

Grid and High Performance Computing: Opportunities for Bioinformatics Research

Albert Y. Zomaya

School of Information Technologies,
The University of Sydney,
Sydney, NSW 2006, Australia
zomaya@it.usyd.edu.au

Over the past few years the popularity of the Internet has been growing by leaps and bounds. However, there comes a time in the life of a technology, as it matures, where questions about its future need to be answered. The Internet is no exception to this case. Often called the “next big thing” in global Internet technology, Grid computing is viewed as one of the top candidates that can shape the future of the Internet. Grid Computing takes collective advantage of the vast improvements in microprocessor speeds, optical communications, raw storage capacity, World Wide Web and the Internet that have occurred over the last five years. Grid technology leverages existing resources and delays the need to purchase new infrastructure. With demand for computer power in industries like the life sciences and health informatics almost unlimited, Grids ability to deliver greater power at less cost gives the technology tremendous potential. Ultimately the Grid must be evaluated in terms of the applications, business value, and scientific results that it delivers, not its architecture. Biology provides some of the most important, as well as most complex, scientific challenges of our times. These problems include understanding the human genome, discovering the structure and functions of the proteins that the genes encode, and using this information efficiently for drug design. Most of these problems are extremely intensive from a computational perspective. One of the principal design goals for the Grid framework is the effective logical separation of the complexities of programming a massively parallel machine from the complexities of bioinformatics computations through the definition of appropriate interfaces. Encapsulation of the semantics of the bioinformatics computations methodologies means that the application can track the evolution of the machine architecture and explorations of various parallel decomposition schemes can take place with minimal intervention from the domain experts or the end users. For example, understanding the physical basis of protein function is a central objective of molecular biology. Proteins function through internal motion and interaction with their environment. An understanding of protein motion at the atomic level has been pursued since the earliest simulations of their dynamics. When simulations can connect to experimental results, the microscopic examinations of the different processes (via simulation) acquire more credibility and the simulation results can then help interpret the experimental data. Improvements in computational power and simulation methods facilitated by the Grid framework could lead to important progress in studies of protein structure, thermodynamics, and kinetics. This talk will review the state of play and show how Grid technology can change the competitive landscape.