Contribution ID: **107** Type: **Demo**

Grid-enabled high throughput virtual screening against neglected and emerging diseases

Tuesday 26 September 2006 17:00 (20 minutes)

Malaria is a dreadful disease affecting 300 million people and killing 1.5 million people every year. Drug resistance has emerged for all classes of antimalarials except artemisinins. This example illustrates the real need for new drugs against neglected diseases. There are millions of chemical compounds available, but it is nearly impossible and very expensive to screen such a high number of compounds in the experimental laboratories by high throughput screening. Besides the high costs, the hit rate is quite low [1]. An alternative is high throughput virtual screening by molecular docking, a technique which can screen millions of compounds rapidly, reliably and cost effectively. 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 [2].

In 2005, for the first time, we have been able to deploy large scale virtual docking within the framework of the WISDOM initiative [3] against plasmepsin, the aspartic protease of Plasmodium, responsible for the initial cleavage of human haemoglobin [4]: more than 46 million ligands were docked in less than 6 weeks using about 80 years of CPU on the EGEE [5] infrastructure. Up to 1700 computers were simultaneously used in 15 countries around the world. Commercial software with a server license was successfully deployed on more than 1000 machines at the same time. At the end of the large scale docking deployment, 100 compounds have been

selected for post processing based on the docking score, the binding mode of the compound inside the binding pocket and the interactions of the compounds to key residues of the protein [6]. Some of the compounds identified were similar to already known plasmepsin inhibitors, like the Urea analogues which were already established as micro molar inhibitors for plasmepsins. This indicates that the overall approach is sensible and large scale docking on computational grids has real potential to identify new inhibitors. In addition to this the Guanidino analogues are very promising and most likely to become a novel class of plasmepsin inhibitors.
This success led to a second computing challenge targeting Avian Flu neuraminidase N1 that required more than 100 CPU years on the EGEE, Auvergrid and TWGrid infrastructures in April and May 2006 [7]. Potential drug compounds against avian flu are now being identified and ranked according to the binding energies of the docked models. At least 50 compounds will be assayed experimentally at identified laboratories.

The WISDOM production environment was designed to achieve production of a large amount of data in a limited time using EGEE, Auvergrid and TWGrid middleware services. Three packages were developed in Perl and Java. Their entry points are a simple command line tool. The first package installs the application components (software, compounds database…) on the grid computing nodes. The second package tests these components. The third package monitors the submission and the execution of the WISDOM jobs thank to the Workload Management System and the Data Management. The used production service is LCG-2.

This abstract has presented pioneering activities in the field of grid enabled virtual screening against neglected and emerging diseases in Europe. These achievements demonstrated the relevance of large scale grids for the drug discovery process and to enable world-wide and multidisciplinary collaboration.
Using the grid to identify the most promising leads for biological tests speeds up the development process, frees up medicinal chemists' time, and concentrates their biological assays in the laboratory on the most promising components.

To illustrate the grid impact, a demonstration [8] will show the number of compounds that can be docked on several grid infrastructures

during the
conference time. Thousands docking jobs are submitted at the beginning of the
conference. The visitor can follow the progress of the experiment during the
conference time by a led display and several statistic figures (success rate, CPU
days consumed, number of jobs vs. site⋯).
The strategy for virtual screening on the grid is presented as well as the grid
infrastructures used. The demonstration visualization will be available during
the conference time on http://wisdom-demo.healthgrid.org. It receives the Best
Demo Award during the Healthgrid 2006 conference.

[1] R.W. Spencer, Highthroughput virtual screening of historic collections on the
file size, biological targets, and file diversity, Biotechnol. Bioeng 61 (1998) 61-
67.
[2] A. Chien et al., Grid technologies empowering drug discovery, Drug
Discovery Today, 7 Suppl 20 (2002) 176-180.
[3] See http://wisdom.eu-egee.fr/
[4] V. Breton, et al., Grid added value to address malaria, Proceedings of the 6-
th IEEE/ACM CCGrid conference (2006).
[5] F. Gagliardi, et al., Building an infrastructure for scientific Grid computing:
status and goals of the EGEE project, Philosophical Transactions: Mathematical,
Physical and Engineering Sciences, 363 (2005) 1729-1742
[6] N. Jacq, et al., Grid-enabled High Throughput Virtual Screening, accepted for
the proceedings of GCCB 2006, (2006)
[7] H.-C. Lee, et al., Grid-enabled High-throughput in silico Screening against
Influenza A Neuraminidase, Proceedings of the NETTAB 2006 workshop, (2006)
[8] N. Jacq, et al., Demonstration of In Silico Docking at a Large Scale on Grid
Infrastructure, Proceedings of Healthgrid conference 2006, Studies in Health
Technology and Informatics, 120 (2006) 155-157, PMID: 16823133.

Summary

The demonstration will show the number of molecules that can be docked on
several grid infrastructures during the conference time. Thousands docking jobs
are submitted at the beginning of the conference. The visitor can follow the
progress of the experiment during the conference time by a led display and
several statistic figures (success rate, CPU days consumed, number of jobs vs.
site⋯). The strategy for virtual screening on the grid is presented as well as the
grid infrastructures used. The demonstration visualization will

be available
during the conference time on http://wisdom-demo.healthgrid.org.

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

Co-authors: Mr LE MAHEC, Gaël (CNRS/IN2P3); Mr SALZEMANN, Jean (CNRS/IN2P3); Mr REICHSTADT, Matthieu (CNRS/IN2P3); Mr VERHAEGHE, Nathanaël (Healthgrid Association); Mr SPALINGER, Nicolas (Healthgrid Association); Mr BERNAT, Pierre (Healthgrid Association); Dr BRETON, Vincent (CNRS/IN2P3); Mr LEGRÉ, Yannick (CNRS/IN2P3)

Presenter: Mr JACQ, Nicolas (CNRS/IN2P3)

Session Classification: Demo session

Track Classification: Users & Applications