Combining simulation and machine learning to recognize function in 4D

Russ Biagio Altman,
Stanford University

As structural genomics projects mature, there is an increasing focus on proteins with novel folds. Unfortunately, the new structures that result often have no known function, and so there is an urgent need for methods to annotate function. Different methods can provide function in many ways, but we are specifically interested in identifying the low-level molecular function of protein structures, such as binding sites and active sites. We have previously reported a method, FEATURE (http://feature.stanford.edu/), based on supervised machine learning, that builds statistical models of sites based on a few examples, and then scans new proteins looking for a match. We have found that sometimes FEATURE misses functions because the crystal form of a protein is not in a conformation to allow us to recognize the site of interest. Working with the SIMBIOS National Center for Physics-Based Simulation (http://simbios.stanford.edu/), we are exploring the use of molecular simulation to generate structural diversity. We have found that molecular dynamics simulations of 1 nanosecond to 10 nanoseconds generate sufficient diversity such that we see conformations that are compatible with functions that are not recognizable in the static crystal structure. In this talk, I will discuss structure-based protein function annotation using machine learning, physics-based simulation of structure, and how they can be profitably combined to improve our understanding of molecular structure and function.

Russ Biagio Altman is professor of bioengineering, genetics, & medicine (and of computer science by courtesy) and chairman of the Bioengineering Department at Stanford University. His primary research interests are in the application of computing technology to basic molecular biological problems of relevance to medicine. He is currently developing techniques for collaborative scientific computation over the Internet, including novel user interfaces to biological data, particularly for pharmacogenomics (e.g. http://www.pharmgkb.org/). Other work focuses on the analysis of functional microenvironments within macromolecules and the application of algorithms for determining the structure, dynamics and function of biological macromolecules (e.g. http://simbios.stanford.edu/). Dr. Altman holds an M.D. from Stanford Medical School, a Ph.D. in medical information sciences from Stanford, and an A.B. from Harvard College. He has been the recipient of the U.S. Presidential Early Career Award for Scientists and Engineers, a National Science Foundation CAREER Award. He is a fellow of the American College of Physicians and the American College of Medical Informatics. He is a past-president and founding board member of the International Society for Computational Biology, an organizer of the annual Pacific Symposium on Biocomputing. He currently directs the Stanford Center for Biomedical Computation, and he won the Stanford Medical School graduate teaching award in 2000.