Computational and Symbolic Systems Biology: Session Introduction

T. Ideker, E. Neumann, and V. Schachter

COMPUTATIONAL AND SYMBOLIC SYSTEMS BIOLOGY

T. IDEKER

Department of Bioengineering
U. C. San Diego
La Jolla, CA 92093
trey@bioeng.ucsd.edu

E. NEUMANN

Beyond Genomics
40 Bear Hill Road, Waltham, MA
eneumann@beyondgenomics.com

V. SCHACHTER

GENOSCOPE (National Consortium for Genomics Research)
2, rue Gaston Crémieux, F-91000 EVRY FRANCE
vs@genoscope.cns.fr

It has become increasingly evident that the use of large-scale experimental data and the invocation of 'Systems Biological' principles are gaining widespread acceptance in mainstream biology. Systems Biology involves the use of global cellular measurements—*i.e.*, genomic, proteomic, and metabolomic—to construct computational models of cellular processes and disease. It typically involves an iterative computational/experimental cycle of 1) inferring an initial model of the cellular process of interest through sequencing, expression profiling, and/or molecular interaction mapping projects; 2) perturbing each model component and recording the corresponding global cellular response to each perturbation; 3) integrating the measured global responses with the current model; and 4) formulating and testing new hypotheses for unexpected observations.

Recent technological developments are enabling us to define and interrogate cellular processes more directly and systematically than ever before, using two complementary approaches. First, it is now possible to systematically measure pathway interactions themselves, such as those between proteins and proteins or between proteins and DNA. Several methods are available for measuring protein-protein interactions at large scale—two of the most popular being the two-hybrid system and protein coimmunoprecipitation in conjunction with tandem mass spectrometry. Protein-DNA interactions, as commonly occur between transcription

factors and their DNA binding sites, are also being measured systematically using the technique of chromatin immunoprecipitation. Other types of molecular interactions and reactions, such as those involving metabolites and drugs, have been culled from the literature and stored in large, publicly-accessible databases such as MetaCyc and KEGG.

A second major approach for interrogating pathways has been to systematically measure the molecular and cellular states induced by the pathway structure. For example, global changes in gene expression are measured with DNA microarrays, while changes in proteins and metabolite concentrations may be quantitated with mass spectrometry, NMR, and other advanced techniques. The amount of quantitative data these experiments yield is on the order of thousands of individual molecular channels, and has been used to successfully identify patterns indicative of biological responses or disease states. However, it has become apparent that single genes or their products do not cause most of the biological phenomena observed. These findings have drawn researchers to the conclusion that the most interesting phenomena in biology result from the interrelated actions of many components within the system as a whole.

Recent computational approaches to Systems Biology have involved formulating both molecular interactions and molecular states into computational pathway models of various types. The amount of research in this area has exploded in recent years, as witnessed by the number of research presentations at meetings such as PSB, RECOMB, the Biopathways Consortium, and the International Conference on Systems Biology. Although much of this research has focussed on systems of differential equations and other numerical pathway simulations, a variety of model types and formalisms are in fact possible. Models may in fact be numerically computable, but they may also be symbolical and accessible to inferential logic. Logical formalisms that describe complex phenomena are just as important as is modeling molecular dynamics, and may lead to faster insight where the computational complexities are too great for a full-scale simulation. These research areas need to be pursued in parallel to more numerically-driven approaches, since they may offer a way to merge much of the symbolic knowledge derived from existing biological research.

In support of this view, almost half of the papers presented in this session involve the use of logical formalisms for modeling pathways, pathway dynamics, and/or network inference. Symbolic logic is used to analyze protein functional domains (Talcott *et al.*); to infer novel metabolic pathways using information on known pathways and the biochemical structures of their metabolites (McShan *et al.*); or to

model cell-cell interactions using a stochastic extension of the pi-calculus (Lecca *et al.*). Many of these papers combine more than one large-scale data type, including gene expression profiles, protein-protein interaction data, and/or pathway databases.

Another group of papers concentrate on either new formal representations for network inference or efficient experimental design, *i.e.* choosing an optimal set of gene deletions, overexpressions, or other experiments to maximize the information gained about the network. Of particular interest here is work by Gat-Viks *et al.* on representing gene regulation by 'chain functions'; inferring a system of differential equations through systematic overexpressions (di Bernardo *et al.*); and methods for decomposing gene expression data into its component cellular processes within a Bayesian framework (Lu *et al.*).

Finally, as an overlapping theme, several papers point to how Systems Biology may be used as part of a high-throughput drug discovery and development platform. For instance, the work by McShan *et al.* might be used to explore how newly developed drugs will be metabolised by the body; the work by di Bernardo *et al.* could be applied to predict primary drug targets based on the pathways they affect; while the work of Kightley *et al.* is a method for network inference submitted by researchers in the biotechnology/pharma industry.

The field of Systems Biology still includes many challenges and holds much promise. By increasing our repertoire of model representations and analytical formalisms, the methods explored here are the starting points for numerous advances in biotechnology, not the least of which is an enhanced ability to target therapeutics appropriately in diseased cells. Thus, we move one step closer to the day in which computational pathway modeling tehniques will have widespread impact and acceptance within basic biological research and replace high-throughput screening as a de-facto standard in "big pharma".