



Computational biology at PNNL focuses on the development, efficient implementation and application of computational tools for the study of complex biological systems.

Computational Biology and Bioinformatics

Researchers at the Pacific Northwest National Laboratory (PNNL) are addressing the computational needs of modern biology by developing the infrastructure, databases and software necessary to formulate and test biological models and collect and manage high-throughput data.

Systems biology research involves vast amounts of “-omics” data, and new technologies are in demand to capture, store, access and analyze it. Managing the huge amounts of biological data and developing new computational methods for analysis and modeling will be key to implementing a systems approach to biology.

PNNL’s innovative approach is user-focused and works with the scientist to design intuitive tools that precisely target the unique needs of the biological science domain.

Computational Modeling

At PNNL, computational biology facilitates the understanding of cell behavior through the creation of sophisticated mathematical and computer-based models. Computational modeling provides validation as well as modeling-enriched analysis for experiments that cannot otherwise be run. Complex models and simulations are created to understand cell behavior using a broad range of biological information, including high-throughput proteomics data. These advanced, mechanistic models allow scientists to create new hypotheses to analyze the ever-expanding mass of biological data.

Network modeling improves the understanding of how cells sense their environment and respond to environmental stimuli. Molecular modeling focuses on molecular dynamics simulation and biomolecular systems analysis. Biomolecular systems currently being studied include protein-protein and protein-DNA complexes, bacterial membranes and membrane-mineral systems.

Research in computational modeling is wide-ranging at PNNL, and includes biomolecular modeling and simulation, kinetic modeling and network analysis of cell signaling and metabolic pathways. Our capabilities include:

- Computational imaging
- Atomic-level molecular modeling
- Process-level biological modeling
- Kinetic simulations and network analysis
- A modeling framework for integrated biological risk assessment.

Bioinformatics

Bioinformatics involves the design, development and application of the computer systems and software that enable scientists to explore high-throughput data from gene expression microarray experiments; mass spectrometry-based peptide and protein identification experiments; and various quantitative measures on metabolic state and metabolites. PNNL bioinformatics scientists are responsible for designing and developing novel computer systems to store and integrate

Pacific Northwest National Laboratory

Operated by Battelle for the U.S. Department of Energy

<http://www.sysbio.org>

such data with prior knowledge of biological pathways, common regulatory mechanisms, etc., to discover the constituents and states of the metabolic and signaling pathways and the connections and dynamic behavior in the genetic transcriptional and protein-protein interaction networks. In other words, the goal is to achieve as full a description of the cellular state as is possible.

This work includes, but is not limited to, reconstruction of the “wiring diagrams” of biological networks (via Bayesian network, dependency network, and other models that incorporate background information of varying confidence); simulation algorithms on network behavior; the development of algorithms linking high-throughput analysis results to biological databases for annotation and interpretation; and text mining of external public data sources.

Bioinformatics scientists at PNNL are also developing diverse computational tools and techniques to facilitate this research, including tools in the areas of high-performance computing, visualization, automated inference and machine learning. Capabilities include:

- Biological data fusion and visualization
- Development of data-mining capabilities for proteome-wide analyses

- Genomic-level data mining
- Peptide permutation and protein prediction (PQuad).

Computational and Data Infrastructure

Infrastructure provides systems and applications for data capture and analysis, as well as storage and metadata retrieval. Scientists at PNNL leverage many tools to develop systematic infrastructures and standardized systems that provide seamless access to data sharing and portals.

The design, development and implementation of computational infrastructures provides researchers with the tools necessary to support data acquisition, metadata tracking, data storage, data retrieval and analysis capabilities in a structured framework. PNNL is engaged in research to support both the experimental and computational processes for high-throughput gene expression, proteomics and metabolomic systems.

This work includes the development of laboratory information systems integrated with data management systems to provide a complete representation of the experimental data sets. Research into the design and development of col-laboratories and problem-solving environments provides resource discovery and access to a heterogeneous suite of

data, modeling, simulation and analysis tools from a common framework. Development of visualization software facilitates the analysis of experimental data and computational results. Capabilities include:

- A collaboratory for the biological sciences
- Statistically rigorous quality control for proteome analysis
- Heuristic entity relationship-building environment (HERBE)
- A computational cell environment (CCE)
- Integrated data structures for mapping cellular networks
- Proteomics research information storage and management (PRISM).

For more information contact:

George S. Michaels
Pacific Northwest National Laboratory
P.O. Box 999; MS P7-50
Richland, WA 99352
Tel: 509-372-6464
E-mail: george.michaels@pnl.gov
Web: www.sysbio.org

August 2004

PNNL-SA-37750