

# **Comparative Modeling of Protein Three-dimensional Structures on a Grid Computing Environment**

J. C. Tong<sup>1</sup>, C. H. Teo<sup>2</sup>, B. B. Jiang<sup>2</sup>, G. S. Ong<sup>2</sup>, T. W. Tan<sup>1</sup>

<sup>1</sup>Department of Biochemistry, National University of Singapore, 8 Medical Drive, Singapore 117597

<sup>2</sup>Singapore Computer Systems, 7 Bedok South Road, Singapore 469272

With the completion of the Human Genome Project (HGP) in 2003, an important challenge facing scientists today is to uncover the functions of all proteins encoded in our genome. With only a fraction of the 30,000 human genes having known structures and functions to date, protein structure prediction plays an important role in advancing our understanding of how those predicted proteins interact with each other and the environment as a system.

It is estimated that examining the entire human genome will require up to 1 million years of computational time on a single up-to-date personal computer [1]. High computing power is necessary for large scale modeling of macromolecular protein structures. A quick benchmark was performed to examine the feasibility of large-scale comparative modelling of protein structures using the state-of-the-art MODELLER software on the NUS' supercomputing grid Tera-Scale Campus Grid (TCG).

The performance of MODELLER to model protein structures using a single Pentium III 600 MHz processor and a grid computing environment of 100 nodes was evaluated. A total of 9,999 jobs consisting of protein sequences with ~200 amino acid residues were submitted to MODELLER installed in both the single Pentium III processor and the grid

cluster. The corresponding processing speed was 3628780 seconds and 259178 seconds respectively with a grid speed-up value of 14.00.

The results indicate that MODELLER can be successfully deployed on a grid environment, with significant speedup by parallel processing. An additional benefit of using a grid-enabled MODELLER and any other grid-enabled software is the ability to submit or retrieve jobs from any location (without the software installed) to and from remote computers (with the software installed) on the grid.

## **REFERENCES**

1. The Human Proteome Folding Project. <http://www.grid.org/projects/hpf>