Contribution ID: **101**   Type: **demo presentation**

Multiple Ligand Trajectory Docking - A Case Study of Complex Grid Jobs Management

Wednesday, 9 May 2007 19:30 (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).

Interaction between large biomolecules and smaller bio-active ligands lies on the
foundation of many biological properties and is of huge interest in the bio-molecular
and pharmaceutical research. The role of the ligand is to influence the reaction that
occurs in an active site of a biomolecule. The interaction is usually studied using
multiple ligand trajectory docking, a computationally intensive process aimed to find
energetically favorable orientation of the ligand within an active site.

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 present a specific thin graphical application built on top of the gLite Job
Provenance service (JP) and Charon Extension Layer (CEL). JP keeps long-term track of
execution of Grid jobs, including annotations, and provides data mining capabilities
on these data. CEL toolkit is a universal framework creating a layer upon the basic
Grid middleware making the process of job preparation and manipulation a relatively
easy task. CEL provides a command-line interface and offers full control over
submitted jobs.

The application supports selection of subsets of both trajectory snapshots and
specific ligands, queries JP, and displays a 2D array of finished Grid jobs matching
the criteria. Their results can be examined, including 3D visualization of the
emerging structures. Bunches of jobs can be prepared in order to fill empty cells of
the matrix to complete the overall docking analysis. Also, the already finished jobs
can be taken from JP, cloned and re-run with modified input parameters.

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)

Job Provenance can handle arbitrary annotations assigned to Grid
jobs. Such
annotations can be understood to belong to the job outputs too.
The demonstration is
based on this approach. However, native data item annotations are
not supported by JP
by design. Presence of general data annotation/provenance service
would streamline
the solution of the presented problem.

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 search is done on snapshots taken from the molecular dynamics
trajectory
describing the dynamic behavior of the biomolecule. For each
snapshot and ligand the
best position of the ligand is calculated, yielding a matrix
containing energies of
snapshot/ligand interactions. The minima correspond to the most
favorable ligands and
provide insight into the specific shape of the biomolecule.

The concrete problem shown in the demo deals with one 2 ns
acetylcholinesterase
trajectory and 3 ligands, requiring some 6000 CPU hours on an
average compute
server. A realistic studies use more and longer trajectories
(tens ns) and higher
number of potential ligands (tens to hundreds).

Performing such computation is infeasible without the Grid
infrastructure. Also,
managing its results is a non-trivial problem, undoable without
semi-automatic
support tools. A sophisticated job submission system coupled with
a community-wide
provenance of already run jobs is a necessary prerequisite.

Primary authors: KRENEK, Ales (CESNET); KMUNICEK, Jan (CESNET); WIESNER, Jiri (CESNET); PETREK, Martin (CESNET)

Co-authors: FILIPOVIC, Jiri (CESNET); SITERA, Jiri (CESNET); MATYSKA, Ludek (CESNET); KULHANEK, Petr (CESNET); SUSTR, Zdenek (CESNET)

Presenters: KRENEK, Ales (CESNET); KMUNICEK, Jan (CESNET); PETREK, Martin (CESNET)

Session Classification: Poster and Demo Session