Quantitative dissection of the simple repression input–output function

Hernan G. Garcia[a] and Rob Phillips[b,1]

Departments of [a]Physics and [b]Applied Physics, California Institute of Technology, Pasadena, CA 91125

Edited* by Curtis G. Callan, Princeton University, Princeton, NJ, and approved May 26, 2011 (received for review October 18, 2010)

We present a quantitative case study of transcriptional regulation in which we carry out a systematic dialogue between theory and measurement for an important and ubiquitous regulatory motif in bacteria, namely, that of simple repression. This architecture is realized by a single repressor binding site overlapping the promoter. From the theory point of view, this motif is described by a single gene regulation function based upon only a few parameters that are convenient theoretically and accessible experimentally. The usual approach is turned on its side by using the mathematical description of these regulatory motifs as a predictive tool to determine the number of repressors in a collection of strains with a large variation in repressor copy number. The predictions and corresponding measurements are carried out over a large dynamic range in both expression fold change (spanning nearly four orders of magnitude) and repressor copy number (spanning about two orders of magnitude). The predictions are tested by measuring the resulting level of gene expression and are then validated by using quantitative immunoblots. The key outcomes of this study include a systematic quantitative analysis of the limits and validity of the input–output relation for simple repression, a precise determination of the in vivo binding energies for DNA–repressor interactions for several distinct repressor binding sites, and a repressor census for Lac repressor in *Escherichia coli*.

physical biology | thermodynamic models | protein copy number | *lac* operon

It is now possible not only to make quantitative, precise, and reproducible measurements on the response of a variety of different genetic regulatory architectures, but also to synthesize novel architectures de novo. These successes have engendered hopeful analogies between the circuits found in cells and those that are the basis of many familiar electronic devices (1, 2). However, in many cases, unlike the situation with the electronic circuit analogy, our understanding of these circuits is based upon enlightened empiricism rather than systematic, quantitative knowledge of the input–output relations of the underlying genetic circuits.

Regulatory biology has shed light on the space–time response of a wide variety of these genetic circuits. Examples range from the complex regulatory networks that govern processes such as embryonic development (3, 4) to the synthetic biology setting of building completely new regulatory circuits in living cells (5). In particular, the dissection of genetic regulatory networks is resulting in the elucidation of ever more complex wiring diagrams (see, as an example, ref. 6). With these advances it is becoming increasingly difficult to develop intuition for the behavior of these networks in space and time. In addition, often, the diagrams used to depict these regulatory architectures make no reference to the census of the various molecular actors (the intracellular number of polymerases, activators, repressors, inducers, etc.) or to the quantitative details of their interactions that dictate their response. As a result, there is a growing need to put the description of these networks on a firm quantitative footing.

Often, the default description of regulatory response is offered by phenomenological Hill functions (7–12), which in the case of repression have the form

$$\text{gene expression level} = \frac{\alpha}{1 + ([R]/K_d)^n} + \beta, \qquad [1]$$

where n is the Hill coefficient that determines the sensitivity of the gene regulatory function, K_d is a dissociation constant, and α and β are constants that determine the maximum and basal levels of expression, respectively. Although such descriptions might provide a satisfactory fit of the data, they can deprive us of insights into the mechanistic underpinnings of a given regulatory response or, worse, can force us into thinking about the behavior of a given circuit in a way that is not faithful to the known architecture.

Alternatively, using thermodynamic models, it has been shown for a wide class of regulatory architectures that for each and every circuit, one can derive a corresponding "governing equation" that provides the fold change in gene expression as a function of the relevant regulatory tuning variables (13–15). The goal of our work is to carry out a detailed experimental characterization of the predictions posed by one such governing equation for the regulatory motif describing simple repression (Fig. 1*A*) in which a repressor can bind to a site overlapping the promoter, resulting in the shutting down of expression of the associated gene. This is a particularly fundamental case study because in *Escherichia coli* alone, there are >400 circuits that are regulated by different transcription factors that repress by binding to a single site in the vicinity of the promoter (16). Indeed, simple repression and activation are often thought of as the elementary ingredients of a much more diverse range of real regulatory circuits (17, 18).

As seen in Fig. 1, the level of expression in circuits governed by simple repression can be tuned by several different parameters. One of the key tuning variables in nearly all regulatory and signaling networks is the numbers (or concentrations) of the relevant molecular players in the process of interest. We use the repressor number as one of the main tunable parameters in the experiments described below, with a 100-fold range of different repressor counts considered. To explore our understanding of how this parameter dictates regulatory response, we need to know how many repressors our strains of interest harbor. A series of beautiful recent experiments has made important progress in carrying out the molecular census, using a variety of clever methods. These molecular counts include the census of all actin-related proteins in *Schizosaccharomyces pombe* cells (19), a count of essentially *all* the proteins in *Saccharomyces cerevisiae* cells (20), a determination of the distribution of both lipids and proteins in synaptic vesicles (21), and several counts of the proteins in *E. coli* (22, 23) and other cell types as well (24). Most relevant to the current work is a recent experiment using a fluctuation-based counting method to determine the number of transcription factors in *E. coli* that control a synthetic circuit of interest (10). Our work adds a twist to protein census taking by using thermodynamic models as a way to count the number of repressors in a simple regulatory motif.

Quantitative control of the absolute number of transcription factors is seldom used in experiments that aim to dissect regulatory architectures even though it is one of the main strategies to verify the predictions from thermodynamic models (13–15). Previous work has usually relied on the control of an external

Author contributions: H.G.G. and R.P. designed research; H.G.G. performed research; H.G.G. and R.P. analyzed data; and H.G.G. and R.P. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: phillips@pboc.caltech.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015616108/-/DCSupplemental.

SYSTEMS BIOLOGY