5th EGEE User Forum

Contribution ID: **168** Type: **Oral**

A Protein Tertiary Structure Prediction Service in the ProGenGrid System

Tuesday, 13 April 2010 11:45 (15 minutes)

Proteins are central to all biological processes: a very important problem in proteomics is the prediction of the three-dimensional (3D) structure of proteins from their amino acid sequence, because this information can be useful for determining the protein function, given by a specific spatial conformation that the protein assumes when it reaches the active state. We have integrated a routinely expert dependent strategy, based on the homology modeling procedure, in an automatic tool that may facilitate the generation of carrier models at low resolution, exploiting the EGEE infrastructure.

Detailed analysis

We have defined an "in silico"workflow for the tertiary structure prediction of carrier proteins based on the homology modeling procedure, starting from the amino acid sequence. As case study, we modelled the structure of the DIcarboxylate Carrier (DIC) of Yeast, using the ADP/ATP carrier protein of Bos Taurus heart mitochondria. The protein tertiary structure prediction process involves several data and software tools in the following phases: a) template/s search with the PSI-BLAST tool; b) secondary structure computing: the topological model is calculated by using the ENSEMBLE predictor; c) multi alignment between the target sequence and the template with ClustalW and Yap tools; d) models building with Modeller; and e) models check with CE and Procheck tools. These tools are all sequential and in particular Modeller is computationally expensive if a great number of models is requested as output. We run the alignment tools on Globus machines of the SPACI grid infrastructure and other tools on gLite, by using our tool of submission and monitoring of jobs. It is a meta scheduler that has several plug-ins for scheduling and monitoring of jobs on different grids.

Conclusions and Future Work

This paper presented a service for tertiary structure prediction implemented in the ProGenGrid system. The validation of the system has been performed by predicting the structure of the DIC protein. ProGenGrid has been used both for designing the workflow prediction graph and for scheduling it onto a Computational Grid, based on gLite and Globus grid resources. Moreover a virtual reality environment, based on the X3D and Ajax3d technologies, has been built in order for interacting with the protein. We plan to extend the service for the prediction to other kind of proteins.

Impact

We have integrated a set of software and data involved in the tertiary structure prediction process of several carrier proteins, producing good models with RMSD equals to 1.2-1.3 A. Several tools have been ported on the EGEE infrastructure and integrated in the ProGenGrid system, a Grid PSE. A main contribution of this work is thus a Grid-based tertiary structure prediction service based on the previous cited grid-enabled bioinformatics tools. The developed service offers the following features: 1) prediction of the protein three-dimensional structures; 2) optimization and 3) visualization of the structures, taking into account several formats such as

PDB and X3D.
With respect to other tools such as Swiss-Model and Geno3d, that are suited for general purpose prediction of the models, our system has been developed and optimized for transmembrane proteins, using specific tools such as ENSEMBLE and YAP for generating the secondary structure prediction.
Tests were made by using the ProGenGrid Grid Portal, that allows the launching and monitoring of the simulation in a Grid environment based on gLite and Globus middleware, by using Biomed VO and SPACI facilities for testing the service.

Keywords

Tertiary Structure Prediction, Grid Problem Solving Environment, Grid Portal, Workflow

URL for further information

https://sara.unisalento.it/cgi-bin/bioinfo/enter

Author: Dr MIRTO, Maria (CMCC, University of Salento, Lecce)

Co-authors: Dr FERRAMOSCA, Alessandra (Di.S.Te.B.A, University of Salento, Lecce); Prof. ALOISIO, Giovanni (CMCC, University of Salento, Lecce); Prof. ZARA, Vincenzo (Di.S.Te.B.A, University of Salento, Lecce)

Presenter: Dr MIRTO, Maria (CMCC, University of Salento, Lecce)

Session Classification: Bioinformatics

Track Classification: End-user environments, scientific gateways and portal technologies