

Reverse Engineering Gene Regulatory Networks

Gene regulatory networks (GRNs) are pathways of genes whose induced proteins regulate the expression of other genes and their products. They orchestrate biochemical processes that specify spatial and temporal patterns or govern the formation of tissues and organs. GRNs reveal the causality of these processes through activation or repression of targets by regulatory proteins. With current technologies, activity levels of each biomolecule in a network can be measured directly, while causal linkages remain unobservable. A challenge for molecular biologists is to identify the causal links that constitute the pathways in a GRN. Identification and ultimately control of these signaling pathways are important first steps in repairing developmental defects, including those that cause tissue-specific cancers.

Mathematical and statistical tools have been employed to *reverse engineer*, or reconstruct, the pathways in a GRN from microarray data. The data, which record expression levels of genes, are typically limited to tens of measurements, while the number n of biomolecules in a GRN is in the hundreds to thousands. A standard reverse-engineering approach can be described as follows: 1) choose a modeling class, such as Boolean networks or linear differential equations; 2) use biological properties to constrain the class, for example, by limiting the number of links per biomolecule; and 3) construct a model defined by a set of n functions that fit the data. The pathway structure can then be extracted from the model.

One limitation of this paradigm is that typically only a small number of models are generated and they may not be sufficiently diverse to accommodate the level of complexity inherent in GRNs. Furthermore, since the data vastly underdetermine the network, the number of models that fit the data may be considerably larger than the number of models that are constructed. Since the true pathway structure may not be known, it is desirable to build a large set of potential models from which the most realistic model can be chosen.

The above limitation was overcome by a recent novel method, proposed by MBI Postdoctoral Fellow Brandilyn Stigler and Virginia Bioinformatics Institute Research Professor Reinhard Laubenbacher. In their setting, the modeling class consists of polynomial dynamical systems (PDSs), which are generalizations of Boolean networks and are state- and time-discrete analogs of continuous dynamical systems. Rooted in computational algebra, the algorithm encodes in a zero-dimensional ideal all PDSs that fit a given data set and selects an optimal PDS by computing a Grobner basis for the ideal. One advantage of this method is that the compact representation of the complete model space provides a rigorous framework within which to run analyses. Stigler is now working with MBI Long-term Visitor Winfried Just on modifications of the algorithm to improve its computational performance. They are also developing new selection strategies that will increase pathway identifiability. The algebraic method is currently being tested by Ohio State University molecular geneticist Helen Chamberlin on a newly discovered regulatory network for embryonal development in the nematode *C. elegans*.

