Enabling Grids for E-sciencE

Contribution ID: **53** Type: **Oral contribution**

Using Grid Computation to Accelerate Structure-based Design Against Influenza A Neuraminidases

Wednesday 1 March 2006 16:30 (15 minutes)

The potential for re-emergence of influenza pandemics has been a great threat since the report of that the avian influenza A virus (H5N1) having acquired the ability to be transmitted to humans. An increase of transmission incidents suggests the risk of human-to-human transmission, and the report of development of drug resistance variants is another potential concern. At present, there are two effective antiviral drugs available, oseltamivir (Tamiflu) and zanamivir (Relenza). Both drugs were discovered through structure-based drug design targeting influenza neuraminidase (NA), a viral enzyme that cleaves terminal sialic acid residue from glycoconjugates. The action of NA is essential for virus proliferation and infectivity; therefore, blocking the actives would generate antivirus effects. To minimize non-productive trial-and-error approaches and to accelerate the discovery of novel potent inhibitors, medicinal chemists can take advantage of using modeled NA variant structures and doing structure-based design.

A key work in structure-based design is to model complexes of candidate compounds to structures of receptor binding sites. The computational tools for the work are based on docking tools, such as AutoDock, to carry out quick conformation search of small compounds in the binding sites, fast calculation of binding energies of possible binding poses, prompt selection for the probable binding modes, and precise ranking and filtering for good binders. Although docking tools can be run automatically, one should control the dynamic conformation of the macromolecular binding site (rigid or flexible) and the spectrum of the screening small organics (building blocks and/or scaffolds; natural and/or synthetic compounds, diversified and/or “drug-like” filtered libraries). This process is characterized by computational and storage load which pose a great challenge to resources that a single institute can afford (For example, using AutoDock to evaluate one compound structure for 10 poses within the target enzyme would take 200 Kilobyte storage and 15 minutes on an average PC). The task to evaluate 1 million compound structures 100 poses each would cost 2 Terabyte and more than hundred years). To support such kind of computing demands, this project was initiated to develop a service prototype for distributing huge amount of computational docking requests by taking the advantages of the LCG/EGEE Grid infrastructure.

According to what we have learned from both the High-Energy Physics experiments and the Biomedical community, an effective use of large scale computing offered by the Grid is very promising but calls for a robust infrastructure and careful preparation. Important points are the distributed job handling, data collection and error tracking: in many cases this might be a limitation due to the need of grid-expert personnel effort. Our final goal is to deliver an effective service to academic researchers who for the most part are not Grid experts, therefore we adopted a light-weight and easy-to-use framework for distributing docking jobs on the Grid. We expect that this decision will benefit future deployment efforts and improve application usability.

Introducing the DIANE framework in building the service is aimed at handling the Grid applications in master-worker model, a native computing model of distributing docking jobs on the Grid. With the skeletal parallelism, applications plugged into the framework inherit the intrinsic DIANE features of distributed job handling such as automatic load balancing, and failure recovery. The python-based implementation also lowers the development effort of controlling application jobs on the Grid. With the hiding of composing JDL and of submitting jobs, users can even easily distribute their application jobs on the Grid without having Grid knowledge. In addition, this system can be used to seamlessly merge local guaranteed resources (like a dedicated cluster) with on-demand power provided by the Grid, allowing researches to concentrate on setting up of their application without facing a heavy entry barrier to move in production mode where more resources are needed.

In a preliminary study, we arranged the work into six tasks: (1) target 3D structure preparation; (2) compound 3D structure preparation and refinement, (3) compound properties and filter, (4) Autodock run (5) probable hits analysis and selection, and (6) complex optimization and affinity re-calculation. The DIANE framework has been applied to distribute about 75000 time-consuming AutoDock processes on LCG for screening possible inhibitor candidates against neuraminidases. In addition to show the distribution efficiency, advantages of adopting DIANE framework in the AutoDock application are also discussed in terms of usability, stability and scalability.

Authors: Mr LEE, Hurng-Chun (Academia Sinica Computing Center); Dr WU, Ying-Ta (Academia Sinica Genomic Research Center)

Co-authors: Mr CHEN, Hsin-Yen (Academia Sinica Computing Center); Mr HO, Li-Yung (Academia Sinica Grid Computing Center)

Presenter: Dr WU, Ying-Ta (Academia Sinica Genomic Research Center)

Session Classification: 1a: Life Sciences

Track Classification: Life Science