Chapter 17
Determining the Properties of Gene Regulatory Networks from Expression Data

Larry S. Liebovitch
Florida Atlantic University, USA

Lina A. Shehadeh
University of Miami, USA

Viktor K. Jirsa
Florida Atlantic University, USA

Marc-Thorsten Hütt
Jacobs University, Germany

Carsten Marr
Helmholtz Zentrum München, Germany

ABSTRACT

The expression of genes depends on the physical structure of DNA, how the function of DNA is regulated by the transcription factors expressed by other genes, RNA regulation, such as that through RNA interference, and protein signals mediated by protein-protein interaction networks. We illustrate different approaches to determining information about the network of gene regulation from experimental data. First, we show that we can use statistical information of the mRNA expression values to determine the global topological properties of the gene regulatory network. Second, we show that analyzing the changes in expression due to mutations or different environmental conditions can give us information on the relative importance of the different mechanisms involved in gene regulation.

DOI: 10.4018/978-1-60566-685-3.ch017

INTRODUCTION

All living things contain a "memory" of the past that explicitly defines their species and implicitly reflects the evolutionary events that led to their species. Typically, this memory is encoded in deoxyribonucleic acid, DNA, although it may also be encoded in ribonucleic acid, RNA (for retroviruses), or in epigenetic coding (such as methylation of DNA), or in three dimensional structures (such as the protein confirmations of prions). Each organism uses this memory as a blueprint to design and maintain itself. But it is not like a blueprint that we use to build buildings which is a smaller symbolic picture of the building. Rather, it is more like a computer code, which when executed generates structures that have a very different form than the code itself. But it is unlike the computer codes that we currently construct. Our computer codes execute their instructions in a preset order. However, which instructions living things execute are chosen by a multilevel cacophony of highly interacting networks.

The Central Dogma of molecular biology (Crick, 1958) was that genetic expression is a one way street from the transcription of DNA into mRNA, and then the translation of mRNA into protein. But we are now beginning to appreciate that multiple processes, both forward and backward, control and edit how the instructions of DNA are executed into the proteins that form the structure and function of cells. In this chapter we explore how networks control DNA expression, from within DNA (depending on the physical structure of DNA and the regulation that one gene exerts on another), and from outside of DNA (depending on the editing of mRNA and protein regulatory networks). We show how understanding the physics of networks can be used to devise methods of analysis that reveal the global and local organization of these networks.

BACKGROUND

Transcription Regulatory Networks (TRN)

In transcriptional regulation, the product of one gene, a transcription factor (TF) protein, binds to the promoter region of another gene and increases or decreases its expression. The discovery of regulatory processes in the *lac* operon (Jacob & Monod, 1961) marked an historic step in biology. Lately, the assembly of many such effects into a full network of genetic interactions has heralded the emergence of a system-wide view on transcriptional regulation (Thieffry et al., 1998): The transcriptional regulatory network (TRN) describes how genes regulate each other through the expression and binding of their TFs. In mathematical terms, the TRN is a directed graph consisting of nodes representing genes and links representing the directed regulatory interaction between two genes, mediated by a TF. The statistics of the topology of these connections are summarized by the in-degree and out-degree distributions which define the number of genes with a given number of incoming and outgoing connections.

In the bacterium *Escherichia coli*, evidence for the regulatory action of TFs is documented in what is ''currently the largest electronically-encoded database of the regulatory network of any free-living organism'' (Salgado et al., 2006), called RegulonDB. The most recent version (5.6) of the publicly available database comprises 2735 interactions between 1345 genes. A small fraction of genes (79) are top-level regulators with no input from other genes, while 1197 nodes are solely target nodes, with no regulatory output to other genes. Various studies used the information contained in RegulonDB to construct the *E. coli* TRN and, e.g., analyzed its motif content (Shen-Orr et al., 2002), the aggregation of such motifs