Pacific Symposium on Biocomputing 4:3-4 (1999)

GENE EXPRESSION AND GENETIC NETWORKS

ROLAND SOMOGYI
Incyte Pharmaceuticals
(rsomogyi@incyte.com)

HIROAKI KITANO
Sony Computer Science Laboratory
(kitano@csl.sony.co.jp)

Biology is currently undergoing a shift from a mostly qualitative to an information rich, quantitative science. Using large-scale biological technologies, we are gaining global views of structural and dynamic information in the form of whole genome sequences and the corresponding gene activity patterns at the RNA and protein level. These data reflect the molecular workings of a complex information processing system. In many ways these systems can be effectively viewed from the perspective of genetic feedback networks, given that the fundamental step of biological information flow resides in gene activation and its control through the activity of regulatory genes.

There are obvious challenges in managing and interpreting these data to generate predictive models, which is the essential goal of any science. At this level of complexity, our analysis must move beyond caricatures depicting the putative interactions of small sets of molecules to systematic, quantitative approaches to functional inference and modeling. For this purpose we seek the appropriate computational strategies based on a solid mathematical and statistical foundation.

The following series of nine papers deals with the modeling of molecular networks, inference of functional relationships from gene activity profiles, and a networks approach to structural evolution. We begin with a review by Szallasi in which he explains why integrative approaches have been ignored in the traditional search for “dominant“ molecular genetic mechanisms, and why this is no longer tenable in light of the evidence for combinatorial molecular causes for e.g. complex human diseases. The discussion covers important topics ranging from deterministic vs. stochastic modeling, network compartmentalization, and data requirements for reliable reconstruction of network architectures.

Theoretical investigations into network behavior require computational tools for network construction, visualization and simulation. In the first of four papers, Samsonova & Serov introduce a Java applet, "NetWork", as a WWW interface for general network modeling based on the Boolean network model. For any given architecture, the applet provides the interaction diagrams and calculates the trajectories and corresponding attractors. Beyond "forward" modeling, serious challenges lie in the inverse problem of inferring network architectures from activity

data. Akutsu et al. present a rigorous mathematical investigation of this problem for discrete networks. In particular, they reach an optimistic conclusion on the practicality of "reverse engineering" in terms of data requirements and computational resources. In a departure from discrete models, Weaver et al. examine a continuous, linear modeling scheme that demonstrates stable and cyclical behavior consistent with biological observations. They also provide a reverse engineering algorithm that performs remarkably well even in the presence of noisy input data. Chen et al. present an extensive non-linear modeling framework based on differential equations. They take into account transcriptional and translational control, and time delays. Like Akutsu et al. and Weaver et al., the authors conclude that only a relatively small number of activity measurements is necessary for the inference of a largely correct model of network architecture. In contrast, Erb & Michaels draw a less optimistic conclusion regarding modeling stability. Using sensitivity analysis within a non-linear differential equation model, small errors in parameter estimates in their three gene model circuit prove to cause large errors in model prediction.

Four papers deal with the analysis of specific biological examples. D'haeseleer et al. apply a linear modeling approach to the first-order inference of regulatory connections from times series of gene expression data in CNS development and injury. Given the limited number of measurements, the analysis results in a remarkably sparse interaction matrix and provided good reconstructions of the training data set. While the specific predictions of gene interactions are speculative at this stage, they may serve as guides for future experimental design. Kyoda & Kitano simulate spatio-temporal patterning in development. Their model of six interacting genes faithfully reproduces the protein expression profiles along the primary axes in Drosophila leg formation and provides two new predictions on gene interactions. Going beyond strictly deterministic systems, Goss & Peccoud introduce a stochastic modeling approach to bacterial plasmid stability. Nagl et al. extend the network perspective to the intramolecular evolution of steroid receptor binding sites. Using an information theoretic approach, they demonstrate that ligand binding residues evolve within a network of covarying positions on the protein sequence.