

## Computing and Modeling for Complex Biological Systems

Ed Uberbacher and Frank Larimer  
Life Sciences Division

The Computational Biology Program is involved in integrated research activities related to the primary focus areas of the Complex Biosystems Initiative and also provides core infrastructure for analysis and data sharing among these components. These major activity areas include (1) genome analysis and data management for model organisms and microbes, (2) Development and application of methods for predicting protein structure and modeling and simulating the organization and behavior of protein complexes, and (3) analysis of systems data and data management related to the mammalian functional genomics program.

In the area of genome analysis, activities have focused on model organisms such as *Ciona intestinalis*, Fugu, mouse, which are of DOE importance as representative early chordates, for elucidating regulatory elements in the human genome (part of Goal 2 of the Genomes to Life Initiative). Additionally organisms such as white rot fungus and poplar are relevant for understanding the carbon cycle. A significant growth area is microbial genomics, where the ORNL program has been the primary analysis engine for about 50 DOE microbes, providing high quality gene calls and functional interpretation down to the level of pathway systems. This forms the foundation of current exploratory efforts to develop methods to derive regulatory-network pathway systems from microbial genome sequences.

In the molecular arena, the effort has focused on threading based protein structure prediction, the use of structural constraints in threading, and application of structure prediction technology for understanding protein complexes (GTL Goal 1). These activities are a cornerstone of the Goal 1 GTL Center recently established at ORNL. Using a combination of methods ranging from threading and homology modeling, to molecular dynamics, this effort expects to have major impact on elucidating the structure and dynamic behavior of protein machines in DOE relevant microbes, and also provide tools and computing capabilities for the larger community interested in this general area.

The mammalian functional genomics effort has developed a regional collaboratory involving seven research institutions in Tennessee and beyond. The Tennessee Mouse Genome Consortium is designed to distribute mutagenized mice and use the expertise at the several sites to analyze phenotype. The Computational Biology effort has responded to this need by constructing a variety of informatics systems to track mice and sample, electronically organize phenotype results, share data and observations, build statistical tools to analysis expression results and other experimental data types, and provide community access to the results of the consortium. A very important component of this effort is the Cryo-preserved Mutant Mouse Bank, where information systems have been built to link mouse pedigree information to phenotype, and the mouse and human genome sequences. A research component of this effort involves elucidation of regulatory networks for skin and other systems, using expression analysis and computational methods based on phylogenetic footprinting to identify regulatory motifs at the sequence level.

High performance computing has been used extensively to provide cross-cutting infrastructure for the Initiative. Many different types of analysis have been ported to HPC environments in areas of genome analysis and protein structure and dynamics. One focus here has been on the development of practical and robust methods to utilize these environments for biological codes and interfaces that allow biologists to configure and execute such analysis. The GIST system has emerged as a paradigm for meeting these needs and forms the basis for planned developments related to mass spectrometry analysis and biophysical calculations in the Goal 1 GTL Center.

Outreach and education are also an important component of the overall computational biology effort and the Complex Biosystems Initiative. The HIGMIS project, long responsible for the Human Genome Newsletter and many other outreach vehicles of DOE's Genome Program, has taken a leading role in helping DOE define the direction of its systems biology program. The HGMIS effort will continue to play an expanding role in developing the program plan for GTL and providing documents and web resources to policy makers and the research community that explain the Program's focus, impact and direction. Additionally, ORNL and the University of Tennessee are providing graduate and postdoctoral training in the area of systems biology through the Graduate Program in Genome Science and Technology. Many Computational Biology Program staff members play a role in this program and about 50% of the students enrolling in the program are pursuing a career in computational aspects of biosystems research.