of
their three-
dimensional structure. This is not trivial when facing 10^5-10^7
protein
sequences. Indeed, on a single CPU it would require years to
predict the
structure of such a large library of protein sequences. On the
other hand, this is
a trivial parallelism problem in which the same computation (i.e.
the prediction
of the 3D structure of a protein sequence) must be repeated
several times (i.e.
on a large number of protein sequences). The use of the GRID
infrastructure
makes feasible to approach this problem in an acceptable time frame.
In addition, once the simulation in a grid environment has been
set up, the
same approach can be used to tackle problems of immediate biomedical
relevance such as the prediction of the structure of the entire
set of proteins of
a virus or a bacterial pathogen.

Author: Prof. POLTICELLI, Fabio (Department of Biology, University Roma Tre, Italy)

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: Poster and Demo Session

Track Classification: Poster session