Mathematical Biosciences 231 (2011) 61–68

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Structure–function relations are subtle in genetic regulatory networks

Michael E. Wall

Computer, Computational, and Statistical Sciences Division, Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87505, USA

ARTICLE INFO

Article history:
Available online 15 February 2011

Keywords:
Gene regulation
Function prediction
Network motifs
Escherichia coli
Computational biology
Synthetic biology

ABSTRACT

Recent studies have yielded insights into structure–function relations in genetic regulatory networks. Models of feed-forward loops show that the input–output behavior depends critically on the input signal as well as transcription interactions. Models of induction of the *lac* operon in *Escherichia coli* reveal the importance of metabolism in determining genetic regulatory network behavior. Combined experimental and computational studies of activation by MarA in *E. coli* show how mechanisms of transcription regulation, hidden at the level of genetic regulatory networks, can influence behavior. Together these studies illustrate that gene regulation is critically influenced by factors beyond the topology of genetic regulatory interactions. Prediction of the specific information processing roles of gene circuits is more difficult than we would like, but it is still possible. Thinking about evolution of proteins and networks might make it easier.

1. Introduction

Cells control their destiny by tuning the production of RNA and proteins. This tuning process is known as gene regulation. Gene regulation works in many different ways and can involve the interaction of a variety of factors including proteins, DNA, RNA, and small molecules. An important mechanism of gene regulation is transcription regulation, in which proteins called transcription factors bind to DNA near target genes and either increase or decrease production of mRNA.

The functional binding between a transcription factor and a specific location on the DNA specifies a transcription interaction. The interaction might have an associated sign to indicate whether it increases (+) or decreases (−) transcription of the target gene. A transcriptional regulatory network (TRN) is defined by a set of these interactions. A major challenge in molecular biology is to predict how the interactions represented by a cell's TRN work together to dynamically govern cellular behavior.

A cell's entire TRN is too large to manage without the aid of computers, so there are databases that hold the information. There are 2594 interactions involving 179 TFs in *Escherichia coli* documented in release 6.7 of the RegulonDB database [1]. Some local regions of the *E. coli* network have been extensively characterized [2]. Some of these regions are wired into gene circuits that have a well-defined input and output signal, and many of these circuits perform similar functions, such as producing catabolic enzymes when substrate levels are high, or blocking production of anabolic enzymes when product levels are high. There are preferred patterns in the wirings of different circuits that perform similar functions. Some wiring patterns perform their functions better than others, which explains some of the preferences [3].

Our knowledge of the transcription regulation network of *E. coli* and other organisms is extensive but incomplete. Nevertheless, this has not held back global statistical analysis of TRNs. The number of TFs controlling each gene in the known yeast network is exponentially distributed, and the number of genes each TF controls is governed by something closer to a power-law distribution [4]. Most TFs in *E. coli* block expression of their own gene [5]. Some recurring wiring patterns in local regions of the *E. coli* TRN, called network motifs, appear to be found more often than would be expected by a random chance model [6]. The same is true for the yeast TRN [7,8].

Even though our knowledge of the TRNs of *E. coli* and yeast is extensive, many of the functions carried out by these and other TRNs remain unknown. Experiments to reveal these functions are generally time consuming and costly. It would help enormously if we could identify the functions from the local wiring patterns of the TRN, like looking up the input–output function of an integrated circuit in the manufacturer's manual.

A major obstacle to identifying the function of a wiring pattern is the need to identify the input signal and how it interacts with the TRN. For example, consider the wiring pattern in Fig. 1A where a transcription factor X represses expression of both X and Y. The function depends on how the signal S interacts with the circuit. If S blocks repression of X (Fig. 1B) then an increase in S leads to an increase in X and a decrease in Y. If instead S blocks repression of Y (Fig. 1C) then an increase in S has no effect on X and leads to an increase in Y. If S blocks repression of both X and Y (Fig. 1D) then an increase in S leads to an increase in X, but the effects of X and S on Y are opposite, making the total effect on Y uncertain in the

E-mail address: mewall@lanl.gov

0025-5564/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2011.02.003

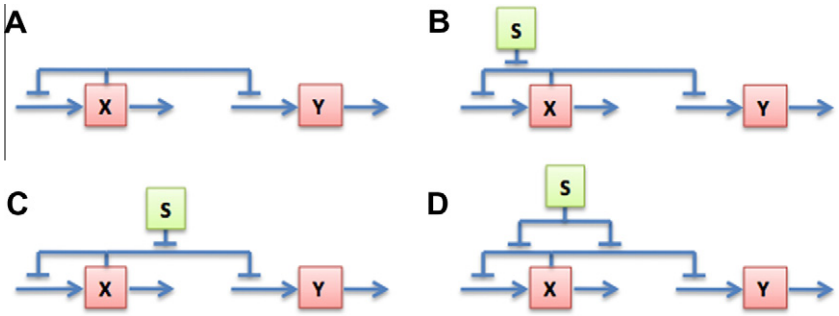

Fig. 1. Multiple functions of an elementary gene circuit. (A) Transcriptional regulatory interactions without a signal interaction. (B) The input signal S blocks repression of X. When S increases X increases and Y decreases. (C) S blocks repression of Y. When S increases X stays the same and Y increases. (D) S blocks repression of both X and Y. When S increases X increases and the effect on Y is uncertain without more information.

absence of additional information. If X is the input signal itself then in this case we have all of the information we need, but there are more complicated circuits involving feedback loops where the effect of X on Y would still be ambiguous without more information.

The example in Fig. 1 raises a general question: What information do we need to determine the functions of a gene circuit? Here I review several studies that address this question. The first section examines the functions of feed-forward loop gene circuits [9,10]. The second examines the role of metabolism in determining gene circuit function [11]. The third examines regulation of different promoters by the same global transcription factor [12,13]. I will conclude by summarizing what these studies tell us about determining gene circuit function, and describe some related work and ideas. In particular, we should seek help from comparative genomics and evolutionary arguments.

2. Multiple functions of feed-forward loops

Shen-Orr et al. [14] found that some recurring wiring patterns in local regions of the *E. coli* TRN, called network motifs, appear to be found more often than would be expected by a random chance model. Similar motifs were found in the yeast TRN [7,8]. One of these motifs is the feed-forward loop (FFL) where Y regulates Z and X regulates both Y and Z. The FFL with all positive regulatory interactions can act as a sign-sensitive accelerator, e.g., in response to a change in an input signal, the level of Z ramps up faster than it ramps down [14]. The FFLs with other combinations of regulatory interactions have been associated with other functions [15]. If the functions of these and other network motifs were robust they would provide enormous insight into the functions of TRNs [16,17]. How robust are they?

To gain insight into the robustness of network motif functions, we analyzed a comprehensive set of FFL models [9]. The models spanned eight combinations of regulatory interactions (positive or negative for each of the three interactions), and 27 combinations of input signal effects (enables, blocks, or has no influence on each of the three regulatory interactions). The dynamics of each of the 216 (= 8×27) models were simulated in response to a timed pattern of step changes between a high and low level of the input signal. Wide ranges of parameter values were sampled for each model. The FFL functions were clustered using a greedy algorithm, and the quality of the clustering vs. the number of clusters was evaluated using the maximum cluster radius as an error measure (see [9] for details of the method). In each case the error decreased sharply as the number of clusters increased to a break point after which the error decreased less sharply. This break point was identified and used to select a single number of clusters among the many possible choices.

The FFL functions clustered into about 15 distinguishable phenotypes (Fig. 2), which is more than 10 times fewer than the number of models, but about twice as many as the number of combinations of genetic regulatory interactions. The robustness of each circuit was quantified by calculating the Shannon entropy of its phenotype distribution, defined as

$$S = -\sum_{i=1}^{N_c} p_i \log_2 p_i, \tag{1}$$

where p_i is the fraction of functional responses that fall into cluster i, and N_c is the total number of clusters. None of the FFLs had a robustly determined phenotype without specifying the input signal effects. Some circuits exhibited just one possible phenotype ($S = 0$) when both the regulatory interactions and the input signal effects were specified and were thus highly robust. Other circuits still exhibited a handful of phenotypes (typically between 2 and 5). For the latter circuits typical entropies of the phenotype distributions were between 0.5 and 2 bits, indicating that their functions are not robust. Our finding that network motifs can exhibit multiple functions was supported by a computational study of the bi-fan network motif by Ingram et al. [18]. Mugler et al. [19] similarly found that the functions of circuits involving three cascading transcription interactions, each controlled by its own signal, could exhibit a rich functional repertoire.

To further characterize the robustness of FFL functions, we analyzed the FFL with all negative regulatory interactions (known as the type 2 incoherent FFL [15]) in greater detail [10]. We identified three possible examples of this type of FFL in *E. coli: galR-galS-galETKM*;*exuR-uxuR-uxuAB*; and *gntRKU-idnDOTR-gntKU*. (The elements of each FFL are listed in the order *x–y–z*.) A comprehensive set of signal effects were considered. Time-dependent patterns of step changes in the input signal, this time including an intermediate as well as high and low signal levels, were used to simulate the dynamics. The resulting time courses were clustered to define a set of dynamical phenotypes. A complementary set of steady-state phenotypes was obtained by calculating the steady-state output pattern that results from a given time ordering of step changes in the input signal and classifying it according to the time ordering of high, low, and intermediate outputs.

The output behaviors of this FFL were highly diverse. The FFL exhibited all possible steady-state phenotypes in response to a low–intermediate–high–intermediate–low pattern of step changes in the input signal. Even specifying the input signal was not sufficient to uniquely determine the phenotype: entropies of the phenotype distributions varied between about 0.6 and 2.5 bits. Many parameter value combinations yielded non-functional circuits: the percentage varied from 17% to 84% depending on the input signal effects. Analysis of the dynamical phenotypes revealed a