Genetic Networks

William F. Loomis and Paul W. Sternberg

The interactions of biological macromolecules and the flow of regulatory information that controls development, behavior, and homeostasis can be considered a genetic network. The nodes in such networks are genes or their RNA and protein products. The connections are the regulatory and physical interactions among the RNAs, proteins, and cis-regulatory DNA sequences of each gene. Modern molecular genetic techniques have greatly increased the rate at which genes are being recognized and their primary sequences determined. The challenge is to link the genes and their products into functional pathways, circuits, and networks. Analyses of regulatory networks (such as those involving signal transduction and transcriptional regulation cascades) illustrate combinatorial action that implements, for example, digital logic, analog-digital conversions, cross-talk and insulation, and signal integration. Although the existence of sophisticated network elements has been suggested by decades of physiological studies, what is new is the scale and detail becoming available for the components. Much of current molecular biology focuses on identifying new components, defining the regulatory inputs and outputs of each node, and delineating the physiologically relevant pathways.

Intensive analysis of individual nodes reveals how many inputs and outputs potentially exist. For example, examination of an individual protein might reveal three distinct binding sites for other proteins, or analysis of the cis-regulatory regions of a gene might reveal that 20 different proteins specifically interact with those sequences. Understanding the substrate specificity of a protein kinase suggests potential targets solely on the basis of primary sequence. Elucidation of networks can benefit from powerful new tools for identifying interacting components, including genetic enhancer and suppressor screens, two-hybrid library screening, and affinity chromatography. Genetic screens for mutations that alter the phenotype of other mutations can identify novel interactions and components but do not guarantee the directness of the connection. Biochemical detection methods also identify novel interactions but do not guarantee their functional relevance. Nevertheless, when used together with independent tests, these two methods are rapidly adding to the wealth of connections that complement the results from genome sequencing projects.

Mutational analysis allows the function of each gene to be examined by determining the effect of its elimination or alteration on the organism's phenotype. Genes can be altered one at a time to assess their individual contribution, or genomes can be generated with multiple changes to understand the relations of the specific genes and their products. In the best cases, the loss of a specific gene results in a clear physiological consequence that indicates the normal role of the gene in a well-defined process. However, site-directed mutants in multicellular organisms often have phenotypes that differ little from wild type, indicating partial functional redundancy or complete irrelevance of the gene in the process under study. The successful analysis of microbial metabolic pathways required all the power and efficiency of microbial genetics to isolate and subtly perturb specific genes before the intricacies of feedback loops and alternate controls became clear.

A sculptural network. Mozart II, outside the library at Stanford University. [Photo by A. Kuspa]

Some networks, like the Brooklyn bridge or the telephone system, string together independent wires, supports, and stays to counter and respond to differing pressures and requirements. Breaking any given connection will not bring down the whole structure, although it may shift it somewhat. Some networks underlie homeostasis; others drive spatial and temporal changes during ontogeny. Biological regulatory processes have been added during evolution to adapt gene expression and function to different needs. Genetic networks are the product (and material) of evolution, and the evolutionary divergence of particular networks can be used to understand their properties. Apparent redundancy is a case in point. Laboratory bioscience is not good at detecting 1 percent differences in output, but natural selection is. Often another organism has expanded the 1 percent such that it can be easily measured in the laboratory.

When mutational and epistatic studies indicate that a simple linear sequence of events is functioning in a specific process, biologists draw a pathway, connecting steps with arrows and assigning different steps to different genes. When the results demand it, branch points and converging pathways are drawn. Such two-dimensional modeling is useful for presenting the results as well as the interpretations, but at some point it will be useful to mathematically abstract the processes. Computer scientists have long simulated networks and have developed symbolic techniques for their display. Sometimes biological problems can be successfully treated as closed systems with simple Boolean logic in electrical engineering terms, a superb recent example being the lytic and lysogenic cycles of phage lambda described by McAdams and Shapiro in this issue (*1*). Modeling of this sort can emphasize aspects that might not be appreciated intuitively and can delineate missing components. For the time being, detailed modeling may be restricted to simple systems because it requires good quantitation of a high proportion of the relevant steps, which is easier said than done in most organisms.

As our knowledge of networks expands, the use of computers will become increasingly important to store and sort the data. Indeed, computer storage and retrieval of biological information is presently routine, as exemplified by the fact that few remember the sequence of a specific gene but most can quickly bring it up on their computer screen along with all related sequences. Sequence comparison and analysis is now an established computer skill, and a similar approach to functional genetic data could become equally routine. Databases centered on the genes of organisms with advanced genome projects are leading the way, many based on the database format designed for *Caenorhabditis elegans*, ACeDB. Determining the function of each gene is more difficult, because different lines of work often lead to quite different visions of how a gene works. Initially, it may be best to include all viewpoints and let the field sort it out. However, as the number of genes in a network increases, it will be awkward if clear definitions of function are not available. Thus, we are brought back to the need for more detailed experimental data on each gene in a wide variety of genetic backgrounds. As these accumulate, general properties of networks may be recognized. When predictions are fullfilled, even if only in rare cases, we can be more confident that we are in the right neighborhood.

W. F. Loomis is in the Department of Biology, University of California, San Diego, La Jolla, CA 92093, USA. P. W. Sternberg is in the Department of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

Reference

1. H. McAdams and L. Shapiro, *Science* **269**, 650 (1995).