



In silico docking on EGEE infrastructure: the case of WISDOM

Wednesday 1 March 2006 16:45 (15 minutes)

Advance in combinatorial chemistry has paved the way for synthesizing large numbers of diverse chemical compounds. Thus there are millions of chemical compounds available in the laboratories, but it is nearly impossible and very expensive to screen such a high number of compounds in the experimental laboratories by high throughput screening (HTS). Besides the high costs, the hit rate in HTS is quite low, about 10 to 100 per 100,000 compounds when screened on targets such as enzymes. An alternative is high throughput virtual screening by molecular docking, a technique which can screen millions of compounds rapidly, reliably and cost effectively. Screening millions of chemical compounds in silico is a complex process. Screening each compound, depending on structural complexity, can take from a few minutes to hours on a standard PC, which means screening all compounds in a single database can take years. Computation time can be reduced very significantly with a large grid gathering thousands of computers.

WISDOM (World-wide In Silico Docking On Malaria) is an European initiative to enable the in silico drug discovery pipeline on a grid infrastructure. Initiated and implemented by Fraunhofer Institute for Algorithms and Scientific Computing (SCAI) in Germany and the Corpuscular Physics Laboratory (CNRS/IN2P3) of Clermont-Ferrand in France, WISDOM has deployed a large scale docking experiment on the EGEE infrastructure. Three goals motivated this first experiment. The biological goal was to propose new inhibitors for a family of proteins produced by Plasmodium falciparum. The biomedical informatics goal was the deployment of in silico virtual docking on a grid infrastructure. The grid goal is the deployment of a CPU consuming application generating large data flows to test the grid operation and services. Relevant information can be found on <http://wisdom.eu-egee.fr> and <http://public.eu-egee.org/files/battles-malaria-grid-wisdom.pdf>.

With the help of the grid, large scale in silico experimentation is possible. Large resources are needed in order to test in a transparent way a family of targets, a large enough amount of possible drug candidates and different virtual screening tools with different parameter / scoring settings. The grid added value lies not only in the computing resources made available, but also already in the permanent storage of the data with a transparent and secure access. Reliable Workload Manager System, Information Service and Data Management Services are absolutely necessary for a large scale process. Accounting, security and license management services are also essential to impact the pharmaceutical community. In a close future, we expect improved data management middleware services to allow automatic update of compound database and the design of a grid knowledge space where biologists can analyze output data.

Finally key issues to promote the grid in the pharmaceutical community include cost and time reduction in a drug discovery development, security and data protection, fault tolerant and robust services and infrastructure, and transparent and easy use of the interfaces.

The first biomedical data challenge ran on the EGEE grid production service from 11

July 2005 until 19 August 2005. The challenge saw over 46 million docked ligands, the equivalent of 80 years on a single PC, in about 6 weeks. Usually in silico docking is carried out on classical computer clusters resulting in around 100,000 docked ligands. This type of scientific challenge would not be possible without the grid infrastructure - 1700 computers were simultaneously used in 15 countries around the world. The WISDOM data challenge demonstrated how grid computing can help drug discovery research by speeding up the whole process and reduce the cost to develop new drugs to treat diseases such as malaria. The sheer amount of data generated indicates the potential benefits of grid computing for drug discovery and indeed, other life science applications. Commercial software with a server license was successfully deployed on more than 1000 machines in the same time. First docking results show that 10% of the compounds of the database studied may be hits. Top scoring compounds possess basic chemical groups like thiourea, guanidino, amino-acrolein core structure. Identified compounds are non peptidic and low molecular weight compounds. Future plans for the WISDOM initiative is first to process the hits again with molecular dynamics simulations. A WISDOM demonstration will be conceived at the aim to show the submission of docking jobs on the grid at a large scale. A second data challenge planned for the fall of 2006 is also under preparation to improve the quality of service and the quality of usage of the data challenge process on gLite.

Primary author: Mr JACQ, Nicolas (CNRS/IN2P3)

Co-authors: Dr MAASS, Astrid (Fraunhofer SCAI); Mrs JACQ, Florence (CNRS/IN2P3); Dr SCHWICHTENBERG, Horst (Fraunhofer SCAI); Mr SALZEMANN, Jean (CNRS/IN2P3); Mr VINOD-KUSAM, Kasam (Fraunhofer SCAI); Mr SRIDHAR, Mahendrakar (Fraunhofer SCAI); Dr ZIMMERMANN, Marc (Fraunhofer SCAI); Dr HOFMANN, Martin (Fraunhofer SCAI); Mr REICHSTADT, Matthieu (CNRS/IN2P3); Dr BRETON, Vincent (CNRS/IN2P3); Mr LEGRÉ, Yannick (CNRS/IN2P3)

Presenter: Mr JACQ, Nicolas (CNRS/IN2P3)

Session Classification: 1a: Life Sciences

Track Classification: Life Science