

## COMPUTATIONAL BIOLOGY

### Where Cornell's Great Breadth in the Sciences and Tradition of Interdisciplinary Research Pays Off

MRI images of the human brain, remote data on landscapes, protein databanks, and genomic databases contain unprecedented detailed, quantitative information about biological systems that are transforming the way researchers do almost all biology. The effective use of these and other new sources of data requires a new breed of scientist with interdisciplinary skills in computation, mathematics, statistics, and the physical sciences, as well as in biology.

Computational biology research at Cornell is concentrated in four areas: bioinformatics and data mining, biomolecular structure, evolutionary biology, and systems biology. These include such topics as DNA databases, protein structure and function, evolutionary genomics, biomechanics, computational neuroscience, and the management of natural and agricultural systems. Computational biology is an area where Cornell's great breadth in the sciences and tradition of interdisciplinary research pays off.

More than 40 faculty representing 36 graduate fields are involved in computational biology. Two faculty groups—the Computational Biology Subcommittee of the CIS and the Computational and Statistical Genomics focus area of the Cornell Genomics Initiative—assist in coordinating their activities. Both groups supplement departmental efforts in identifying and recruiting new faculty into these areas, maximizing synergy between existing faculty, and increasing the interdisciplinary research opportunities for students.

In response to the growing importance of the area, a new Department of Biological Statistics and Computational Biology was formed in the College of Agriculture and Life Sciences in 2002. This continues a pattern of expansion that began three years ago with the initiation of the Tri-Institutional Program in Computational Biology and Medicine. This program, which combines expertise from Cornell, including the Weill Cornell Medical College, the Sloan-Kettering Institute, and Rockefeller University, supports the recruitment of new computational biology faculty at each of the institutions. These faculty members will have a commitment to communication and interaction with member institutions. It also facilitates the sharing of the cutting-edge high-performance computing and visualization resources of the Cornell Theory Center among computational biologists at these institutions.

Cornell's expansive scope in the sciences and predilection for interdisciplinary activity provides a major advantage to undergraduate and graduate students. Following the recent development of new programs, undergraduates can now choose from three programs in computational biology offered by the Office of Undergraduate Biology, the

Department of Computer Science, and the Department of Biological Statistics and Computational Biology, as well as a concentration in mathematical biology in the Department of Mathematics. Also, a new graduate field specifically concentrated in computational biology is expected to be formed this year.

Descriptions of the research of a select few of Cornell's many computational biologists provides insight into the breadth of research in computational biology.

#### Bioinformatics and Data Mining

The vast amounts of data generated by human and other DNA sequencing projects are useless without computational tools that can identify the important patterns and relationships. The Computational Biology Service Unit (CBSU) of the Cornell Theory Center, supervised by Ron Elber, Computer Science, develops general and specific computational tools needed by Cornell's experimental biologists. For example, the CBSU has created new software, in collaboration with experimental biologist David B. Stern, Boyce Thompson Institute for Plant Research, to uncover potential functions of so-called "junk DNA." In collaboration with Martin Wiedmann, Food Science, CBSU has developed software to track the sources of food-borne disease outbreaks. CBSU also supports computational biology researchers at all member institutions of the Tri-Institutional Program in Computational Biology and Medicine.

#### Biomolecular Structure

Proteins carry out cellular functions ranging from muscle contraction to immune defense to vision. This is where the "rubber meets the road" at the molecular level. Knowing the structures, motions, and functions of proteins is advantageous, when not essential, for understanding their cellular functions and for practical advances in bioengineering and pharmaceutical development.

However, it is enormously more difficult (when it is possible at all) to experimentally study proteins than to determine the DNA sequences of their genes. Thus, computational tools that can theoretically predict these properties are in extreme demand. These are very challenging problems. For example, the "protein folding problem"—predicting the structure of a protein from the DNA sequence of its gene—was designated as a Grand Challenge Problem by the National Science Foundation. Cornell has been a central player in this area, beginning with the pioneering work of Harold A. Scheraga, Chemistry and Chemical Biology, more than 40 years ago. Faculty in Computer Science, L. Paul

Chew, Ron Elber, Klara Kedem, Jon M. Kleinberg, and Golan E. Yona; and in *Molecular Biology and Genetics*, David I. Shalloway, collaborate intensively, particularly in the development of new multiscale hierarchical methods for analyzing and predicting protein structure and dynamics.

### Evolutionary Biology

The complete sequences of the genomes of many organisms are now available. Comparing them in the context of evolution is a powerful way to determine gene functions and the mechanisms that govern evolution and shape genetic diversity. Analyzing single genomes—each typically containing many billions of nucleic acid constituents—and comparing multiple genomes can provide insights with important applications in medicine (for example, tailoring therapies to individual genetics) and agriculture.

In particular, genomes permit the results of experimental research that is often most effectively performed in model systems such as yeast, fruit flies, and mice to be extrapolated to humans, domestic animals, and crops. Effectively utilizing this information requires new computational and statistical tools. Faculty in Biological Statistics and Computational Biology, Carlos Bustamante, Rasmus Nielsen, and Martin T. Wells, and in *Molecular Biology and Genetics*, Andrew Clark, are developing such tools.

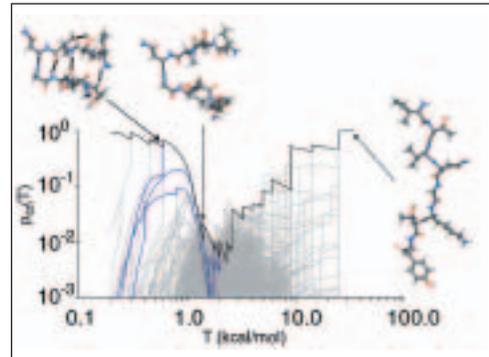
### Systems Biology

At every size scale—proteins, cells, organs, and animals—biological components act together in complex, highly organized networks whose behavior can be mathematically modeled. To list a few examples, Stephen P. Ellner, Ecology and Evolutionary Biology, studies the dynamics governing the rise and fall of insect and other animal populations. John M. Guckenheimer, Mathematics, and Steven H. Strogatz, Theoretical and Applied Mechanics, focus on understanding the emergence of nonlinear bifurcations and synchronous behavior in biological systems. Kelvin H. Lee, Chemical Engineering, mathematically models intracellular signaling.

Other examples include Shimon J. Edelman and David J. Field, Psychology; John M. Guckenheimer, Mathematics; and Christiane Linster, Neurobiology and Behavior, who study computational neuroscience. Richard H. Rand, Theoretical and Applied Mechanics, studies cardiac arrhythmias and retinal dynamics. Andy L. Ruina and Jane Wang, Theoretical and Applied Mechanics, and Francisco Valero-Cuevas, Mechanical and Aerospace Engineering, study the mechanics of biological motion.

*David I. Shalloway*

Chair, CIS Subcommittee on Computational Biology  
Greater Philadelphia Professor of Biological Sciences,  
Molecular Biology and Genetics



New hierarchical methods for understanding protein structure studied by Cornell's computational, structural biologists

▶ **For more information:**  
<http://www.cis.cornell.edu/cb/>