

Computation and Informatics in Biology and Medicine Training Program

Hamid Eghbalnia, Ph.D.

Assistant Scientist
Departments of Biochemistry and Mathematics
University of Wisconsin-Madison

Using NMR data for biological investigations: toward a probabilistic approach

Abstract:

Nuclear Magnetic Resonance (NMR) is uniquely powerful in its ability to provide detailed chemical and structural information at an atomic level about molecules in solution -- even when they are present in living cells or organisms. The wealth of spectral information present in NMR experiments has found a broad range of applications ranging from protein structure determination to screening in drug discovery and recently in system biology.

The potential reach of NMR applications has stimulated a substantial effort toward development of data analysis approaches that enable efficient, objective, reliable and mostly automated interpretation of data. NMR data analysis faces a set of challenges common to many areas of biological investigation. Is the resulting model consistent with experimental data? If so, is it robust to additional plausible but untested perturbations? How consistent are the models at different scales, or resolutions? What is the most informative experiment that would strengthen or reject the model? The conventional view concerning computational intractability of these questions has forced the translation of biological inquiries into less-than-optimal terms in order to use available algorithms. This is undesirable and potentially avoidable. A key insight is that organisms are highly constrained in that they have not just developed and evolved, but done so in ways that are robust to uncertainties in their environment. Therefore, modeling and inference tools need not handle arbitrary, presumably worst-case, problems but only subsets that are biologically meaningful.

This talk will draw upon published work to illustrate common themes and challenges in developing a methodology and computational infrastructure to address data analysis in NMR. We will discuss the use of probabilistic strategies for data collection and peak identification (HIFI-NMR), automated algorithms for assignments (PISTACHIO), secondary structure determination ('PECAN'), and data validation (LACS). A brief discussion of work in progress, as well as recent applications of these methods to investigation of molecular interactions will be presented.

Tuesday, September 27th, 2005
4:00 p.m.

Genetics/Biotechnology Center Auditorium
425 Henry Mall