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

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

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

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. 1A) 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

---

SYSTEMS BIOLOGY

Fig. 1. The simple repression motif. (*A*) States and weights of the thermodynamic model describing this regulatory motif. We assume that Lac repressor sterically excludes RNA polymerase from the promoter, although that assumption is not critical to our analysis. *P* and *R* are the numbers of RNA polymerase and Lac repressor molecules inside the cell, respectively. N_{NS} is the number of nonspecific sites, which we assume to be the size of the genome. $\Delta\varepsilon_{pd}$ and $\Delta\varepsilon_{rd}$ are the difference in energy between being specifically and nonspecifically bound for RNA polymerase and Lac repressor, respectively. The difference in color in the repressor binding site denotes an overlap of the binding site with the promoter. (*B*) The tuning variables that can be varied in the model and controlled experimentally are the binding strength (by changing the Lac repressor operator

sequence) and the number of Lac repressors (by changing its mRNA ribosomal binding site). The effect of tuning these parameters on the fold change in gene expression is shown in the graphs. Note that stronger repressor binding corresponds to a larger fold change. For a detailed derivation of the expression and discussion of the assumptions used see *SI Text* and Fig. S2.
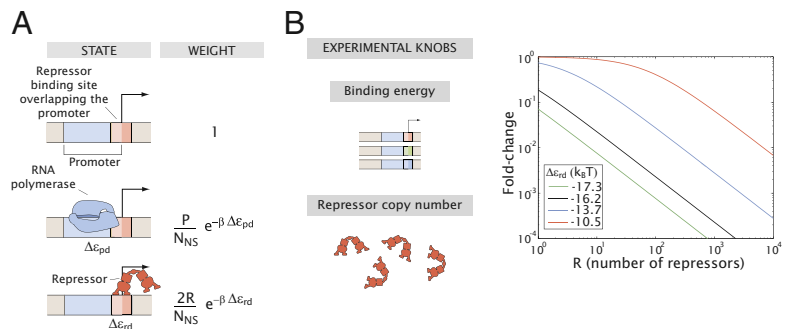
inducer to vary the regulatory response of a genetic circuit (5, 9, 11, 12, 25, 26). However, the use of inducer molecules, although experimentally convenient, adds another layer of complexity to the modeling approach and has been systematically characterized in only a few cases (11).

Recent measurements (10, 23, 27–31,) have also often focused on the variability or "noise" associated with transcriptional regulation. Although there has been great recent interest in this gene expression variability, we argue that a crucial quantitative prerequisite to fully dissecting the properties of genetic networks is a viable description of their mean response, and any conceptual frameworks used to describe the noise must first be consistent with these mean responses.

In this work we test these thermodynamic models of transcriptional regulation by generating *parameter-free predictions* for the level of gene expression as a function of the regulatory tuning variables of the simple repression architecture. We show significant agreement between the theoretical description and the measurements over multiple orders of magnitudes of the inputs and outputs of the system. We conclude that through thermodynamic models we can accurately predict the level of regulation due to simple repression, opening the door to the design of synthetic genetic circuits where the level of gene expression can be tuned theoretically and to the better interpretation of the transcriptional response of naturally occurring circuits.

Theory and Experimental Design

Although our analysis should be relevant generically for simple repression, the reasoning behind our experiments is based upon a series of earlier measurements and calculations on the level of repression in the specific case of the *lac* operon (32, 33). In particular, we consider the case where there is only a single specific binding site for the Lac repressor (Fig. 1). The wild-type *lac* operon was rewired such that only the main operator was present and then, in turn, different strains were constructed in which the strength of that main operator was systematically weakened according to the progression Oid to O1 to O2 to O3 shown in Fig. S1.

Thermodynamic models assume that the processes leading to transcription initiation by RNA polymerase (RNAP) are in quasi-equilibrium. This assumption means that we can use the tools of statistical mechanics to describe the binding of RNA polymerase and transcription factors (TFs) to DNA. Further, the level of gene expression is assumed to be proportional to the probability that RNA polymerase is bound to the promoter of interest (13, 34). This probability is determined, in turn, by the interactions between polymerase and the promoter and competition for those binding sites by repressors. In Fig. 1*A* we show the thermodynamic states and weights corresponding to a minimal model of the simple repression regulatory motif. In this simplified model the promoter can be found in only one of three states: (*i*) empty, (*ii*) occupied by RNA polymerase, and (*iii*) occupied by Lac repressor. The partition function for this system is obtained by summing over the statistical weights of each of these states and is given by

$$Z = \underbrace{1}_{\text{promoter empty}} + \underbrace{\frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}}}_{\text{RNA polymerase bound}} + \underbrace{\frac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rd}}}_{\text{LacI bound}}, \quad \textbf{[2]}$$

where *P* is the number of RNA polymerase molecules, *R* is the number of Lac repressor tetramers, and $N_{NS} \sim 5 \times 10^6$ is the number of nonspecific DNA sites (the length of the genome), corresponding to the reservoir for both molecules. $\beta = (k_B T)^{-1}$ with k_B being the Boltzmann constant and *T* the absolute temperature. The energies $\Delta\varepsilon_{pd}$ (RNA polymerase–DNA) and $\Delta\varepsilon_{rd}$ (repressor–DNA) correspond to the difference between specific and nonspecific binding for RNA polymerase and Lac repressor, respectively, where we make the simplifying assumption of a homogeneous nonspecific background. The factor of 2 in front of the number of Lac repressors stems from the fact that this molecule is a tetramer, a dimer of dimers, with two binding heads. Therefore, 2*R* corresponds to the number of binding heads inside the cell. For a complete derivation of these terms, please refer to refs. 14 and 35, *SI Text* and Fig. S2.

The probability of finding RNA polymerase bound to the promoter is then given by

$$p_{\text{bound}} = \frac{\dfrac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}}}{Z}, \quad \textbf{[3]}$$

where *Z* is the partition function defined in Eq. **2**. A much more convenient quantity to measure is the fold change or relative change in gene expression due to the presence of the transcription factor; namely,

$$\text{fold change} = \frac{p_{\text{bound}}(R \neq 0)}{p_{\text{bound}}(R = 0)} = \frac{1 + \dfrac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}}}{1 + \dfrac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}} + \dfrac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rd}}}. \quad \textbf{[4]}$$

The great advantage of this quantity is that it is easily accessible both theoretically and experimentally. It is unitless and can be measured by comparing the levels of gene expression (in any arbitrary or absolute units) when Lac repressor is present and absent. We define this fold change in gene expression with respect to the absence of transcription factor and not with respect to a state where the transcription factor is fully induced such as in the presence of saturating concentrations of Isopropyl β-D-1-thiogalactopyranoside (IPTG). Using inducers would require us to consider the induction process explicitly (11). In the case of a weak promoter such as *lac*UV5 used in this work (ref. 15 and *SI Text*) the term $(P/N_{NS}) e^{-\beta\Delta\varepsilon_{pd}} \ll 1$. This outcome results in the fold change collapsing to the simpler form,

$$\text{fold change} = \frac{1}{1 + \frac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rd}}}. \qquad [5]$$

This last expression serves as the basis of our experimental design where we identify two tuning variables that can be controlled experimentally in a systematic fashion: the binding energy and the number of Lac repressors. In Fig. 1*B* we show the predicted fold change as a function of these two experimentally accessible parameters. Alternatively, the binding of Lac repressor can be described by a dissociation constant, the concentration of Lac repressor for which the fold change in gene expression is 1/2. This approach is explained in *SI Text*. Throughout the text we report both binding energies and approximate dissociation constants, although all of our measurements and analysis are built around binding energies and repressor numbers. Approximate concentrations and dissociation constants are provided merely as rough estimates for the purposes of comparison with literature values in which sometimes these quantities are favored. For details of the estimation of the concentrations and dissociation constants, see *SI Text*, *Connecting* $\Delta\varepsilon_{rd}$ *to* K_d.

Earlier hints as to how simple repression plays out quantitatively were offered by Oehler et al. (32, 33) who measured the fold change in gene expression for constructs bearing each one of the four operators and for two different numbers of Lac repressor per cell. Using Eq. **5** or equivalent expressions (15, 36), the binding energy of Lac repressor to each one of the operators can be estimated. It must be noted, though, that these original measurements were not performed with the intention of the kind of quantitative dissection advocated here and that therefore the uncertainties in the parameters are substantial. In *SI Text*, Figs. S3 and S4, and Table S1, we give a detailed analysis of the extent to which these earlier results are consistent with our own data. Additionally, in *SI Text* and Fig. S5 we show that our results do not depend on the particular choice of quantification protocol for our enzymatic reporter. Finally, as shown in *SI Text* and Table S2, even if we replace our reporter with a fluorescent protein the results are essentially unaltered.

For the measurements reported here, we created ∼30 strains of bacteria where we systematically tuned the number of repressors, using a recently developed scheme for controlling ribosomal binding strength (37). Although this scheme provides a rough expectation for the number of repressors in each one of those strains, we had no precise or accurate a priori knowledge of the actual intracellular numbers of Lac repressors. These strains bear reporter constructs regulated by simple repression such as those shown in Fig. S1, for which we measure the fold change in gene expression. If we are to believe the input–output function from Eq. **5**, once we know the binding energy of the operator in question there is a direct and unequivocal relation between the fold change in gene expression and the number of repressor molecules. Testing these predictions requires an accurate and precise quantification of the absolute levels of repressor inside the cell. In fact, we view this approach as a way to count molecules by inference by looking at levels of gene expression and passing these levels of expression through the theoretical filter of Eq. **5**.

In the following sections we test these parameter-free predictions over a wide range of both expression and repressor numbers and show that they largely jibe with our experimental observations. The logic advocated here is that if Eq. **5** is shown to be predictive, it will open the door to creating synthetic gene regulatory circuits whose level of gene expression can be precisely tuned a priori and to being able to predict the regulation of a particular promoter by just looking at its regulatory sequence. In addition, a predictive understanding of the input–output relation of these architectures will serve as a jumping-off point for the design and understanding of more complex circuits such as those involving DNA looping, cooperative repression, etc. (15).

Results

Eq. **5** represents a provocatively simple expression purporting to describe the response of a bacterial cell to a wide variety of perturbations such as altering the DNA target sites (with the K_ds changing by three orders of magnitude or, equivalently, $\Delta\varepsilon_{rd}$ changing by 7 $K_B T$) (15, 36, 38, 39) and repressor copy numbers (with the copy numbers changing by several orders of magnitude). If we take this equation seriously, it implies that once we have determined the parameter $\Delta\varepsilon_{rd}$, there is a quantitative relation between the fold change in gene expression and the corresponding number of Lac repressors. Namely, once we know one quantity we can *predict* the other.

To exploit Eq. **5** we designed *lac*UV5 promoters with a single binding site for Lac repressor at the wild-type position of O1. These promoters bore Oid, O1, O2, or O3 and controlled the expression of the enzymatic reporter gene *lacZ* (*Materials and Methods* and Fig. S1), although as reported in *SI Text*, we also examined many of the same constructs using fluorescence as well, resulting in nearly identical results. We integrated each one of these simple repression constructs such as the one shown in in Fig. 1*A* in the chromosome of a strain bearing no Lac repressor and in six different strains that we systematically designed to express different constitutive levels of Lac repressor. As mentioned above, although we had a qualitative expectation about the number of Lac repressors present in each strain, we had no previous quantitative information about that magnitude.

Taking the Repressor Census Through Thermodynamic Models. We measured the fold change in gene expression of our simple repression constructs bearing the operators Oid, O1, O2, or O3 in the six different strain backgrounds we created. There are several different ways to explore the results in conjunction with Eq. **5**. As noted above, one scheme is to determine the absolute number of repressor molecules within one strain and to combine this with the measured fold change to obtain the in vivo binding energy for each of the different operators through Eq. **5**. With these binding energies in hand, a way to put the predictability of the thermodynamic model on the stand is to predict the number of repressors in the other strains. An alternative concept is simply to use all of the fold-change and repressor count measurements and to see how well they agree with the functional form provided by Eq. **5** by making one global fit to the in vivo binding energy for each operator.

Regardless of the scheme chosen it is necessary to possess an absolute count of the repressor number in each one of our strain backgrounds. Details of this determination are given below. To carry out the first scheme presented above we used strain RBS1027 as the basis of the calculation of the binding energies. The resulting fold change in gene expression for each operator in this strain background and the calculated binding energies are shown in Fig. 2*A*. Using these binding energies we plot the fold change in gene expression as a function of binding energy for all strains and choices of operators in Fig. 2*B* (the corresponding absolute values of gene expression measured for each strain are shown in Fig. S6). The data in Fig. 2*B* are fitted to Eq. **5** to generate a prediction for the number of repressors within each one of the five remaining strains. These predictions are shown in Fig. 2*C*.

Because the majority of our strains were created for this particular work, the resulting predicted cellular numbers cannot be compared with any external standard. However, strain HG104 expresses wild-type levels of repressor from the native *lacI* gene. Indeed, for this strain we predict 8.8 ± 0.8 repressor tetramers per cell, comparable to the previous and, to our knowledge, only available absolute measurement (40).

To bring the predictions of the model for simple repression to fruition we need to directly measure the number of repressors in each one of our six strains. We measured the in vivo number of Lac repressors in these six strains by performing quantitative immunoblots (19, 41, 42) from cell lysates such as those shown in Fig. 3*A*. To get an absolute count of the amount of Lac repressor in each strain a series of dilutions of a purified Lac repressor standard of a known concentration was used (Fig. 3*B*). Quantification of the luminescence of the immunoblots was performed using a cooled CCD camera. Care was taken to account for spatial nonuniformities in the light collection as depicted in Fig. 3*C*. We can reliably detect a wide range of purified Lac repressor stand-
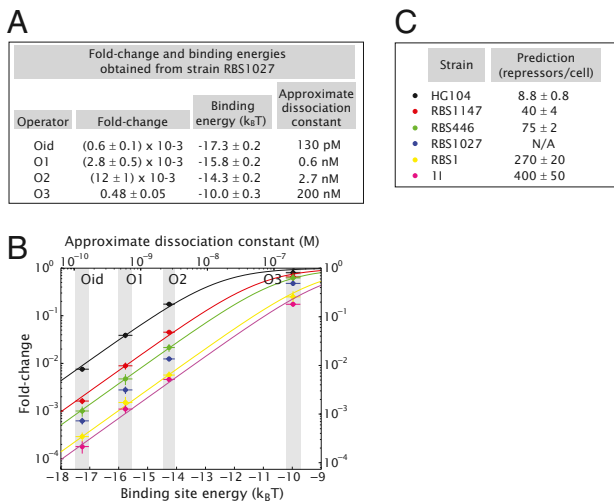
Fig. 2. Single-site binding energies and prediction of the number of repressors for different strains. (A) The operator binding energies and approximate dissociation constants are deduced from the measurement of the fold change for the different operators in strain RBS1027 combined with our knowledge of its intracellular number of repressors, using Eq. 5. (B) The fold change in gene expression is measured for all four operators in six different strain backgrounds (including RBS1027). Using the binding energies from A, we fit the data to Eq. 5 to make a parameter-free prediction of the number of repressors present in each strain shown in C. Errors in the predictions represent the SE of the corresponding fit. The errors in the binding energies are here denoted as gray shaded regions. Estimated dissociation constants are shown for convenience for comparison with literature values. The basis for these estimates is explained in *SI Text*.

ards using our immunoblots (as low as 50 pg, corresponding to around five repressors per cell). This result increases our confidence in the method as a way of precisely quantifying protein counts in bulk even at very low levels (*Materials and Methods* and Fig. 3D). It is important to note, however, that counting methods based on purification, such as immunoblots, have the inherent caveat that some proteins might have stayed behind in the different fractions. Although we took action to reduce this effect, the

results from immunoblots should be viewed as a lower bound on the actual number of proteins in vivo.

Our predictions for the number of Lac repressors in each strain can now be compared with the direct measurements of this quantity, which are shown in Fig. 4A. In Fig. 4B we compare the predictions and direct measurements explicitly. The direct measurements are comparable to the predictions within experimental error, giving us confidence that the proposed input–output function from Eq. 5 appropriately describes the input–output properties of the simple repression regulatory motif. This result suggests in turn that once we know the binding energy for an operator, we have predictive power. Although this analysis yielded results that are largely consistent between theory and experiment, it appears that we systematically underestimate the number of repressors in the two strains with the highest repressor number. The reader is referred to *SI Text* for a further discussion of these two strains.

Direct Determination of the in Vivo Lac Repressor Binding Energies. The scheme for exploring the limits and validity of the thermodynamic model advocated in the previous section is based on using one strain to determine the binding energy of Lac repressor to its operator DNA. However, as noted earlier, an alternative approach is to simply use the entirety of our data to evaluate global fits of Eq. 5 to the data corresponding to a given operator. Implementation of this concept is shown in Fig. S7B, where we combine all of our measurements to determine the best values of the different in vivo binding energies. On the other hand, one might choose to use the information about fold change and repressor copy number for one particular strain to derive the different binding energies. This analysis can be done, in turn, for all strains created for this work in an analogous way to what we did with strain RBS1027 in the previous section. In Fig. 5 we compare such fits with the binding energies that can be obtained from analyzing a single strain. Additionally, we show the energies obtained from the Oehler et al. data (33) (*SI Text* and Fig. S8) and from Fig. S7B for comparison. These multiple approaches for obtaining the binding energies, all leading to essentially comparable results (for example, Fig. S7A), increase our confidence in the simple model of Eq. 5 and in the minimalist modeling philosophy used to obtain it as a quantitative and predictive tool.

Finally, it is common in the theoretical treatment of experiments on transcriptional regulation to include a constant level of expression dubbed the "leakiness". Such leakiness is usually un-
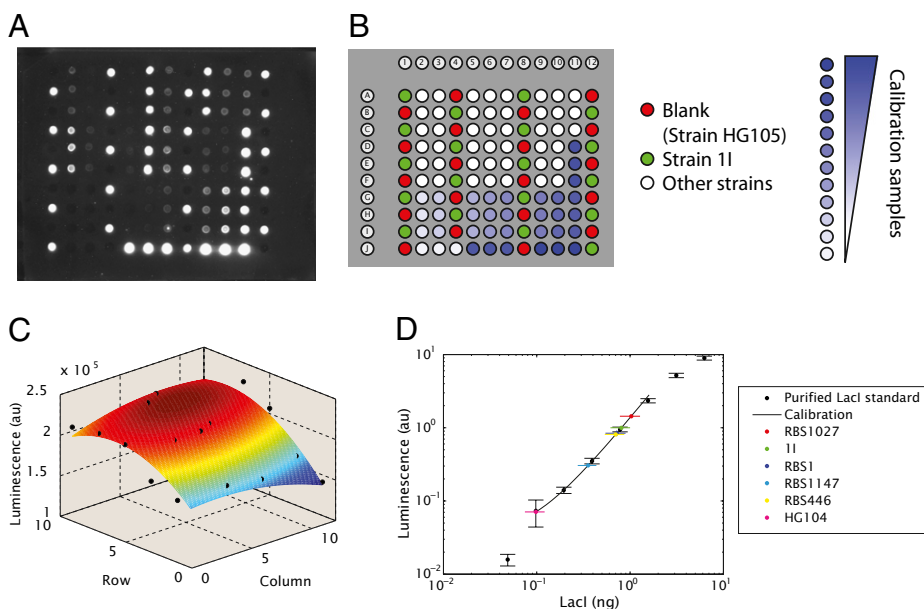

Fig. 3. Immunoblots for the measurement of the in vivo number of Lac repressors. (A) Typical luminescence image obtained from an immunoblot. (B) Map of the samples loaded on the membrane shown in A. The blank (HG105) and 1I samples are used to create a normalization map by subtracting the blank luminescence from all samples and dividing by 1I. White spots correspond to the cell lysates measured and the blue spots correspond to the different concentrations of purified Lac repressor standard. (C) Normalization map generated by fitting a 2D polynomial to 1I samples scattered around the membrane (black dots) after removing the blank. This map was used to account for nonuniformities in the collection of luminescence from the membrane. (D) Luminescence vs. quantity of LacI loaded. The calibration samples are used to construct a power law fit. The luminescence of the measured samples is shown as well. The unknown amounts of repressor loaded are determined by using the calibration curve. Samples 1I and RBS1 have been diluted 1:8 to match them to the dynamic range of the assay and therefore appear to have less signal within a spot (*SI Text*).
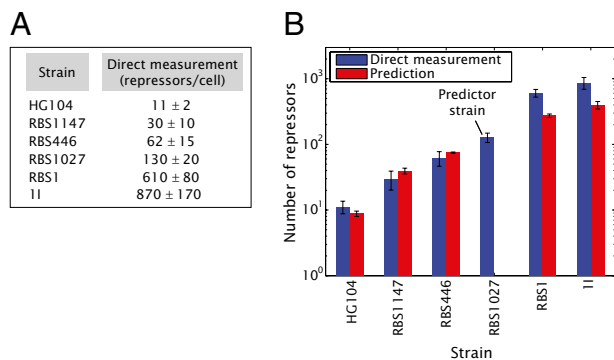
Fig. 4. Experimental and theoretical characterization of repressor copy number. (*A*) Immunoblots were used to measure the number of Lac repressors in six strains with different constitutive levels of Lac repressor. Each value corresponds to an average of cultures grown on at least 3 different days. The error bars are the SD of these measurements. (*B*) The fold-change measurements in Fig. 2*B* were combined with the binding energies obtained from Fig. 2*A* (derived from strain RBS1027) to predict the number of Lac repressors per cell in each one of the six strains used in this work. These predictions were examined experimentally by counting the number of Lac repressors using quantitative immunoblots.

Fig. 5. Determination of the in vivo binding energies. For each strain we combine the measurements of the fold change in gene expression with the corresponding number of repressors and solve Eq. **5** to obtain an estimate of the binding energies (dots). The energies obtained from the Oehler et al. data (33) are also shown. The lines correspond to using all measurements of the fold change in gene expression with their corresponding repressor numbers to fit Eq. **5** to obtain the best possible estimate for the binding energies. This fit is shown in Fig. S7*B*. The results of this approach are shown as horizontal lines and the shaded region captures the uncertainty.

derstood as a low level of activity that is independent of any regulation. The reader is referred to *SI Text* and Fig. S9 for a more detailed description of leakiness where we show that the values obtained for the binding energies do not change significantly for reasonable values of the leakiness.

Discussion

Theoretical models of gene expression, especially in bacteria, have reached a very high level of sophistication. Similarly, measurements of gene expression have come to the point where they are both reproducible and quantitative enough to serve as the basis for explicit attempts at confronting theory and experiment and to explore the merits of these theoretical perspectives as a conceptual framework for describing regulatory response. Indeed, such measurements have now reached the point where in our view it is no longer appropriate to use just words to describe them—they call for a theoretical response that is commensurate with the level of quantitative detail in the experiments themselves. To that end, we have undertaken a detailed study of one of the most important and fundamental regulatory building blocks found in living organisms from all three domains of life, namely, simple repression. Simple repression and its positive regulation counterpart, namely simple activation, serve as the paradigmatic building blocks of the much richer regulatory strategies that are used in the growing list of both natural and synthetic networks now being explored.

In recent years, the governing equations characterizing the transcriptional response of these elementary regulatory building blocks and much more complicated assemblies of them have been worked out in detail using the ideas of statistical mechanics. The work described here provides a template for the kind of rich interplay between theory and experiment that should be demanded of these other networks as well. In particular, the governing equations describing regulatory architectures feature certain key tuning variables that serve to elicit different biological responses. In the experiments described here, we have explored two of the elementary tuning parameters that govern the simple repression motif, namely, the strength of the transcription factor binding sites and the molecular counts of the repressors themselves. We have shown that an input–output function for simple repression obtained from thermodynamic assumptions, which depends on those two tuning parameters, can indeed predict in a parameter-free manner the regulatory outcome over roughly four orders of magnitude in the transcriptional output.

Using the thermodynamic model approach coupled tightly with precise measurements we have been able to perform a systematic
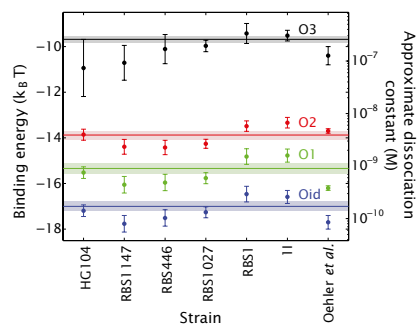
quantitative dissection of the input–output relation for simple repression and believe that similar analyses should be carried out for each of the other governing equations describing key regulatory motifs. As a by-product of these measurements, we have been able to make a precise determination of the in vivo binding energies for DNA–repressor interactions. In addition, these results provide a census of the repressor content for Lac repressor in *E. coli* over a large dynamic range (roughly two orders of magnitude in repressor counts). The predictive power revealed by this model on the basis of a few parameters is one of the first steps toward having a standardized description of a regulatory architecture on the basis of its microscopic parameters (1, 2). Harkening back to the electronic circuit analogy, the results presented here are analogous to illustrating that for a resistor there is a value for the resistance that is necessary and sufficient to predict the current given the voltage. In our case specification of the binding energy $\Delta\varepsilon_{rd}$ is necessary and sufficient to predict the fold change in gene expression given the number of repressors.

Further characterization of this architecture should explore the role of promoter copy number and operator position as these architectural features are known to alter the expression profile as well (43–45). In addition, with these insights in hand for the case of simple repression in the *lac* operon, it is now important to examine a suite of similar architectures in *E. coli* and other bacteria with the idea being to explore the extent to which the successes found in this case can be expected to apply to other genes.

Materials and Methods

DNA Constructs and Strains. The construction of all plasmids and strains is described in detail in *SI Text*.

In short, plasmids pZS25O1+11, pZS25O2+11, pZS25O3+11, and pZS25Oid +11 have a *lac*UV5 promoter controlling the expression of a LacZ reporter as shown schematically in Fig. S1.

Plasmid pZS3*1-lacI expresses Lac repressor off of a $p_{LtetO-1}$ promoter (46). The ribosomal binding site of this construct was weakened following ref. 37, using site-directed mutagenesis (Quikchange II; Stratagene) in order to generate constructs expressing LacI at different levels as described in *SI Text* and Table S3.

The *E. coli* strains used in this experiment are shown in Table S4. HG105 is wild-type *E. coli* (MG1655) with a complete deletion of the **lacIZYA** genes. HG104 is also wild-type *E. coli* with a deletion of the **lacZYA** genes. We therefore expect strain HG104 to express wild-type levels of Lac repressor.

Reporter constructs and Lac repressor constructs were integrated into the **galK** and **ybcN** regions, respectively, using recombineering (47) and combined using P1 transduction. Please refer to *SI Text* for details.

Growth Conditions and Gene Expression Measurements. Strains to be assayed were grown in M9 minimal medium plus 0.5% glucose and harvested during

exponential growth. Our protocol for measuring LacZ activity is basically a slightly modified version of the one described in refs. 48 and 49. Details are given in *SI Text*.

Measuring in Vivo Lac Repressor Number. Cell lysates of our different strains bearing Lac repressor were obtained as described in *SI Text*. Calibration samples using a known concentration of purified Lac repressor (courtesy of Stephanie Johnson, California Institute of Technology, Pasadena, CA) diluted in a lysate of HG105 strain (strain without Lac repressor) were used.

A nitrocellulose membrane was prepared for sample loading and afterward blocked and treated with anti-LacI primary monoclonal antibody and HRP-linked secondary antibody as discussed in *SI Text*. Two microliters of each sample was spotted on the membrane in a pattern similar to that of a 96-well plate. The resulting drops had a typical size of 3 mm. All samples were loaded in triplicate with the exception of samples 1I and HG105. Both of them were loaded on the order of 20 times on different positions of the membrane to obtain a spatial standard that would allow for corrections of nonuniformities in the light collection (see below).

The membrane was dried and developed with Thermo Scientific Super-Signal West Femto Substrate and imaged in a Bio-Rad VersaDoc 3000 system with an exposure of 5 min. A typical raw image of one of the membranes is shown in Fig. 3*A* and the corresponding loading map can be seen in Fig. 3*B*. Custom Matlab code was written to detect the spots and calculate their total luminescence. The luminescence coming from the HG105 blank samples was

fitted to a second-degree polynomial, which was in turn subtracted from all other luminescence values. After this procedure another second-degree polynomial was fitted to the 1I samples, resulting in a typical surface such as the one shown in Fig. 3*C*. Note that differences of up to 25% could be observed between different positions on the membrane. This last polynomial was used to normalize the intensity of all other samples.

The luminescence corresponding to the calibration samples was overlaid with the luminescence from the strains. The calibration samples were fitted to a power law using only the calibration data points in the range of the samples that were to be measured. An example of this calibration is shown in Fig. 3*C*. For additional details please refer to *SI Text*.

Finally, the amount of Lac repressor found in a spot was related to the number of Lac repressor molecules per cell by calibration of the OD readings of the original cultures to cell density as described in *SI Text*.

1. Endy D (2005) Foundations for engineering biology. *Nature* 438:449–453.
2. Voigt CA (2006) Genetic parts to program bacteria. *Curr Opin Biotechnol* 17:548–557.
3. Ben-Tabou de-Leon S, Davidson EH (2007) Gene regulation: Gene control network in development. *Annu Rev Biophys Biomol Struct* 36:191.
4. Gregor T, Tank DW, Wieschaus EF, Bialek W (2007) Probing the limits to positional information. *Cell* 130:153–164.
5. Cox RS, 3rd, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. *Mol Syst Biol* 3:145.
6. Peter IS, Davidson EH (2009) Modularity and design principles in the sea urchin embryo gene regulatory network. *FEBS Lett* 583:3948–3958.
7. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–338.
8. Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 31:64–68.
9. Setty Y, Mayo AE, Surette MG, Alon U (2003) Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci USA* 100:7702–7707.
10. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.
11. Kuhlman T, Zhang Z, Saier MH, Jr., Hwa T (2007) Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc Natl Acad Sci USA* 104:6043–6048.
12. Kaplan S, Bren A, Zaslaver A, Dekel E, Alon U (2008) Diverse two-dimensional input functions control bacterial sugar genes. *Mol Cell* 29:786–792.
13. Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. *Proc Natl Acad Sci USA* 100:5136–5141.
14. Bintu L, et al. (2005) Transcriptional regulation by the numbers: Models. *Curr Opin Genet Dev* 15:116–124.
15. Bintu L, et al. (2005) Transcriptional regulation by the numbers: Applications. *Curr Opin Genet Dev* 15:125–135.
16. Gama-Castro S, et al. (2008) RegulonDB (version 6.0): Gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res* 36(Database issue):D120–D124.
17. Alon U (2007) *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall/CRC Mathematical and Computational Biology Series (Chapman & Hall/CRC, Boca Raton, FL).
18. Alberts B (2008) *Molecular Biology of the Cell* (Garland Science, New York), 5th Ed.
19. Wu JQ, Pollard TD (2005) Counting cytokinesis proteins globally and locally in fission yeast. *Science* 310:310–314.
20. Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
21. Takamori S, et al. (2006) Molecular anatomy of a trafficking organelle. *Cell* 127:831–846.
22. Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 25:117–124.
23. Taniguchi Y, et al. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329:533–538.
24. Malmström J, et al. (2009) Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* 460:762–765.
25. Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U (2006) Plasticity of the cis-regulatory input function of a gene. *PLoS Biol* 4:e45.
26. Guido NJ, et al. (2006) A bottom-up approach to gene regulation. *Nature* 439:856–860.
27. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
28. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* 123:1025–1036.
29. Zenklusen D, Larson DR, Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* 15:1263–1271.
30. Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440:358–362.
31. Choi PJ, Cai L, Frieda K, Xie XS (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322:442–446.
32. Oehler S, Eismann ER, Krämer H, Müller-Hill B (1990) The three operators of the *lac* operon cooperate in repression. *EMBO J* 9:973–979.
33. Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Müller-Hill B (1994) Quality and position of the three *lac* operators of *E. coli* define efficiency of repression. *EMBO J* 13:3348–3355.
34. Ackers GK, Johnson AD, Shea MA (1982) Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci USA* 79:1129–1133.
35. Han L, et al. (2011) Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS One* 4(5):e5621.
36. Vilar JM, Leibler S (2003) DNA looping and physical constraints on transcription regulation. *J Mol Biol* 331:981–989.
37. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27:946–950.
38. Winter RB, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli* repressor–operator interaction: Equilibrium measurements. *Biochemistry* 20:6948–6960.
39. Hsieh WT, Whitson PA, Matthews KS, Wells RD (1987) Influence of sequence and distance between two operators on interaction with the *lac* repressor. *J Biol Chem* 262:14583–14591.
40. Gilbert W, Müller-Hill B (1966) Isolation of the Lac repressor. *Proc Natl Acad Sci USA* 56:1891–1898.
41. Borggrefe T, Davis R, Bareket-Samish A, Kornberg RD (2001) Quantitation of the RNA polymerase II transcription machinery in yeast. *J Biol Chem* 276:47150–47153.
42. Kobiler O, et al. (2005) Quantitative kinetic analysis of the bacteriophage lambda genetic network. *Proc Natl Acad Sci USA* 102:4470–4475.
43. Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci USA* 85:8973–8977.
44. Elledge SJ, Davis RW (1989) Position and density effects on repression by stationary and mobile DNA-binding proteins. *Genes Dev* 3:185–197.
45. Ryu S, Fujita N, Ishihama A, Adhya S (1998) GalR-mediated repression and activation of hybrid *lac*UV5 promoter: Differential contacts with RNA polymerase. *Gene* 223:235–245.
46. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25:1203–1210.
47. Sharan SK, Thomason LC, Kuznetsov SG, Court DL (2009) Recombineering: A homologous recombination-based method of genetic engineering. *Nat Protoc* 4:206–223.
48. Miller JH (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
49. Becker NA, Kahn JD, Maher LJ, 3rd (2005) Bacterial repression loops require enhanced DNA flexibility. *J Mol Biol* 349:716–730.