Structural Analysis of Cellular Networks

Jörg Stelling

Institute of Computational Science, ETH Zurich

The relationship between structure, function and regulation in complex cellular networks is a still largely open question. For an integrated and predictive description, detailed dynamic mathematical modeling meets difficulties because mechanisms and kinetic parameters are rarely available. In contrast, structure-oriented approaches only require network topology, which can often be derived from genomic data. Here, concepts and methods of structural network analysis will be discussed, in particular the computation and characteristics of elementary flux modes (EFMs). EFMs are defined as the smallest sub-networks enabling a system to operate in steady state. Importantly, they allow one to investigate the space of all metabolic states that are meaningful for the cell. The first part of the talk will focus on algorithms for EFM computation (which descend from computational geometry and are equivalent to the enumeration of extreme rays of polyhedral cones) and their application to investigate the effects of network perturbations and to predict control features. The second part will deal with extensions of EFM applications to analyze systems dynamics based on the chemical reaction network theory developed by Martin Feinberg. For an example network inspired by yeast cell cycle control, the approach allows for model discrimination, identification of key mechanisms for multistationarity, and robustness analysis. Structural network analysis, thus, provides challenging theory problems as well as broad perspectives for uncovering the organization and functionality of cellular networks.

Toric Dynamical Systems

Bernd Sturmfels

Department of Mathematics, University of California at Berkeley

We discuss a class of dynamical systems for which both the steady state loci and the parameter spaces are toric varieties. They arise in chemical mass-action kinetics as those systems that possess a complex balancing state, which is conjectured to be a global attractor. This is a mathematics lecture, aimed at outlining a possible bridge to systems biology.

Toward Genotype-Phenotype Maps for Transcription Regulation

Erik van Nimwegen

Division of Bioinformatics, Biozentrum, University of Basel

Gene expression patterns are determined to a significant extent by transcription iniiation rates, which are controlled by a 'transcription regulatory code' of small sequence motifs in the DNA that are recognized by a large array of DNA binding proteins. I will discuss Bayesian inference procedures that are used to infer the locations of regulatory motifs in DNA genome-wide by combining comparative genomic sequence data with high-throughput data such as chip-on-chip and microarray expression data. Analysis of the resulting genome-wide annotations of regulatory motifs unveils a number of qualitative and quantitative `grammatical' patterns that provide insights into the way the transcriptional regulatory code is read out to determine gene expression. I will finish by showing some preliminary work on deep-sequencing data in which quantitative models are used to predict gene expression patterns in terms of the inferred constellations of regulatory motifs at each promoter.

LC/MS Alignment by Non-negative Generalized Canonical Correlation Analysis

Bernd Fischer[1], Volker Roth[2], Joachim M. Buhmann[1]

[1]Institute of Computational Science, ETH Zurich; [2]Department of Computer Science, University of Basel

For a quantitative analysis of differential protein expression, one has to overcome the problem of aligning time series of measurements from liquid chromatography coupled to mass spectrometry. When repeating experiments one typically observes that the time axis is deformed in a non-linear way. We propose a technique to align the time series based on generalized canonical correlation analysis (GCCA) for multiple datasets. The monotonicity constraint in time series alignment is incorporated in the GCCA algorithm. The alignment function is learned both in a supervised and a semi-supervised fashion. We compare our approach with previously published methods for aligning mass spectrometry data on a large proteomics dataset.

Getting to the Root of Cell Identity

Philip N. Benfey, Siobhan Brady, Ji-Young Lee, Hongchang Cui, Richard Twigg, Jose Dinneny, and Terri Long

Biology Department and IGSP, Duke University

Most eukaryotic development begins with a fertilized egg, which undergoes a series of cell divisions to generate a multicellular organism in which the diverse cell types function in harmony. Central processes in development include creating distinctions between cells and producing coordination among different cells so that they function as units. In plants, both processes have been shown to rely heavily on cell-to-cell communication and activation and/or repression of subsets of genes. While signaling and transcription are equally important for development, high throughput techniques for identifying the nodes and links in transcriptional networks have matured more rapidly. For plants the simplifying aspects of development in an organ such as the root, make it highly tractable for the application of these approaches.

In an effort to identify the transcriptional networks that regulate cell identity in the root we are generating three datasets: 1) global expression profiles; 2) cellular localization of transcription factors; and 3) transcription factor targets. To understand the role of transcriptional networks in development, each of these datasets needs to be at cell-type specific resolution. To acquire global expression profiles we developed technology that uses sorted marked populations of cells with subsequent hybridization of the labeled RNA to microarrays. Employing this methodology, we now have data that cover most of the cell types in the root. We have also developed methods for identifying expression profiles along the developmental axis of the root. Combining these datasets gives us spatial and temporal expression patterns at high resolution. Analysis of these data provides insight into the regulatory basis of cell identity under standard growth conditions.

To investigate the response to environmental stimuli at the cellular level, we are profiling expression after cell sorting in response to a number of stresses. Many of the genes that are induced or repressed in response to these stresses in a tissue-specific manner have not been previously characterized. We are also developing technology that will allow us to follow the effects of stimuli on hundreds of genes at cell-type specific resolution over time. Initial results suggest that there is a core set of functions for each cell type overlaid with a set of cell-type specific responses to different stimuli.

This work is supported by grants from the NIH, NSF and DARPA.