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

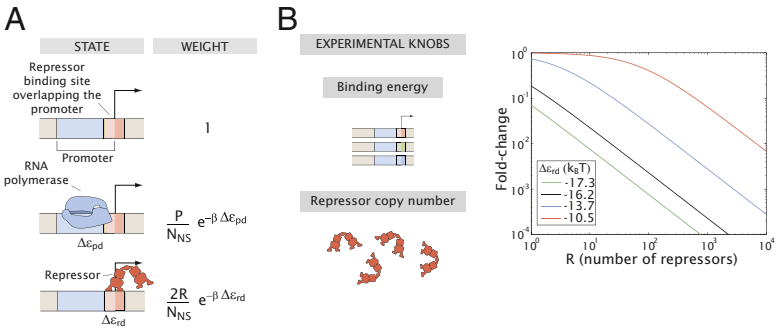

Fig. 1. The simple repression motif. (*A*) States and weights of the thermodynamic model describing this regulatory motif. We assume that Lac repressor sterically excludes RNA polymerase from the promoter, although that assumption is not critical to our analysis. *P* and *R* are the numbers of RNA polymerase 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

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

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

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

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,

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

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 $(P / N_{NS})e^{-\beta\Delta\varepsilon_{pd}} << 1$. This outcome results in the fold change collapsing to the simpler form,

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

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_BT) (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-

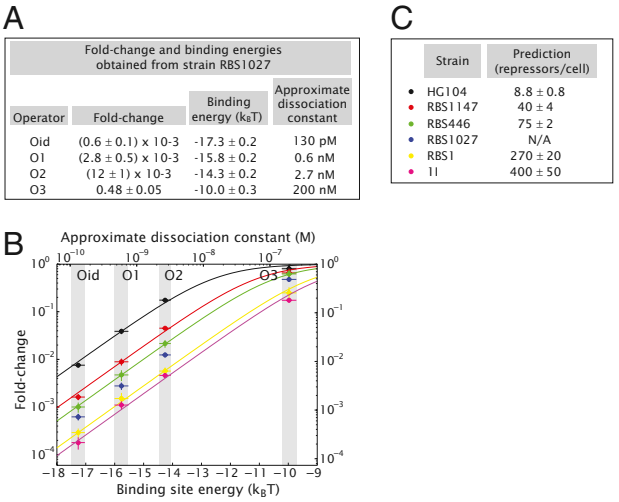

Fig. 2. Single-site binding energies and prediction of the number of repressors for different strains. (*A*) The operator binding energies and approximate dissociation constants are deduced from the measurement of the fold change for the different operators in strain RBS1027 combined with our knowledge 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. 3*D*). 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. 4*A*. In Fig. 4*B* 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. S7*B*, 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. S7*B* for comparison. These multiple approaches for obtaining the binding energies, all leading to essentially comparable results (for example, Fig. S7*A*), 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-

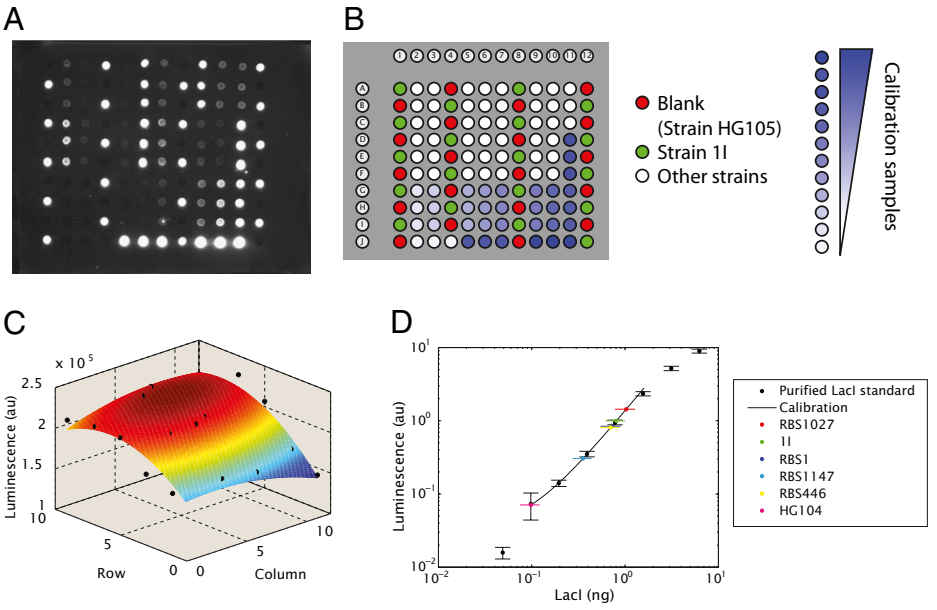

Fig. 3. Immunoblots for the measurement of the in vivo number of Lac repressors. (*A*) Typical luminescence image obtained from an immunoblot. (*B*) Map of the samples loaded on the membrane shown in *A*. The blank (HG105) and 1I samples are used to create a normalization map by subtracting the blank 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*).

A

Strain	Direct measurement (repressors/cell)
HG104	11 ± 2
RBS1147	30 ± 10
RBS446	62 ± 15
RBS1027	130 ± 20
RBS1	610 ± 80
11	870 ± 170

B

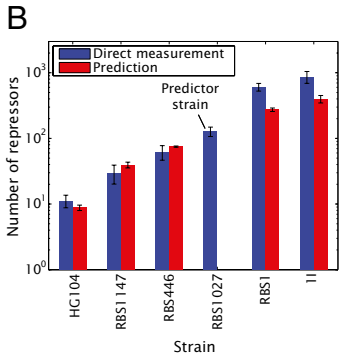

Fig. 4. Experimental and theoretical characterization of repressor copy number. (*A*) Immunoblots were used to measure the number of Lac repressors in six strains with different constitutive levels of Lac repressor. Each value corresponds to an average of cultures grown on at least 3 different days. 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.

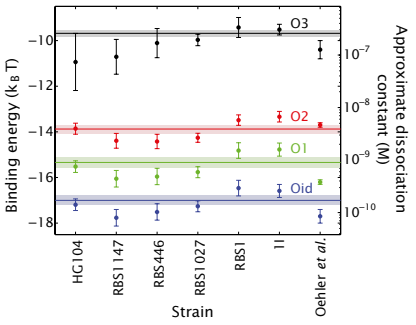

Fig. 5. Determination of the in vivo binding energies. For each strain we combine the measurements of the fold change in gene expression with the corresponding number of repressors and solve Eq. **5** to obtain an estimate of the binding energies (dots). The energies obtained from the Oehler et al. data (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

SYSTEMS BIOLOGY

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*.

ACKNOWLEDGMENTS. We thank Rob Brewster, Stephanie Johnson, Jane Kondev, Tom Kuhlman, Kathy Matthews, Ron Milo, Alvaro Sanchez, Paul Wiggins, and Bob Schleif for enlightening discussions over the course of many years and comments on the manuscript, and to Thomas Gregor and Ted Cox for lending their respective laboratory spaces for further experiments. We thank Franz Weinert, James Boedicker, Heun Jin Lee, and Maja Bialecka for help with the cell counts calibration. We thank the National Institutes of Health for support through Grant DP1 OD000217 (Director's Pioneer Award) and Grant R01 GM085286, and La Fondation Pierre Gilles de Gennes (R.P.).

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.

Supporting Information

Garcia and Phillips 10.1073/pnas.1015616108

SI Text

S1. Theoretical Background. In the following sections we explore the theoretical background leading to the different predictions explored in the main text. We start by introducing thermodynamic models in general and arrive at an expression for the fold change in gene expression due to repression by Lac repressor.

S1.1. "Thermodynamic models" of transcriptional regulation. Thermodynamic models of transcriptional regulation are based on computing the probability of finding RNA polymerase (RNAP) bound to the promoter and how the presence of transcription factors (TFs) modulates this probability. These models and their application to bacteria are reviewed in (1, 2).

These models make two key assumptions. First, the models assume that the processes leading to transcription initiation by 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 TFs to DNA. Second, they assume that the level of gene expression of a gene is proportional to the probability of finding RNAP bound to the corresponding promoter.

We start by analyzing the probability that RNAP will be bound at the promoter of interest in the absence of any transcription factors. We assume that the key molecular players (RNAP and TFs) are bound to the DNA either specifically or nonspecifically. In particular, this question has been addressed experimentally in the context of RNAP (3) and the Lac repressor (4, 5), our two main molecules of interest in this paper. The reservoir for RNAP is therefore the background of nonspecific sites. To determine the contribution of this reservoir we sum over the Boltzmann weights of all of the possible configurations. For P RNAP molecules inside the cell with N_{NS} nonspecific DNA sites we get

$$Z^{\mathrm{NS}}(P;N_{\mathrm{NS}}) = \frac{N_{\mathrm{NS}}!}{P!(N_{\mathrm{NS}}-P)!} e^{-\beta\varepsilon_{pd}^{\mathrm{NS}}} \simeq \frac{(N_{\mathrm{NS}})^P}{P!} e^{-\beta\varepsilon_{pd}^{\mathrm{NS}}}, \qquad \text{[S1]}$$

where $\beta = 1/K_{\mathrm{B}}T$. The first factor in the first expression accounts for all of the possible configurations of RNAP on the reservoir. Examples of such configurations are shown diagrammatically in Fig. S2*A*. The second factor assigns the energy of binding between RNAP and nonspecific DNA, $\varepsilon_{pd}^{\mathrm{NS}}$ (the subscript *pd* stands for RNA polymerase–DNA interaction), which, as a theoretical convenience that may have to be revised in quantitatively dissecting real promoters, is taken to be the same for all nonspecific sites. A more sophisticated treatment of this model to account for the differences in the nonspecific binding energy has been addressed by ref. 6. Finally, the last expression corresponds to assuming that $N_{\mathrm{NS}} \gg P$, a reasonable assumption given that the *E. coli* genome is ~5 Mbp long and that the number of σ^{70} RNAP molecules, the type of RNAP we are interested in for the purposes of this paper, is on the order of 1,000 (7).

We calculate the probability of finding one RNAP bound to a promoter of interest in the presence of this nonspecific reservoir. Two states are considered: Either the promoter is empty and P RNAPs are in the reservoir or the promoter is occupied leaving $P-1$ RNAP molecules in the reservoir. The corresponding total partition function is

$$Z(P;N_{\mathrm{NS}}) = \underbrace{Z^{\mathrm{NS}}(P;N_{\mathrm{NS}})}_{\text{Promoter unoccupied}} + \underbrace{e^{-\beta\varepsilon_{pd}^{\mathrm{S}}} Z^{\mathrm{NS}}(P-1;N_{\mathrm{NS}})}_{\text{Promoter occupied}}, \qquad \text{[S2]}$$

where, in analogy to the nonspecific binding energy, we have defined $\varepsilon_{pd}^{\mathrm{S}}$ as the binding energy between RNAP and the promoter. Our strategy in these calculations is to write the total partition function as a sum over two sets of states, each of which has its own partial partition function. The probability of finding the promoter occupied, p_{bound} is then

$$p_{\mathrm{bound}}(P) = \frac{e^{-\beta\varepsilon_{pd}^{\mathrm{S}}} Z^{\mathrm{NS}}(P-1;N_{\mathrm{NS}})}{Z^{\mathrm{NS}}(P;N_{\mathrm{NS}}) + e^{-\beta\varepsilon_{pd}^{\mathrm{S}}} Z^{\mathrm{NS}}(P-1;N_{\mathrm{NS}})} = \frac{1}{1+\frac{N_{\mathrm{NS}}}{P} e^{\beta\Delta\varepsilon_{pd}}}, \qquad \text{[S3]}$$

with $\Delta\varepsilon_{pd} = \varepsilon_{pd}^{\mathrm{S}} - \varepsilon_{pd}^{\mathrm{NS}}$, the difference in energy between being bound specifically and nonspecifically. With these results in hand we can now turn to regulation by Lac repressor.

S1.2. Simple repression by Lac repressor. In its simplest form, repression is carried out by a transcription factor that binds to a site overlapping the promoter. This binding causes the steric exclusion of RNAP from that region, decreasing the level of gene expression. Additionally, these transcription factors might be multimeric, resulting in the presence of two DNA binding heads on the protein and leading to DNA looping if extra binding sites are present. In the case of Lac repressor, for example, the protein is already in its multimeric form before binding to DNA (8).

We begin by analyzing the case of repressors that require binding only to a single site to repress expression for the case of a repressor with only one binding head. This case study will allow us to develop key concepts like the role of nonspecific binding, which will be useful when addressing the case of repression by Lac repressor tetramers.

S1.2.1. Repression by Lac repressor dimers. We use the simpler case of a repressor with just one binding head to build some key concepts. In analogy to section *S1.1* for the case of RNAP we consider Lac repressor to be always bound to DNA, either specifically or nonspecifically. This assumption is consistent with the available experimental data (5). Our aim is to examine all of the different configurations available to P RNA polymerase molecules, R LacI dimers, and N_{NS} nonspecific sites. If the binding energies of RNAP and the LacI head to nonspecific DNA are $\varepsilon_{pd}^{\mathrm{NS}}$ and $\varepsilon_{rd}^{\mathrm{NS}}$, respectively, the nonspecific partition function becomes

$$Z^{\mathrm{NS}}(P,R_2) = \underbrace{\frac{N_{\mathrm{NS}}^P}{P!} e^{-P\beta\varepsilon_{pd}^{\mathrm{NS}}}}_{Z^{\mathrm{NS}}(P)} \underbrace{\frac{N_{\mathrm{NS}}^{R_2}}{R_2!} e^{-R_2\beta\varepsilon_{rd}^{\mathrm{NS}}}}_{Z^{\mathrm{NS}}(R_2)}, \qquad \text{[S4]}$$

where we have assumed that both LacI dimers and RNAP are so diluted in the reservoir that they do not interact with each other and we use the notation R_2 with the subscript 2 as a reminder that we are considering the case of dimers.

Our model states that we can find three different situations when looking at the promoter: (*i*) both sites can be empty, (*ii*) one RNAP can be taken from the reservoir and placed on its site, and (*iii*) a LacI dimer can be taken from the reservoir and placed on the main operator. These states and their corresponding normalized weights, which we derive below, are shown in Fig. S2*B*. This model assumes that LacI sterically excludes RNA polymerase from the promoter, which is supported by the results from ref. 9. However, it can be easily modified to accommodate a state where both LacI and RNAP are bound simultaneously, for example.

The total partition function is

$$Z_{\text{total}}(P,R_2) = \underbrace{Z^{\text{NS}}(P,R_2)}_{\text{promoter free}} + \underbrace{Z^{\text{NS}}(P-1,R_2)e^{-\beta\varepsilon^{\text{S}}_{pd}}}_{\text{RNAP on promoter}} + \underbrace{Z^{\text{NS}}(P,R_2-1)e^{-\beta\varepsilon^{\text{S}}_{rd}}}_{\text{LacI dimer on operator}}, \quad \textbf{[S5]}$$

where $\varepsilon^{\text{S}}_{pd}$ and $\varepsilon^{\text{S}}_{rd}$ are the binding energies of RNAP and a Lac repressor head to their specific sites, respectively. We factor out the term corresponding to having all molecules in the reservoir and define $\Delta\varepsilon_{pd} = \varepsilon^{\text{s}}_{pd} - \varepsilon^{\text{NS}}_{pd}$ and $\Delta\varepsilon_{rd} = \varepsilon^{\text{s}}_{rd} - \varepsilon^{\text{NS}}_{rd}$ as the energy gain of RNAP and dimeric LacI when switching from a nonspecific site to their respective specific sites, respectively. The probability of finding RNAP bound to the promoter is given by

$$p_{\text{bound}} = \frac{\frac{P}{N_{\text{NS}}} e^{-\beta\Delta\varepsilon_{pd}}}{1 + \frac{P}{N_{\text{NS}}} e^{-\beta\Delta\varepsilon_{pd}} + \frac{R_2}{N_{\text{NS}}} e^{-\beta\Delta\varepsilon_{rd}}}. \quad \textbf{[S6]}$$

This expression can be rewritten as

$$p_{\text{bound}} = \frac{1}{1 + \frac{N_{\text{NS}}}{P \cdot F_{\text{reg}}(R_2)} e^{\beta\Delta\varepsilon_{pd}}}, \quad \textbf{[S7]}$$

where we have defined the regulation factor

$$F_{\text{reg}}(R_2) = \frac{1}{1 + \frac{R_2}{N_{\text{NS}}} e^{-\beta\Delta\varepsilon_{rd}}}. \quad \textbf{[S8]}$$

Note that in the absence of repressor ($R_2 = 0$), p_{bound} turns into Eq. **3**. The regulation factor can be seen as an effective rescaling of the number of RNAP molecules inside the cell (1) and, in the case of repression, it is just the probability of finding an empty operator in the absence of RNAP.

One of the key assumptions in the thermodynamic class of models is that the level of gene expression is linearly related to p_{bound}. This assumption allows us to equate the fold change in gene expression to the fold change in promoter occupancy:

$$\text{fold change}(R_2) = \frac{p_{\text{bound}}\,(R_2 \neq 0)}{p_{\text{bound}}\,(R_2 = 0)}. \quad \textbf{[S9]}$$

If we substitute p as shorthand for $\frac{P}{N_{\text{NS}}} e^{-\beta\Delta\varepsilon_{pd}}$ in the expression for p_{bound}, we find

$$\text{fold change}(R_2) = \frac{p+1}{p + \frac{1}{F_{\text{reg}}(R_2)}}. \quad \textbf{[S10]}$$

The fold change becomes independent of the details of the promoter in the case of a weak promoter, where $p \ll 1,\ 1/F_{\text{reg}}(R_2)$, which permits us to write the approximate expression

$$\text{fold change}(R_2) \simeq F_{\text{reg}}(R_2) = \left(1 + \frac{R_2}{N_{\text{NS}}} e^{-\beta\Delta\varepsilon_{rd}}\right)^{-1}. \quad \textbf{[S11]}$$

In the case of the *lac* promoter if one considers in vitro binding energies of RNAP to the promoter, p has the approximate value $\sim 10^{-3}$ (1). The case of the *lac*UV5 promoter used in this work is explored in section *S1.4*, where we show that although it is a stronger promoter than the wild-type *lac* promoter, p is still a small value. Repression always bears a regulation factor <1, suggesting that we can use the weak promoter approximation for the *lac*UV5 promoter.

In much the same way done in this work, Oehler et al. (10) created different constructs by varying the identity of the Lac repressor binding site. For each one of these constructs they measured the fold change in gene expression as a function of the concentration of LacI dimers inside the cell.

In Fig. S2*C* we present a fit of their measured fold change as a function of the number of Lac repressor molecules inside the cell. This fit is made by determining the parameters in Eq. **S11**. Note that for each construct there is only one unknown: the in vivo binding energies, $\Delta\varepsilon_{rd}$. The results are summarized in Table S1.

S1.2.2. The nonspecific reservoir for Lac repressor tetramers. We now consider the differences in the case where experiments are performed using tetramers rather than dimers (as in the present study). When dealing with Lac repressor tetramers only one head has to be bound to the DNA. In principle, it is not clear what the state of the other head will be. For example, that extra head could be "hanging" from the DNA without establishing contact with DNA. Another option is that the extra head will also be exploring different nonspecific sites. For the purposes of this section we assume that the second head can also bind to DNA.

Even though only one head bound to the operator is necessary for repression, we will see that it is important to account for the presence of the second head. In analogy to the dimer case, we assume that both Lac repressor binding heads are bound to DNA at all times, either specifically or nonspecifically. This choice is arbitrary and the final results do not depend on the particular model for the state of the second head. We work with this particular formulation of the problem because it is both concrete and analytically tractable and makes the counting of the accessible states more transparent.

The model for the nonspecific reservoir is depicted in Fig. S2*D*. For LacI dimers we assumed that the molecules were exploring all possible nonspecific sites. For the case of tetramers, in contrast, LacI will be exploring all possible DNA loops between two different nonspecific sites. We start by considering only one LacI molecule. We count the possible ways in which we can arrange the two heads on different nonspecific sites on the DNA. We label the site where one of the heads binds i and the other site j. For every choice of sites an energy $\varepsilon^{\text{NS}}_{rd}$ is gained for each head that is nonspecifically bound. A cost in the form of a looping free energy $F_{\text{loop}}(i,j)$ is also paid for bringing sites i and j together. The sum over all nonspecific states can be written as

$$Z^{\text{NS}}(R_4 = 1) = \frac{1}{2} \underbrace{\sum_{i=1}^{N_{\text{NS}}} e^{-\beta\varepsilon^{\text{NS}}_{rd}}}_{\text{head 1, site } i} \underbrace{\sum_{j=1}^{N_{\text{NS}}} e^{-\beta\varepsilon^{\text{NS}}_{rd}}}_{\text{head 2, site } j} \underbrace{e^{-\beta F_{\text{loop}}(i,j)}}_{\text{Looping between sites } i \text{ and } j}. \quad \textbf{[S12]}$$

Note that a factor of $\frac{1}{2}$ has been introduced not to overcount loops. This is equivalent to assuming that the two binding heads on a repressor are indistinguishable. Our model assumes that the binding of a tetramer head is independent of the state of the other head. Therefore, the interaction between a head and DNA is the same in the tetramer and the dimer case.

Because the bacterial genome is circular, we can choose a particular binding site for the first head, i_0, and sum over all possible positions for the second head. This analysis can now be done for the different N_{NS} positions that can be chosen for i_0, resulting in

$$Z^{\text{NS}}(R_4 = 1) \simeq \frac{1}{2} \underbrace{N_{\text{NS}}}_{\text{choices for } i_0} e^{-\beta 2\varepsilon^{\text{NS}}_{rd}} \sum_j e^{-\beta F_{\text{loop}}(i_0,j)}. \quad \textbf{[S13]}$$

Finally, we bury the term $\sum_j e^{-\beta F_{\text{loop}}(i_0,j)}$ into an effective nonspecific looping free energy $e^{-\beta F^{\text{NS}}_{\text{loop}}}$.

To obtain the partition function for R_4 tetramers (where now the subscript 4 is a reminder that the repressor is a tetramer) we assume that all repressors are independent and indistinguishable. We therefore extend the partition function to the case of R_4 noninteracting tetramers in the reservoir by computing

$$Z^{\text{NS}}(R_4) = \frac{\left[Z^{\text{NS}}(R_4 = 1)\right]^{R_4}}{R_4!} = \frac{1}{2^{R_4}} \frac{(N_{\text{NS}})^{R_4}}{R_4!} e^{-\beta R_4 2\varepsilon^{\text{NS}}_{rd}} e^{-\beta R_4 F^{\text{NS}}_{\text{loop}}}, \quad \textbf{[S14]}$$

where the binding energy is still defined as in section *S1.2.1*.

From this point on we consider only Lac repressor tetramers. As a result, for notational compactness we replace R_4 with R. We obtain the complete nonspecific partition function by multiplying the factor corresponding to repressors with a factor corresponding to RNAP being bound nonspecifically shown in Eq. **S4** resulting in

$$Z^{\text{NS}}(P,R) = \frac{(N_{\text{NS}})^P}{P!} e^{-\beta P \varepsilon^{\text{NS}}_{pd}} \frac{1}{2^R} \frac{(N_{\text{NS}})^R}{R!} e^{-\beta R 2\varepsilon^{\text{NS}}_{rd}} e^{-\beta R F^{\text{NS}}_{\text{loop}}}, \quad \textbf{[S15]}$$

which now allows us in the next section to address the case of repression by tetramers.

S1.2.3. Repression by Lac repressor tetramers. We begin by taking one head of one Lac repressor tetramer out of the nonspecific reservoir shown in Eq. **S14** and binding it specifically to the operator. This analysis can be easily done by going back to Eq. **S12**. We label the position on the genome corresponding to the specific site i_0. We choose only those terms in the summation corresponding to the binding site of interest. Because either one of the heads can reach the position labeled by i_0, we obtain the following partition function for a single tetramer bound to a specific site:

$$Z^{O,\text{NS}}_R = \frac{1}{2} e^{-\beta \varepsilon^{\text{S}}_{rd}} e^{-\beta \varepsilon^{\text{NS}}_{rd}} \left(\sum_{i=1}^{N_{\text{NS}}} e^{-F_{\text{loop}}(i,i_0)} + \sum_{j=1}^{N_{\text{NS}}} e^{-F_{\text{loop}}(i_0,j)} \right). \quad \textbf{[S16]}$$

Because both sums are identical, we can reduce this to

$$Z^{O,\text{NS}}_R = e^{-\beta \varepsilon^{\text{S}}_{rd}} e^{-\beta \varepsilon^{\text{NS}}_{rd}} \sum_{j=1}^{N_{\text{NS}}} e^{-F_{\text{loop}}(i_0,j)} = e^{-\beta \varepsilon^{\text{S}}_{rd}} e^{-\beta \varepsilon^{\text{NS}}_{rd}} e^{-\beta F^{\text{NS}}_{\text{loop}}}. \quad \textbf{[S17]}$$

We are now ready to calculate the total partition function. We consider the three states from Fig. 1*B*. The weights corresponding to the first two states will be the same as in the LacI dimer case. The third state corresponds to the partition function term we just calculated. The total partition function is then

$$Z_{\text{total}}(P,R) = Z^{\text{NS}}(P,R) + Z^{\text{NS}}(P-1,R) e^{-\beta \varepsilon^{\text{S}}_{pd}} + Z^{\text{NS}}(P,R-1) \times Z^{O,\text{NS}}_R. \quad \textbf{[S18]}$$

The last term corresponds to having $R - 1$ repressors in the reservoir and having one repressor with one head bound specifically. After rewriting these equations using Eq. **S17**, and using the weak promoter approximation, we get a fold change

$$\text{fold change}(R) \simeq \left(1 + 2 \frac{R}{N_{\text{NS}}} e^{-\beta \Delta \varepsilon_{rd}} \right)^{-1}. \quad \textbf{[S19]}$$

Even though the contribution from the nonspecific loops just vanished, we see that there is a factor of 2 difference in front of the number of LacI tetramers. This result is different from the fold change in gene expression for dimers shown in Eq. **S9**. It can be easily understood if we think about the actual number of binding heads that are now present. In the case of dimers we have R_2 binding heads whereas for tetramers there are $2R_4$ binding heads inside the cell. As a result, no information about the nonspecific looping background can be obtained by doing the experiment described in the main text. We see that as long as the number of binding heads is the same the fold change will not vary. Interestingly, this is one of the conclusions from the data by Oehler et al. (10). They compared repression for two different numbers of monomers of each kind of LacI, such that $2R_4 = R_2$. The fold change in gene expression obtained for each monomer concentration is comparable for dimers and tetramers as long as this condition is met. An alternative way to look at this is by comparing the binding energies obtained for dimers and tetramers. These two sets of energies, obtained from Eqs. **S11** and **S19**, are shown in Table S1.

S1.3. Connecting $\Delta\varepsilon_{rd}$ to K_d. We can also describe the fold change in perhaps the more familiar language of dissociation constants (2). We think of the two reactions shown in Fig. S2*E* where the DNA can be bound either by RNA polymerase or by Lac repressor. In steady state we can relate the concentrations of the different molecular players to the respective dissociation constants through

$$\frac{[P][D]}{[P-D]} = K_{\text{P}} \quad \textbf{[S20]}$$

and

$$\frac{[R][D]}{[R-D]} = K_{\text{d}}. \quad \textbf{[S21]}$$

In these equations we have defined $[P]$ and $[R]$ as the concentrations of RNA polymerase and Lac repressor that are not bound to the promoter, respectively. The concentrations of their respective protein DNA complexes are $[P - D]$ and $[R - D]$. $[D]$ is the concentration of free DNA. Finally, K_{P} and K_{d} are the dissociation constants for RNA polymerase and repressor, respectively.

We want to determine p_{bound}, the probability of finding the promoter occupied by RNA polymerase. This probability can also be expressed as the fraction of DNA molecules occupied by RNA polymerase and given by

$$p_{\text{bound}} = \frac{[P-D]}{[D] + [R-D] + [P-D]}. \quad \textbf{[S22]}$$

If we divide by $[D]$ and use Eqs. **S20** and **S21**, we arrive at

$$p_{\text{bound}} = \frac{[P]/K_{\text{P}}}{1 + [R]/K_{\text{d}} + [P]/K_{\text{P}}}. \quad \textbf{[S23]}$$

By comparing this expression to, for example, Eq. **3** we can relate the repressor binding energy $\Delta\varepsilon_{rd}$ to the tetramer dissociation constant through

$$\frac{[R]}{K_{\text{d}}} = \frac{2R}{N_{\text{NS}}} e^{-\beta \Delta \varepsilon_{rd}}, \quad \textbf{[S24]}$$

where we have assumed that the concentration of free repressor, $[R]$, is approximately equal to the total concentration of repressor in the cell. Throughout the text we express the binding energies also in the language of approximate dissociation constants. To do this we assume an estimated *E. coli* volume of 1 fL such that a repressor per cell corresponds to a concentration of 1.7 nM. It is important to note, however, that there are many subtleties involved in the correct determination of the cytoplasmic volume of

E. coli. As a result we view all of the dissociation constants reported in this work as approximate values suitable only for the purposes of making order of magnitude comparisons to other literature values that often use that language for describing in vitro experiments.

***S1.4. Weak promoter approximation for the lac*UV5 promoter.** A key assumption leading to the simple expression for the fold change in gene expression from Eq. **5** is that the weight corresponding to RNA polymerase being bound to the promoter is much smaller than 1, meaning that the promoter will be unoccupied. Mathematically, we express this as $P/N_{NS}e^{-\beta\Delta\varepsilon_{pd}} \ll 1$. Following ref. 1 we can write the binding energy as

$$\Delta\varepsilon_{pd} = \varepsilon^{S}_{pd} - \varepsilon^{NS}_{pd} = \frac{K^{S}_{d}}{K^{NS}_{d}}, \qquad \text{[S25]}$$

where K^{S}_{d} and K^{NS}_{d} are the dissociation constants of RNA polymerase to specific and nonspecific DNA, respectively. In vitro values for the nonspecific dissociation constant are $K^{NS}_{d} \approx 10{,}000$ nM (11), whereas the specific dissociation constant for the *lac*UV5 promoter $K_{d}{}^{S}$ has been measured to be between 6 nM (12) and 80 nM (13). This result corresponds to a binding energy range between −4.8 and −7.4 k_BT. In exponentially growing *E. coli* there are ∼500 σ^{70} RNA polymerase molecules available (7). This polymerase count results in a range for the factor $(P/N_{NS})e^{-\beta\Delta\varepsilon_{pd}}$ of 0.01–0.16. Therefore, we conclude that not neglecting the term corresponding to RNA polymerase binding to the promoter from our expression for the fold change would result in only a small correction at the most.

S2. Predictions Generated by Our Analysis of the Oehler et al. Data. Oehler et al. (10) measured the fold change in gene expression for all four operators considered in our experiments in two different strain backgrounds expressing different numbers of repressor molecules. In sections *S1.2.1* and *S1.2.3* we showed how through Eq. **5** we can obtain in vivo binding energies for each of those four operators by exploiting measured fold changes. The energies resulting from this procedure for the data of Oehler et al. (10) are shown in Table S1.

It is interesting to ask to what extent the binding energies derived from these earlier measurements can be used to make "predictions" about our own strains. That is, despite the dearth of quantitative information in these earlier measurements, as noted above, they still provide enough hints to actually extract estimated binding energies that can then be used in conjunction with measured fold changes to estimate the number of repressors in the strain of interest.

In Fig. S3*B* we show the fits of our model to the fold-change data assuming the energies obtained from the Oehler et al. data. The resulting predictions are shown in Fig. S3*C*. These predictions can be now put to test by contrasting them with the direct measurements of the absolute number of repressors in each of our strain backgrounds. These direct measurements are shown once again in Fig. S4*A* and their comparison with the predictions is presented in Fig. S4*B*. As can be seen from Fig. S4, even the case in which we use binding energies obtained from data stemming from an independent experiment yields surprisingly reasonable predictions for the number of repressors harbored in our strains.

S3. Global Fit to All Our Data and Sensitivity of the Predictions. One of the approaches followed in this work was to use the data on fold change and absolute number of repressors for one strain (RBS1027) to obtain the binding energies. These energies were in turn used to generate predictions. This analysis was done because we intended to test the predictions generated by the thermodynamic model. A legitimate alternative is to combine all of our available data for the fold change in gene expression with the corresponding data on the number of Lac repressors in each strain to obtain the best possible estimate for the Lac repressor binding energies. The corresponding fit and resulting energies are shown in Fig. S7*B*.

To get a better sense of how well this fit constrains the values of the binding energies we wished to analyze the "sensitivity" of the fit. To do this we plotted the data corresponding to the binding site O1 and overlaid it with curves for the fold change in gene expression where we have chosen different values for the binding energy. In Fig. S8 we show the data for the O1 binding site together with its best fit and several other curves with different choices of the binding energy. It is clear from Fig. S8 that the fit is constraining the value of the binding energy relatively well (within $<1k_BT$) and that the error in the parameter resulting from the fit captures this.

S4. Repression for Strains RBS1 and 1I. In the main text we hint multiple times at a slight discrepancy between our theoretical predictions and the results measured for the fold change in strains RBS1 and 1I. We do not believe that this discrepancy is due to a problem with the determination of the concentration of Lac repressor because we were able to reliably detect higher and lower concentrations of the purified standard than those corresponding to these two strains. Another alternative is that we did not quantify the level of gene expression correctly. Indeed, the measurements for Oid correspond to the lowest levels of gene expression quantified in this work. For example, could there be some constant transcription level or "leakiness" that cannot be repressed by Lac repressor? However, the shift is also present in the other operators where the levels of gene expression are such that a constant leakiness would have a negligible effect. Additionally, the measurements of these two strains for all other operators are well between the range of the rest of the data which shows no such systematic shift. We are then forced to conclude that the discrepancy, if real and not just an unfortunate experimental systematic error unaccounted for, is due to the fact that these strains have a much higher level of Lac repressor. This line of logic would lead us to conclude that affinity of Lac repressors to DNA can somehow be affected if its intracellular number is too high. However, further experimentation will be necessary to confirm this assertion.

S5. Accounting for Leakiness. One interesting property of Eq. **5** is that it predicts that the fold change in gene expression will go down indefinitely as the number of repressors is increased. However, at some point one would expect to have some constant level that is, in principle, independent of any regulation. This is called leakiness and is usually attributed to transcription that is independent of the promoter of interest. Such nondesired transcription could stem, for example, from RNA polymerase escaping from a nearby promoter and generating a transcript.

We wish to determine whether our results are being contaminated by such leakiness and, if so, what its effect on the estimation of the binding energies would be. The smallest absolute value of LacZ activity detected in our strains corresponds to binding site Oid in strain 1I. This combination has an activity of ∼1 Miller unit (MU). This activity level sets a bound on the maximum value of the leakiness: Because we can measure activities down to 1 MU, the leakiness cannot be any higher than that and, in the worst possible case, it would be equal to 1 MU.

The fold change in gene expression was calculated throughout this work using the following formula:

$$\text{fold change} = \frac{\text{expression}\,(R \neq 0)}{\text{expression}\,(R = 0)}. \qquad \text{[S26]}$$

However, if there was leakiness in our measurements, this result would mean that we are overestimating the expression measurements. If *leak* corresponds to the value of this leakiness, then the corrected fold change in gene expression is

$$\text{fold change} = \frac{\text{expression}\,(R \neq 0) - \text{leak}}{\text{expression}\,(R = 0) - \text{leak}}. \qquad \text{[S27]}$$

Here we have made the implicit assumption that the leakage does not depend on the presence of Lac repressor. Correcting our measurements for leakiness would then result in lower values of the fold change. To determine how much of a difference this correction could make to our calculation of the binding energies we performed an analysis analogous to the one shown in Fig. S8*B* for different proposed values of leakiness ranging between 0 and 1 MU. The results of these different fits are shown in Fig. S9*A*. It is clear from Fig. S9*A* that there would not be a significant change in the binding energies for any of the considered values of leakiness. In Fig. S9 we show the relative change in binding energy between the worst-case scenario (leakiness of 1 MU) and the case where we do not correct for leakiness. It is clear that even in this extreme case the corrections to the binding energies are negligible. We conclude that leakiness, if present, would not be affecting our results in any measurable way.

S6. SI Materials and Methods. ***S6.1. Plasmids.*** Plasmid pZS22-YFP was kindly provided by Michael Elowitz (California Institute of Technology, Pasadena, CA). The EYFP gene comes from plasmid pDH5 (University of Washington Yeast Resource Center) (14). The main features of the pZ plasmids are located between unique restriction sites (15). The sequence corresponding to the *lac*UV5 promoter (16) between positions −36 and +21 was synthesized from DNA oligos and placed between the EcoRI and XhoI sites of pZS22-YFP to create pZS25O1+11-YFP. Note that we follow the notation of Lutz and Bujard (15) and assign the promoter number 5 to the *lac*UV5 promoter. The O1 binding site in pZS25O1+11-YFP was changed to O2, O3, and Oid using site-directed mutagenesis (Quikchange II; Stratagene), resulting in pZS25O2+11-YFP, pZS25O3+11-YFP, and pZS25Oid+11-YFP. These plasmids are shown diagrammatically together with the promoter sequence in Fig. S1.

The *lacZ* gene was cloned from *E. coli* between the KpnI and HindIII sites of all of the single-site constructs mentioned in the previous paragraph. The O2 binding site inside the *lacZ* coding region was deleted without changing the LacZ protein (17), using site-directed mutagenesis. Successful mutagenesis was confirmed by sequencing the new constructs around the mutagenized area.

After we generated these constructs and integrated them on the *E. coli* chromosome, we determined that the different LacZ constructs had acquired some mutations. On average there were three different point mutations in each construct, although pZS25O3+11-lacZ lost both the KpnI and HindIII sites. All these constructs still expressed functional LacZ. This problem did not present itself in the case of the YFP constructs. We attribute this higher number of mutations in part to possible problems in the PCR amplification of the lacZ coding region.

Every time the fold change in gene expression is calculated, the expression of a strain is normalized by the expression of another strain bearing the exact same mRNA sequence. Therefore, we do not believe that the different mRNA sequences and potential different absolute LacZ activities have a considerable effect on the fold change. This assertion is in part also supported by the fact that our experimental data and theoretical predictions match reasonably well. If there is an effect on the fold change due to the differences in the coding region, it seems to be of the same magnitude as the experimental error.

A construct bearing the same antibiotic resistance, but no reporter, was created by deleting YFP from one of our previous constructs. This construct serves to determine the spontaneous hydrolysis or background of our enzymatic measurements.

Plasmid pZS21-lacI was kindly provided by Michael Elowitz (California Institute of Technology, Pasadena, CA). This plasmid has kanamycin resistance. The chloramphenicol resistance gene flanked by FLIP recombinase sites was obtained by PCR from plasmid pKD3. The insert was placed between the SacI and AatII sites of pZS21-lacI to generate pZS3*1-lacI. For this work we wished to have additional concentrations than those provided by pZS3*1-lacI, for which we mutated the ribosomal binding regions. These new ribosomal binding regions were designed using a recently developed thermodynamic model of translation initiation (18). First, the original RBS ("WT") was deleted using site-directed mutagenesis (Quikchange II; Stratagene), using primer 15.29 and its reverse complementary. This primer deleted the sequence between the EcoRI site and the transcription start. From here we proceeded to add new ribosomal binding sequences by mutagenesis using primers 15.2, 15.31, 15.37, and 15.39. All of the primer sequences are shown in Table S4. These primers gave rise to new ribosomal binding regions named RBS1, RBS446, RBS1027, and RBS1147.

S6.2. Strains. Chromosomal integrations were performed using recombineering (19). Primers used for these integrations are shown in Table S4. The reporter constructs were integrated into the galK region (20) of strain HG105, using primers HG6.1 and HG6.3. Note that our reporter gene was integrated in the opposite direction to the neighboring genes to avoid spurious readthrough of the LacZ coding region by RNA polymerase molecules transcribing from nearby promoters. Constructs expressing Lac repressor with the different RBS were integrated into the phage-associated protein *ybcN* (21), using primers HG11.1 and HG11.3.

This integration resulted in strains HG105::ybcn < > 3*1-lacI, HG105::ybcn < > 3*1RBS1-lacI, HG105::ybcn < > 3*1RBS446-lacI, HG105::ybcn < > 3*1RBS1027-lacI, and HG105::ybcn < > 3*1RBS1147-lacI. For simplicity we call these strains 1I, RBS1, RBS446, RBS1027, and RBS1147, respectively. In Table S3 we show the predicted strength from the model and the corresponding number of Lac repressors once the constructs were chromosomally integrated. We can see that even though the predicted and measured values do not correlate too well, the constructs chosen span a wide range of expression levels. This result does not necessarily contradict the results reported in ref. 19 as they claim they can predict the RBS strength within a factor of 2.3.

The reporter constructs were then combined with the different strains expressing varying amounts of Lac repressor, using P1 transduction (openwetware.org/wiki/Sauer:P1vir_phage_transduction). All integrations and transductions were confirmed by PCR amplification of the replaced chromosomal region and by sequencing.

S6.3. Growth conditions and gene expression measurements. Strains to be assayed were grown overnight in 5 mL of LB plus 30 μg/mL kanamycin and chloramphenicol (when needed) at 37 °C and 300 rpm shaking. The cells were then diluted 1:4,000- to 1:1,000-fold into 4 mL of M9 minimal medium plus 0.5% glucose in triplicate culture tubes. Antibiotics were not added at this step. These cells were grown for 6–9 h until an OD_{600} of (approx.) 0.3 was reached after which they were once again diluted 1:10 and grown for another 3 h to 0.3 OD_{600} for a total of >10 cell divisions. At this point cells were harvested and their level of gene expression was measured. Details of our protocol for measuring LacZ activity are given below.

S6.4. β-Galactosidase assay. Our protocol for measuring LacZ activity is basically the one described in refs. 22 and 23 with some slight modifications as follows. A volume of the cells between 2.5 μL and 200 μL was added to Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, pH 7.0) for a total volume of 1 mL. The volume of cells was chosen such that the yellow color would develop in no less than 15 min (and up to several hours). For the case of the no-reporter constructs 200 μL of cell culture was used. Additionally, we included a blank sample with 1 mL of Z-buffer. The whole assay was performed in 1.5-mL Eppendorf tubes.

To lyse the cells, 25 μL of 0.1% SDS and 50 μL of chloroform were added and the mixture was vortexed for 10 s. Finally, 200 μL of 4 mg/mL 2-nitrophenyl β-D-galactopyranoside (ONPG) in Z-buffer was added to the solution and its color, related to the concentration of the product ONP, was monitored visually. Once enough yellow developed in a tube, the reaction was stopped by adding 200 μL of 2.5 M Na_2CO_3 instead of adding 500 μL of a 1-M solution as done in other protocols. At this point the tubes were spun down at >13,000 × *g* for 3 min to reduce the contribution of cell debris to the measurement.

A total of 200 μL of solution was read for OD_{420} and OD_{550} on a Tecan Safire2 and blanked using the Z-buffer sample. The OD_{600} of 200 μL of each culture was read with the same instrument. The absolute activity of LacZ was measured in Miller units using the formula

$$\mathrm{MU} = 1,000\frac{\mathrm{OD}_{420} - 1.75\times\mathrm{OD}_{550}}{t\times v\times\mathrm{OD}_{600}}0.826, \qquad \textbf{[S28]}$$

where *t* is the reaction time in minutes and *v* is the volume of cells used in milliliters. The factor of 0.826 is not present in the usual formula used to calculate Miller units. It is related to using 200 μL Na_2CO_3 as opposed to 500 μL. When using 500 μL, the final volume of the reaction is 1.725 mL (1 mL Z-buffer, 25 μL 0.01% SDS, 200 μL ONPG, 500 μL Na_2CO_3). However, when using only 200 μL of Na_2CO_3, the total volume is 1.425 mL. The factor of 0.826 adjusts for the difference in concentration of ONP.

All reactions were performed at room temperature. No significant difference in activity was observed with respect to performing the assay at 25 °C in an incubator.

S6.4.1. An alternative method to perform the β-galactosidase assay. Even though the β-galactosidase protocol used to obtain the results in the main text is very common, one of the reviewers suggested an alternative approach that could potentially yield more reliable results. One assumption in the protocol described above is that the absolute activity of a culture scales linearly with its cell content. However, instead of measuring the Miller units for a culture at a particular value of OD_{600} one could take various samples at different OD_{600} values and measure the magnitude

$$1,000\frac{\mathrm{OD}_{420} - 1.75\times\mathrm{OD}_{550}}{t\times\upsilon}0.826 \qquad \textbf{[S29]}$$

for each point on the growth curve. The absolute activity from such a procedure can be plotted as a function of the corresponding OD_{600} and from its slope the Miller units can be computed. Conceptually, this method is more compelling because the Miller units are obtained from a fit to multiple points rather than from a single measurement. For simplicity, we call this protocol the "slope" method. The alternative of measuring the activity at only one OD_{600} point is called the "end-point" method.

In Fig. S5 *A* and *B* we show the data for several strains combined with linear fits for each such strain. We repeated this analysis for each strain in our work. As can be seen in Fig. S5, the data fit nicely on a line. We were also interested in the errors incurred in both the slope method and the end-point method. One way to check for differences in these methods is to compare the Miller units obtained from the slope method with those using the last data point (i.e., that obtained for the highest OD_{600} value) as the input for the end-point method. By using the slope method, we are able to estimate an error on the basis of the goodness-of-fit of the straight line. However, this error does not exist in the case of the end-point method. Instead, the error associated with this method originates from uncertainties in the absorption measurements. Fig. S5*C* shows a direct comparison of the two methods over four orders of magnitude in Miller units. The resulting data can be fit to a line with slope 1 nearly perfectly. Additionally, if we perform a linear fit with the intercept fixed to zero we obtain a slope of 1.033 ± 0.005. From this plot we conclude that, at least in terms of mean values, the two methods are basically indistinguishable.

A second way to compare these two methods is through their respective uncertainties. What is the relative importance of errors found in the slope method and those arising from multiple repeats of the same experiment? We estimate this reproducibility by measuring three repeats of each strain. We then compare the following magnitudes: (*i*) Each repeat gives a Miller unit value. We calculate the SD between those three values and its coefficient of variation (CV). We call this "repeat error". (*ii*) Each repeat has an error associated with it as a result of the linear fit. For each repeat we then calculate the CV and take the mean of this CV for a given strain. We call this "fit error". These two errors are plotted as a function of the mean level of expression in Fig. S5*D*. From this plot we conclude that the two errors are similar in magnitude although the repeat error is slightly higher than the fit error in some cases. As a result it appears that there is no extra reliability of the results using the slope method because the sample-to-sample variability induces comparable errors of its own.

Another source of error accounted for in our article is the day-to-day variability. The point here is that when repeating the whole experiment on different days, there will be another kind of variability in the results. For the end-point method we can then compare the repeat error defined above to the "day error". Besides the fact that performing the experiment over multiple days gives a better sense of the reproducibility of the results, for the experiments described in the main text, multiple-day experiments were a necessity as a consequence of the sheer magnitude of the data that was required. Measuring the level of gene expression of all our strains in triplicate and performing the protein purification steps to quantify their absolute content of Lac repressor were not feasible within 1 d. As a result, different strains were quantified over different days, always making sure that each strain had been quantified on at least 4 different days. The corresponding error is calculated by taking the SD of the mean values obtained on different days (which were themselves obtained from averaging over three repeats) and calculating the corresponding CV. In Fig. S5*E* we show both errors as a function of the mean level of gene expression. In this case, we conclude that even though the repeat and day-to-day errors are comparable in some cases, in the majority of the cases the day-to-day variability will be higher than the variability within a day.

As a result of the data presented here we conclude that both methods agree in the mean level of gene expression over four orders of magnitude in Miller units. Their accuracy seems to be comparable.

S6.4.2. Measuring repression using fluorescence. As another check on the reliability of our measurements, we were curious about the quantitative implications inherited with a particular choice of reporter of the level of gene expression. Even though β-galactosidase is one of the most common reporters of gene expression, in recent years, fluorescence reporters have increasingly become the method of choice for many experiments. As a result, we were interested in the extent to which the in vivo binding energies depend upon the method used for the quantification of gene expression. To check this dependence we built constructs bearing O2, O1, and Oid regulating the expression of YFP in the same simple repression circuit considered in the main text (see ref. 24 for details). We measured the corresponding fold change in strain HG104. By using the information from our immunoblots on the number of repressors in that strain we can once again calculate the binding energies just by inverting Eq. **5**. In Table S2 we show a comparison of the fluoresence-derived energies to the binding energies obtained when considering the data for HG104 using the LacZ reporter. As seen in Table S2, the binding energies that are obtained on the basis of fluorescence are comparable to those resulting from the LacZ assay in all cases and have values that fall within error

bars of each other. A more detailed comparison between these two reporters is published elsewhere (24).

S6.5. Measuring in vivo Lac repressor copy number. The Lac repressor purification protocol used in this work is an adaptation of the one published in ref. 25. The strains to be assayed were first grown to saturation in LB + 20 μg/mL of chloramphenicol. They were then diluted 1:40,000 into 50 mL of M9 minimal medium + 0.5% glucose and grown to an OD_{600} of ∼0.6. Cells were spun down (6,000 × *g* for 10 min) and resuspended in 36 μL of breaking buffer (BB) (0.2 M Tris-HCl, 0.2 M KCl, 0.01 M magnesium acetate, 5% glucose, 0.3 mM DTT, 50 μg/L PMSF, 50 mg/100 mL lysozyme, pH 7.6) per milliliter of culture and per OD. Typically, ∼45 mL of culture would be spun down and resuspended in 900 μL of BB. Cells were slowly frozen by placing them at −20 °C, after which they were slowly thawed on ice. At this point 4 μL of a 2,000 Kunitz/mL DNase solution (Sigma) and 40 μL of a 1 M $MgCl_2$ solution were added and the samples were incubated at 4 °C with mixing for 4 h. Samples were frozen, thawed, and incubated with mixing at 4 °C two more times after which they were spun down at 15,000 × *g* for 45 min. At this point the supernatant was obtained and its volume measured. The pellet was subsequently resuspended with 900 μL of BB and spun down again. This sample serves as a control that most Lac repressor was in the original supernatant. The luminescence of these sample resuspensions was compared with respect to the luminescence of the samples corresponding to the first spin. On average, the resuspension signal was ∼12% of the first spin signal. However, some samples showed signals as high as 35%. We chose to discard any data coming from samples showing a resuspension signal >20%.

Additionally to the cell lysates, calibration samples were prepared before performing a measurement. Purified Lac repressor (courtesy of Stephanie Johnson, California Institute of Technology, Pasadena, CA) was diluted into a lysate of strain HG105 to different concentrations. The concentration of purified repressor in our stock solution was determined by spectroscopy using the available extinction coefficient (26). To have all samples within the dynamic range of our methods (see below) cell lysates corresponding to strains 1I and RBS1 were diluted 1:8 in HG105 lysate.

A nitrocellulose membrane was prewetted in TBS (20 mM Tris-HCl, 500 mM NaCl, pH 7.5) for 10 min and then left to air dry. After loading the samples the immunoblots were blocked using blocking solution, which consists of 5% dry milk and 2% BSA in TBST (20 mM Tris Base, 140 mM NaCl, 0.1% Tween 20, pH 7.6), with mixing at room temperature for 1 h. After that the membrane was incubated in a 1:1,000 dilution of Anti-LacI monoclonal antibody (from mouse; Millipore) in blocking solution at 4 °C overnight. The membrane was subsequently incubated in a 1:2,000 dilution of HRP-linked anti-mouse secondary antibody (GE Healthcare) for 1 h at room temperature. Finally, the membrane was washed by incubating in TBST for 5 min twice and by a final incubation of 30 min.

As described in the text, we obtain the total luminescence corresponding to each spot using Matlab image analysis custom code. This information is stored in a matrix Lum(x, y), where the coordinates on the membrane are given by x and y. The values corresponding to the HG105 blank sample are them fitted to a second-degree 2D polynomial. This polynomial can be represented as Background(x, y). Finally, we can also fit such a polynomial to the luminescence of the samples corresponding to strain 1I. This results in the polynomial $1I(x, y)$. In Fig. 3*C* we plot the polynomial $1I(x, y)$ – Background(x, y). The normalized luminescence matrix is then calculated in the following way:

$$\mathrm{Lum}_{\mathrm{norm}}(x,y) = \frac{\mathrm{Lum}(x,y) - \mathrm{Background}\,(x,y)}{1I(x,y) - \mathrm{Background}\,(x,y)}. \qquad \textbf{[S30]}$$

All further analysis is then done on the normalized matrix $\mathrm{Lum}_{\mathrm{norm}}(x, y)$.

The calibration standards are fitted to a power law $\mathrm{LacI}_{\mathrm{lum}} = A \times \mathrm{LacI}_{\mathrm{mass}}^{B} + C$, where $\mathrm{LacI}_{\mathrm{lum}}$ is the luminescence collected from the spots on the membrane and $\mathrm{LacI}_{\mathrm{mass}}$ is their corresponding masses. We are interested in obtaining an interpolation between the calibration samples to get an estimate of the amount of Lac repressor loaded in each spot on the membrane. Therefore, we perform the fit on only the calibration data that are directly in the range of our unknown samples, as shown by the calibration line in Fig. 3*D*.

Once the amount of Lac repressor in each spot was obtained, the corresponding number of Lac repressors per cell was calculated. This calibration between mass detected on the membrane and the corresponding intracellular number of Lac repressors depends on the concentration of cells in the cultures assayed and the volume recovered from the various concentration and lysis steps. As such, there is no calibration factor. As an example, we consider the case where there is one repressor tetramer per cell and estimate the expected amount of repressor on the membrane. We typically start with a 45-mL culture at an OD_{600} of 0.6. This, in turn, is concentrated down to 900 μL after the purification process. A total of 2 μL of these concentrated cells is loaded on the membrane. In this case, we can now calculate the amount loaded on the membrane, resulting in

$$\begin{aligned} N_{\text{cells loaded}} = {} & \underbrace{0.8\times10^{9}\text{cells/ mL}}_{\text{OD}_{600}\text{ to cell density calibration}} \times \underbrace{0.6}_{\text{OD}_{600}} \\ & \times \underbrace{45\text{ mL}}_{\text{culture volume}} \times \underbrace{\frac{2\mu\text{L}}{900\mu\text{L}}}_{\text{final purified volume and amount loaded}} \\ = {} & 48\times10^{6}\text{ cells}. \end{aligned} \qquad \textbf{[S31]}$$

The calibration of OD_{600} to cell density was performed by plating serial dilutions of a culture at a known OD_{600} and counting colonies. This calibration was comparable to $(7.9 \pm 0.5) \times 10^{8}$ cells/mL/OD_{600} obtained using a microfluidic chip where single cells in a culture could be counted by microscopy. The molecular mass of a tetramer is 154 kDa. This mass results in a mass of ∼12 pg in a spot. Of course, there is an uncertainty associated with the calculation of the number of cells loaded that will propagate into the measurement of the number of repressors per cell. However, this uncertainty stems from errors in measuring volumes and in calibrating the OD_{600} readings and they are no larger than 5–10%. On the other hand, the day-to-day variation of the readings was on the order of 20–30%. As a result we chose to report only the day-to-day variation as our error in the measurement of the intracellular concentration of Lac repressor.

1. Bintu L, et al. (2005) Transcriptional regulation by the numbers: Models. *Curr Opin Genet Dev* 15:116–124.
2. Bintu L, et al. (2005) Transcriptional regulation by the numbers: Applications. *Curr Opin Genet Dev* 15:125–135.
3. Rünzi W, Matzura H (1976) *In vivo* distribution of ribonucleic acid polymerase between cytoplasm and nucleoid in *Escherichia coli*. *J Bacteriol* 125:1237–1239.
4. von Hippel PH, Revzin A, Gross CA, Wang AC (1974) Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The *lac* operon: Equilibrium aspects. *Proc Natl Acad Sci USA* 71:4808–4812.
5. Kao-Huang Y, et al. (1977) Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: Measurement of DNA-bound *Escherichia coli lac* repressor *in vivo*. *Proc Natl Acad Sci USA* 74:4228–4232.
6. Gerland U, Moroz JD, Hwa T (2002) Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc Natl Acad Sci USA* 99: 12015–12020.
7. Jishage M, Ishihama A (1995) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: Intracellular levels of sigma 70 and sigma 38. *J Bacteriol* 177: 6832–6835.

8. Barry JK, Matthews KS (1999) Thermodynamic analysis of unfolding and dissociation in lactose repressor protein. *Biochemistry* 38:6520–6528.
9. Schlax PJ, Capp MW, Record MTJ Jr. (1995) Inhibition of transcription initiation by *lac* repressor. *J Mol Biol* 245:331–350.
10. 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.
11. Record MTJ, Reznikoff W, Craig M, McQuade K, Schlax P (1996) *Escherichia coli* RNA polymerase (sigma70) promoters and the kinetics of the steps of transcription initiation. In *Escherichia coli and Salmonella Cellular and Molecular Biology*, eds Neidhardt FC et al. (ASM, Washington, DC), pp 792–821.
12. Buc H, McClure WR (1985) Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the *lac* UV5 promoter. Evidence for a sequential mechanism involving three steps. *Biochemistry* 24:2712–2723.
13. Matlock DL, Heyduk T (1999) A real-time fluorescence method to monitor the melting of duplex DNA during transcription initiation by RNA polymerase. *Anal Biochem* 270: 140–147.
14. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.
15. 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.
16. Müller-Hill B (1996) *The lac Operon: A Short History of a Genetic Paradigm* (Walter de Gruyter, Berlin).
17. 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.
18. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27:946–950.
19. Sharan SK, Thomason LC, Kuznetsov SG, Court DL (2009) Recombineering: A homologous recombination-based method of genetic engineering. *Nat Protoc* 4: 206–223.
20. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
21. Pósfai G, et al. (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science* 312:1044–1046.
22. Miller JH (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
23. Becker NA, Kahn JD, Maher LJ, 3rd (2005) Bacterial repression loops require enhanced DNA flexibility. *J Mol Biol* 349:716–730.
24. Garcia HG, Lee HJ, Boedicker JQ, Phillips R (2011) The limits and validity of methods of measuring gene expression for the testing of quantitative models. *Biophys J*, in press.
25. Xu J, Matthews KS (2009) Flexibility in the inducer binding region is crucial for allostery in the *Escherichia coli* lactose repressor. *Biochemistry* 48:4988–4998.
26. Butler AP, Revzin A, von Hippel PH (1977) Molecular parameters characterizing the interaction of *Escherichia coli lac* repressor with non-operator DNA and inducer. *Biochemistry* 16:4757–4768.

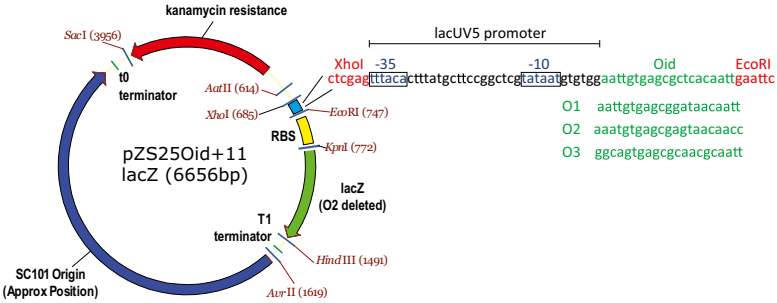

Fig. S1. Plasmid diagram and promoter sequence. The main features of the plasmid pZS25Oid+11-lacZ are shown flanked by unique restriction sites (the features are not to scale). The particular promoter sequence based on the *lac*UV5 promoter is shown together with the sequences of the different Lac repressor binding sites used.

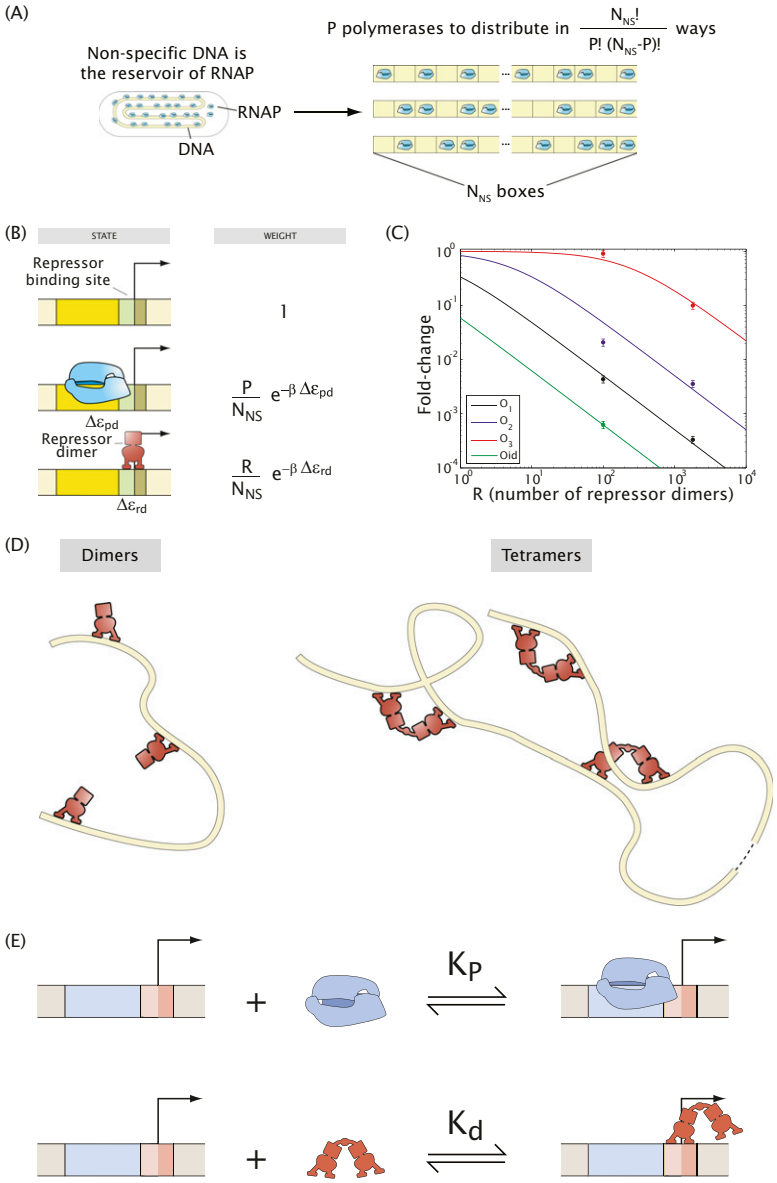

Fig. S2. Thermodynamic model of transcription and simple repression. (*A*) Model for the RNA polymerase reservoir. The nonspecific sites on the genome are assumed to be the reservoir for RNAP. Different arrangements of RNAP on this reservoir are shown. (*B*) Schematic listing of the different states and their respective weights for repression by Lac repressor dimers, when RNAP and the dimeric repressor have overlapping sites. (*C*) Repression for four different strengths of the main repressor binding site (Om) as a function of the number of dimers inside the cell. The binding energy of dimeric Lac repressor to each operator is calculated by fitting each dataset to the repression expression from Eq. S11 and is presented in Table S1. (*D*) Model for the nonspecific looping background. Possible states of nonspecific DNA bound by Lac repressor dimers, which will explore all available nonspecific sites, and tetramers, which will explore all possible loops between nonspecific sites. (*E*) Repression as a set of chemical reactions. The two reactions involved in regulation by simple repression are shown. K_P and K_d are dissociation constants. These reactions are also described by Eqs. **S20** and **S21**.

A

Binding constants obtained from data of Oehler *et al.*		
Operator	Binding energy (k_BT)	Approx. dissociation constant
Oid	-17.7 ± 0.3	90 pM
O1	-16.2 ± 0.1	380 pM
O2	-13.7 ± 0.1	4.7 nM
O3	-10.4 ± 0.4	130 nM

C

Strain	Prediction (repressors/cell)
HG104	9 ± 2
RBS1147	48 ± 8
RBS446	60 ± 20
RBS1027	130 ± 40
RBS1	220 ± 70
11	400 ± 100

B

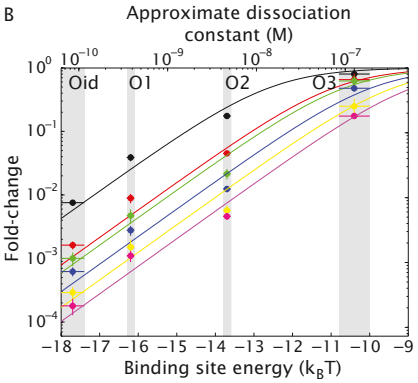

Fig. S3. Single-site binding energies and prediction of the number of repressors for different strains using energies deduced from the Oehler et al. data. (*A*) The operator binding energies and dissociation constants are deduced from the data by Oehler et al. (10) using Eq. **5**. The error bars are calculated assuming an error in the fold-change measurement of 30% and assuming no error in the number of repressor molecules. (*B*) The fold change in gene expression is measured for all four operators in six different strain backgrounds. 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.

(A)

Strain	Direct measurement (repressors/cell)
HG104	11 ± 2
RBS1147	30 ± 10
RBS446	62 ± 15
RBS1027	130 ± 20
RBS1	610 ± 80
11	870 ± 170

(B)

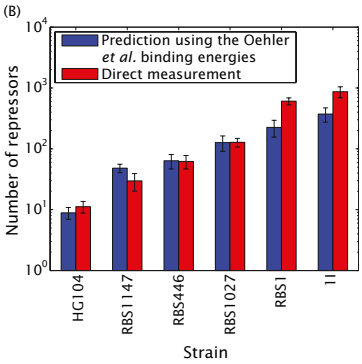

Fig. S4. Experimental and theoretical characterization of repressor copy number. (*A*) Immunoblots were used to measure the cellular concentration of Lac repressor in six strains with different constitutive levels of Lac repressor. Each value corresponds to an average of cultures grown on at least 3 different days. The error bars are the SD of these measurements. (*B*) The fold-change measurements in Fig. 2 were combined with the binding energies obtained from Fig. S3*A* (derived from previous experimental results) (10) 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.

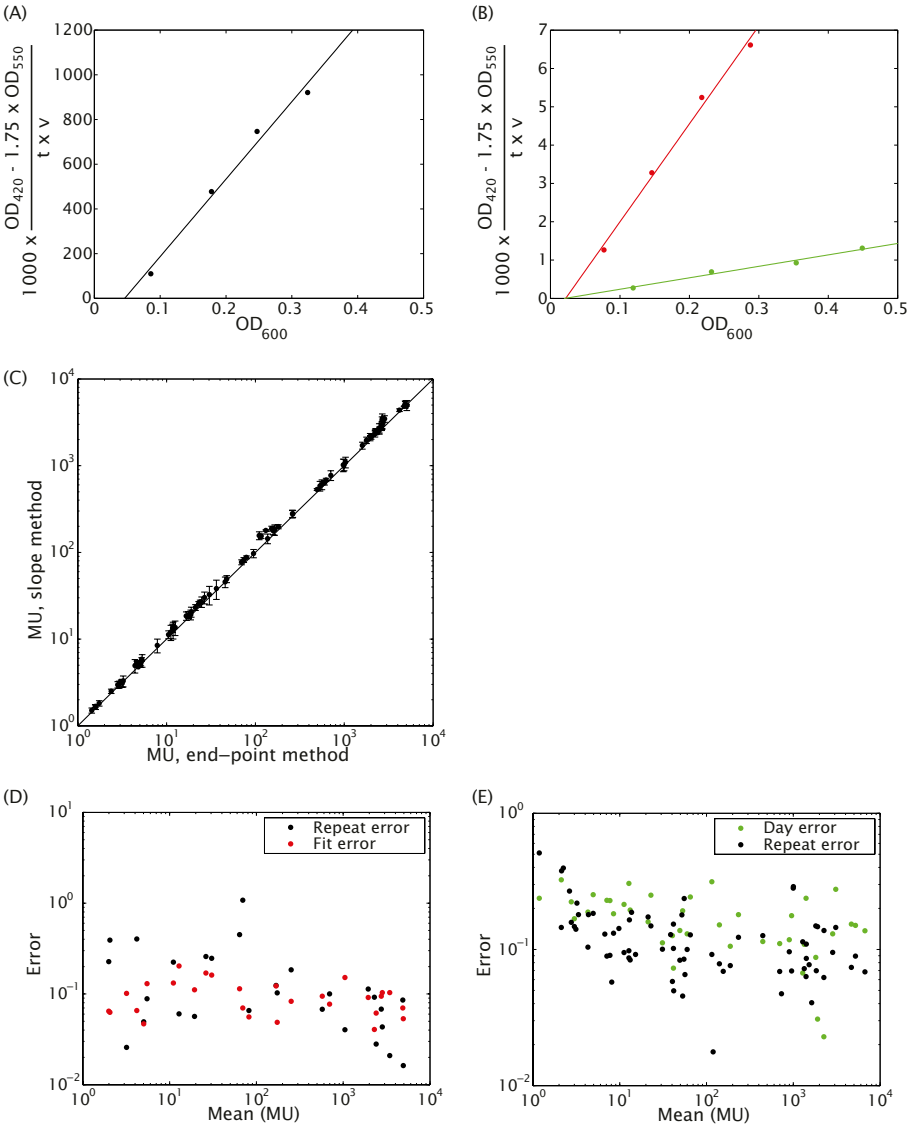

Fig. S5. The "slope" method to calculate β-galactosidase activity. (*A* and *B*) The quantity $1{,}000((OD_{420} - 1.75 \times OD_{550})/(t \times v))$ is plotted for three different samples as a function of OD_{600} for several representative strains spanning the whole range of expression covered in our experiment. The curves in all cases are linear fits to the data. (*C*) Explicit comparison of the two methods to quantify β-galactosidase activity over four orders of magnitude. The results of the slope and end-point methods are plotted against each other. The line has a slope of 1. A linear fit to the data with a fixed zero intercept yields a slope of 1.033 ± 0.005. (*D* and *E*) Errors associated with day-to-day variability and repeat variability vs. linear fits. (*D*) The repeat and fit errors are shown as a function of the mean level of expression. From this plot it is clear that both errors are comparable with a slight bias of the repeat error to be higher than the fit error. (*E*) The average relative error stemming from averaging repeats over 1 d is compared with the relative error originating from averaging over multiple days and plotted as a function of the mean level of gene expression. Both errors seem to be comparable with a slight bias for the day-to-day variability to be higher.

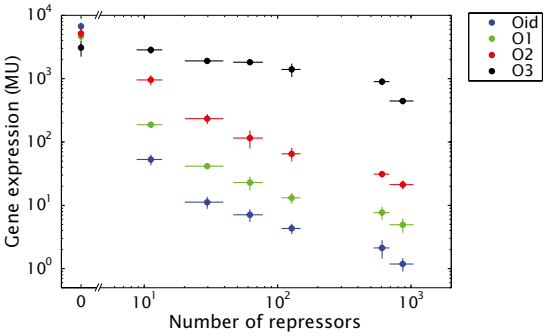

Fig. S6. Average absolute levels of expression. The absolute levels of expression corresponding to our different constructs in the different strain backgrounds are shown in Miller units. By the ratio of the activity of a given construct in a given strain with respect to the activity of the same construct in strain HG105 we calculate the fold change in gene expression. Note that throughout this work the repression values correspond to the average of the repression measured on different days. In this case we plot the average of the absolute expression of each strain and construct over different days. The error bars correspond to the SD of the repeats.

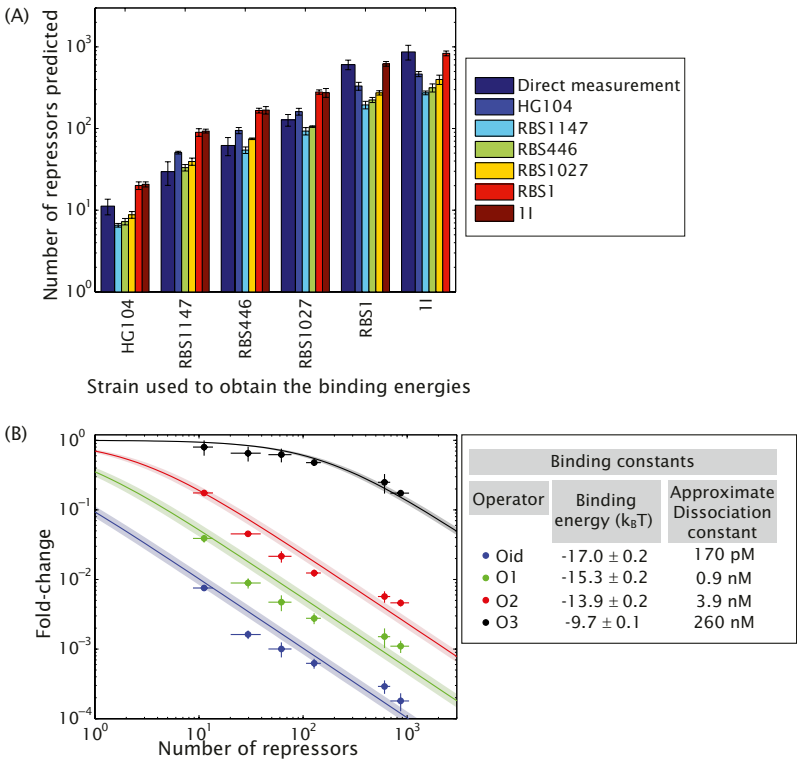

Fig. S7. Different ways of calculating the binding energies give comparable predictions. (*A*) For each strain noted by a group of bars the binding energies were obtained by taking the number of repressors obtained through immunoblots as a given and combining this number with the fold-change measurements for the same strain. With these binding energies we predict the number of repressors for all of the remaining strains. For comparison, the actual direct measurement done using immunoblots is also included. (*B*) Using all measurements of the fold change in gene expression with their corresponding repressor concentration we fit Eq. **5** to obtain the best possible estimate for the binding energies. The results of the fits are expressed in units of k_BT.

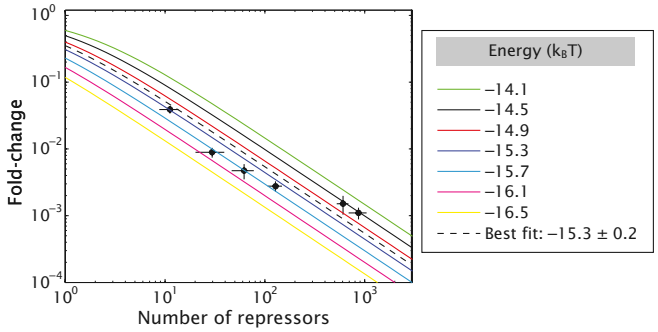

Fig. S8. Sensitivity in the determination of the binding energies. The data for binding site O1 are shown with its best fit along with several other choices of the binding energy parameter, which reveal how the positions of the curves depend upon this choice. Visual inspection of the curves constrains the value of the binding energy to within $<1k_BT$ of the fit value.

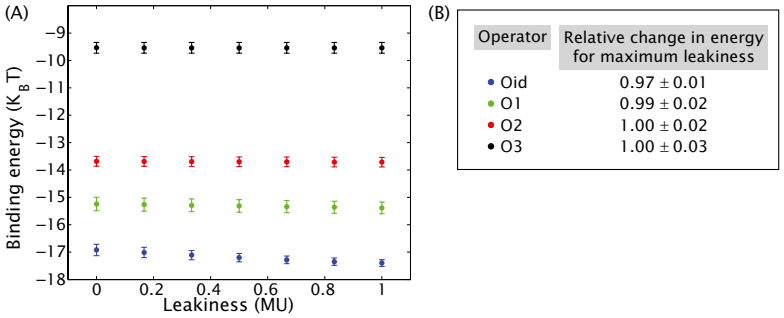

Fig. S9. Potential effects of leakiness on the calculation of binding energies. (*A*) A variable leakiness in the level of gene expression was assumed and the fold change in gene expression was reanalyzed using Eq. **S27**. The resulting binding energies are shown as a function of the assumed leakiness. (*B*) Relative change in binding energies for each operator corresponding to the case without any assumed leakiness and to the worst possible leakiness of 1 MU.

Table S1. Single-site binding energies for repressor dimers and tetramers using the data by Oehler et al.

Operator	Dimers (k_BT)	Tetramers (k_BT)
Oid	−18.2 ± 0.3	−17.7 ± 0.3
O1	−16.1 ± 0.2	−16.2 ± 0.1
O2	−13.7 ± 0.5	−13.7 ± 0.1
O3	−10.0 ± 0.4	−10.4 ± 0.4

The energies are obtained using the data by Oehler et al. (10) and Eq. **S11** and Eq. **5** for the dimers and tetramers, respectively. The error bars are calculated assuming an error in the fold-change measurement of 30%.

Table S2. Binding energies calculated using YFP and LacZ as reporters of gene expression

Operator	Energy from YFP (k_BT)	Energy from LacZ (k_BT)
Oid	−16.8 ± 0.4	−17.2 ± 0.2
O1	−15.1 ± 0.2	−15.5 ± 0.3
O2	−13.8 ± 0.4	−13.9 ± 0.2

The fold change of constructs bearing O2, O1, and Oid and either a LacZ or a YFP reporter were measured in strain HG104. By combining these measurements with our knowledge of the number of repressors within the strain we can compute the corresponding binding energies. In all cases the obtained binding energies are comparable within error bars.

Table S3. Predicted and measured strength of the different ribosomal binding sequences used to generate constitutive levels of Lac repressor

RBS	Normalized predicted strength (au)	Normalized measured strength (repressors/cell)
"WT"	1	1 ± 0.2
RBS1	0.88	0.7 ± 0.2
R1027	0.58	0.15 ± 0.04
R446	0.25	0.07 ± 0.02
R1147	0.64	0.03 ± 0.01

The ribosomal binding sequence denoted as "WT" corresponds to the original found in pZS3*1-lacI (16). The measured strength corresponds to the resulting level of repressor once these constructs are integrated on the chromosome. The predicted strengths are calculated from ref. 19. Both the predicted and the measured strengths are normalized by this RBS.

Table S4. Primers and *E. coli* strains used throughout this work

Primer number and name	Sequence	Description
15.29-RBSDelete	gacgcactgaccgaattcatggtgaatgtgaaaccag	Delete the RBS from pZS3*1-lacI
15.2-tetR-RBS1	cgcactgaccgaattcattaaagaTTT gaaaggtaccatatggtg	
15.31-RBS446	cgcactgaccgaattc TCTAGACAGTATAGAGTAGAGAGACTAA atggtgaatgtgaaac	
15.37-RBS1027	cgcactgaccgaattc TCTAGATATTTAAGAGGACAATACTGG atggtgaatgtgaaac	
15.39-RBS1147	cgcactgaccgaattc TCCCCACATTAAACAGGGAAGACTGG atggtgaatgtgaaac	
HG6.1	gtttgcgcgcagtcagcgatatccattttcgcgaatccggagtg taagaa ACTAGCAACACCAGAACAGCC	Integration of the lacZ reporter constructs into the *galK* gene between positions 1,504,078 and 1,505,112.
HG6.3	ttcatattgttcagcgacagcttgctgtacggcaggcaccagct cttccg GGCTAATGCACCCAGTAAGG	
HG11.1	acctctgcggaggggaagcgtgaacctctcacaagacggcatca aattac ACTAGCAACACCAGAACAGCC	Integration of lacI constructs into the *ybcN* gene between positions 1,287,628 and 1,288,047.
HG11.3	ctgtagatgtgtccgttcatgacacgaataagcggtgtagccat tacgcc GGCTAATGCACCCAGTAAGG	

Strain	Genotype	Comment
HG104	Δ*lacZY A*	Deletion in MG1655 from 360,483 to 365,579.
HG105	Δ*lacZY A*, Δ*lacI*	Deletion in MG1655 from 360,483 to 366,637.

The first five primers and their respective reverse complement were used to modify the RBS of the different constructs. The inserted RBS regions are denoted by uppercase bases. The remaining primers are used for integration. Lowercase indicates the portion of the primer that is homologous to the *E. coli* gene where the integration is made and uppercase indicates primer homology to the plasmid where PCR was carried out. Chromosomal positions correspond to the sequence in GenBank accession no. U00096.