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

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

where $\beta = 1/K_B T$. 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, ε_{pd}^{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 $N_{NS} \gg P$, 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

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

where, in analogy to the nonspecific binding energy, we have defined ε_{pd}^{S} as the binding energy between RNAP and the pro-

moter. 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

$$\begin{aligned}p_{bound}(P) &= \frac{e^{-\beta\varepsilon_{pd}^{S}}Z^{NS}(P-1; N_{NS})}{Z^{NS}(P; N_{NS}) + e^{-\beta\varepsilon_{pd}^{S}}Z^{NS}(P-1; N_{NS})}\\ &= \frac{1}{1 + \frac{N_{NS}}{P}e^{\beta\Delta\varepsilon_{pd}}},\end{aligned} \quad \textbf{[S3]}$$

with $\Delta\varepsilon_{pd} = \varepsilon_{pd}^{s} - \varepsilon_{pd}^{NS}$, 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 ε_{pd}^{NS} and ε_{rd}^{NS}, respectively, the nonspecific partition function becomes

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

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

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

where ε_{pd}^S and ε_{rd}^S 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 $\Delta\varepsilon_{pd} = \varepsilon_{pd}^S - \varepsilon_{pd}^{\text{NS}}$ and $\Delta\varepsilon_{rd} = \varepsilon_{rd}^S - \varepsilon_{rd}^{\text{NS}}$ as the energy gain of RNAP and dimeric LacI when switching from a non-specific site to their respective specific sites, respectively. The probability of finding RNAP bound to the promoter is given by

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

This expression can be rewritten as

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

where we have defined the regulation factor

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

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:

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

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

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

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

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

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. S2C 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. S2D. 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 $\varepsilon_{rd}^{\text{NS}}$ 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

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

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

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

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

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

$$Z^{\text{NS}}(R_4) = \frac{\left[Z^{\text{NS}}(R_4 = 1)\right]^{R_4}}{R_4!} = \frac{1}{2^{R_4}} \frac{(N_{\text{NS}})^{R_4}}{R_4!} e^{-\beta R_4 2\varepsilon_{rd}^{\text{NS}}} e^{-\beta R_4 F_{\text{loop}}^{\text{NS}}},$$

[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

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

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:

$$Z_R^{O, \text{NS}} = \frac{1}{2} e^{-\beta \varepsilon_{rd}^S} e^{-\beta \varepsilon_{rd}^{\text{NS}}} \left(\sum_{i=1}^{N_{\text{NS}}} e^{-F_{\text{loop}}(i, i_0)} + \sum_{j=1}^{N_{\text{NS}}} e^{-F_{\text{loop}}(i_0, j)} \right).$$

[S16]

Because both sums are identical, we can reduce this to

$$Z_R^{O, \text{NS}} = e^{-\beta \varepsilon_{rd}^S} e^{-\beta \varepsilon_{rd}^{\text{NS}}} \sum_{j=1}^{N_{\text{NS}}} e^{-F_{\text{loop}}(i_0, j)} = e^{-\beta \varepsilon_{rd}^S} e^{-\beta \varepsilon_{rd}^{\text{NS}}} e^{-\beta F_{\text{loop}}^{\text{NS}}}.$$

[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

$$Z_{\text{total}}(P, R) = Z^{\text{NS}}(P, R) + Z^{\text{NS}}(P-1, R) e^{-\beta \varepsilon_{pd}^S} + Z^{\text{NS}}(P, R-1) \times Z_R^{O, \text{NS}}.$$

[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

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

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

$$\frac{[P][D]}{[P-D]} = K_P \qquad [S20]$$

and

$$\frac{[R][D]}{[R-D]} = K_d. \qquad [S21]$$

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

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

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

$$p_{\text{bound}} = \frac{[P]/K_P}{1 + [R]/K_d + [P]/K_P}. \qquad [S23]$$

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

$$\frac{[R]}{K_d} = \frac{2R}{N_{\text{NS}}} e^{-\beta \Delta \varepsilon_{rd}}, \qquad [S24]$$

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

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

where K_{d}^{S} and K_{d}^{NS} are the dissociation constants of RNA polymerase to specific and nonspecific DNA, respectively. In vitro values for the nonspecific dissociation constant are $K_{d}^{NS}\approx 10,000$ 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 $k_{B}T$. 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 $<1k_{B}T$) 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

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

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 plas-mid 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$_2$CO$_3$ 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

$$MU = 1,000 \frac{OD_{420} - 1.75 \times OD_{550}}{t \times v \times OD_{600}} 0.826, \quad \textbf{[S28]}$$

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$_2$CO$_3$ 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$_2$CO$_3$). However, when using only 200 μL of Na$_2$CO$_3$, 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

$$1,000 \frac{OD_{420} - 1.75 \times OD_{550}}{t \times v} 0.826 \quad \textbf{[S29]}$$

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. S5C 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. S5D. 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. S5E 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. 3C we plot the polynomial $1I(x, y) - Background(x, y)$. The normalized luminescence matrix is then calculated in the following way:

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

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

The calibration standards are fitted to a power law $LacI_{lum} = A \times LacI_{mass}^{B} + C$, where $LacI_{lum}$ is the luminescence collected from the spots on the membrane and $LacI_{mass}$ 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. 3D.

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

$$N_{cells\ loaded} = \underbrace{0.8 \times 10^9 \text{cells/ mL}}_{OD_{600}\ \text{to cell density calibration}} \times \underbrace{0.6}_{OD_{600}}$$

$$\times \underbrace{45\ \text{mL}}_{\text{culture volume}} \times \underbrace{\frac{2\mu L}{900\mu L}}_{\text{final purified volume and amount loaded}}$$

$$= 48 \times 10^6 \text{ cells.}$$

$$\textbf{[S31]}$$

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.
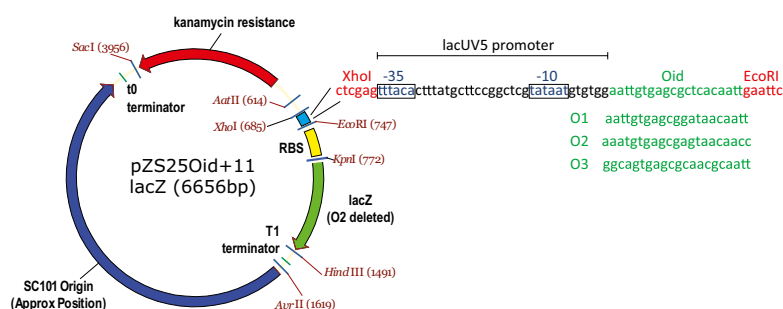
Fig. S1. Plasmid diagram and promoter sequence. The main features of the plasmid pZS25Oid+11-lacZ are shown flanked by unique restriction sites (the features are not to scale). The particular promoter sequence based on the *lac*UV5 promoter is shown together with the sequences of the different Lac repressor binding sites used.
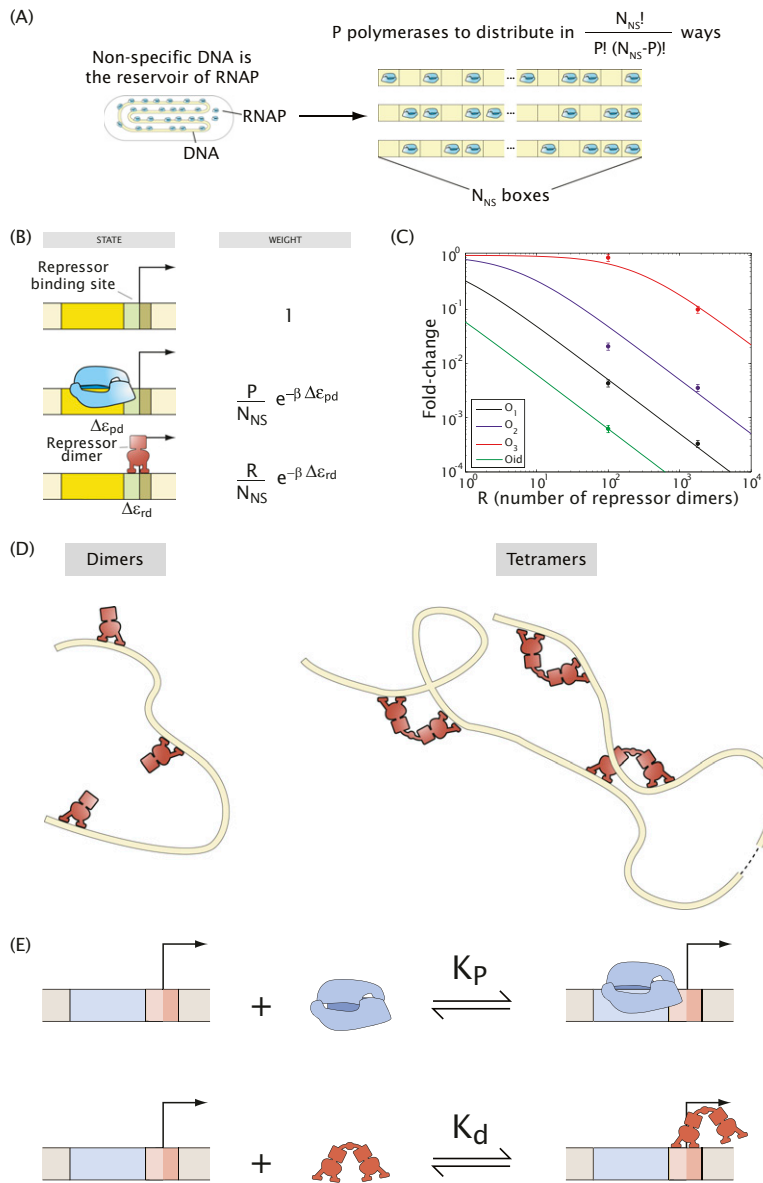
Fig. S2. Thermodynamic model of transcription and simple repression. (*A*) Model for the RNA polymerase reservoir. The nonspecific sites on the genome are assumed to be the reservoir for RNAP. Different arrangements of RNAP on this reservoir are shown. (*B*) Schematic listing of the different states 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**.
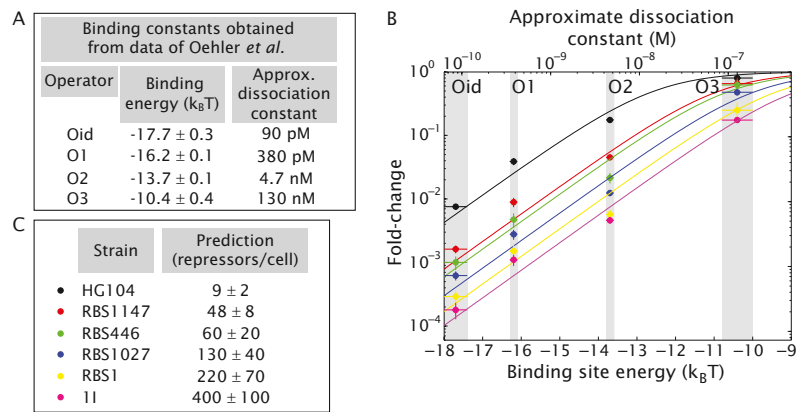
Fig. S3. Single-site binding energies and prediction of the number of repressors for different strains using energies deduced from the Oehler et al. data. (*A*) The operator binding energies and dissociation constants are deduced from the data by Oehler et al. (10) using Eq. **5**. The error bars are 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.
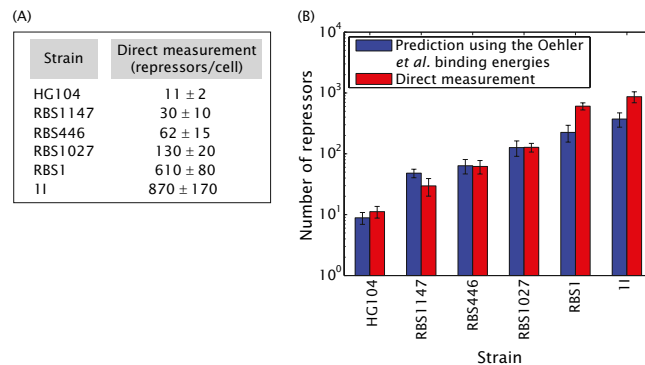

Fig. S4. Experimental and theoretical characterization of repressor copy number. (*A*) Immunoblots were used to measure the cellular concentration of Lac repressor in six strains with different constitutive levels of Lac repressor. Each value corresponds to an average of cultures grown on at least 3 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.
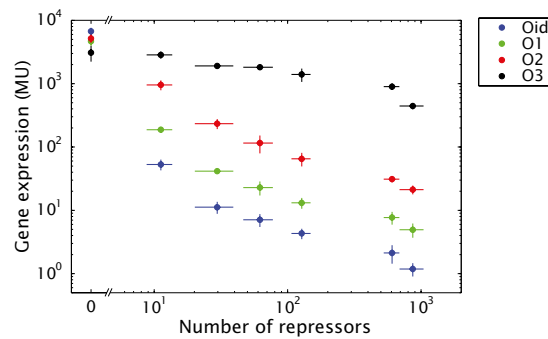
Fig. S5. The "slope" method to calculate β-galactosidase activity. (*A* and *B*) The quantity $1,000((OD_{420} - 1.75 \times OD_{550})/(t \times v))$ is plotted for three different samples as a function of OD_{600} for several representative strains spanning the whole range of expression covered in our 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.
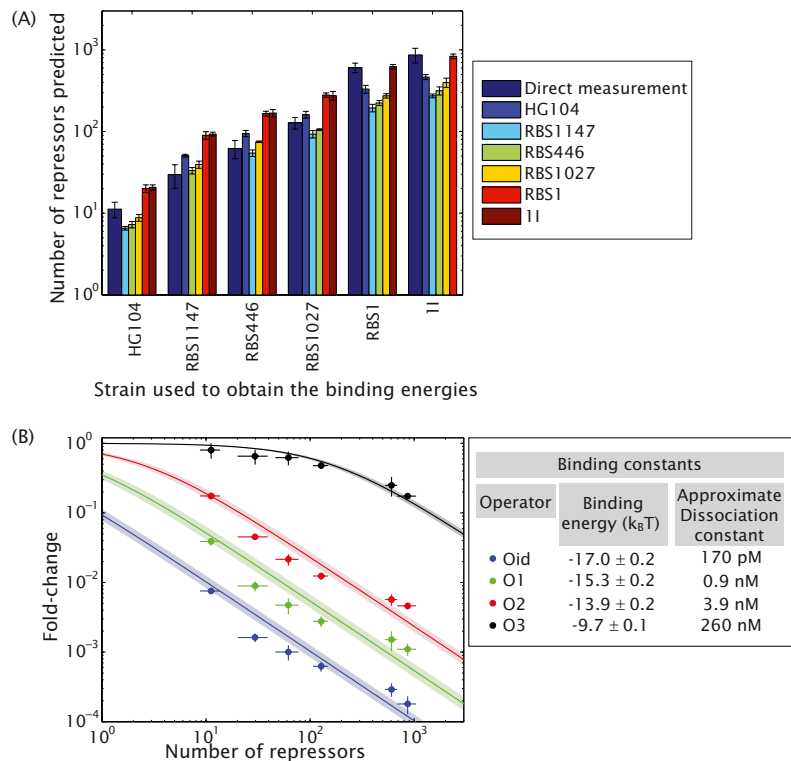

Fig. S7. Different ways of calculating the binding energies give comparable predictions. (*A*) For each strain noted by a group of bars the binding energies were obtained by taking the number of repressors obtained through immunoblots as a given and combining this number with the fold-change measurements 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.
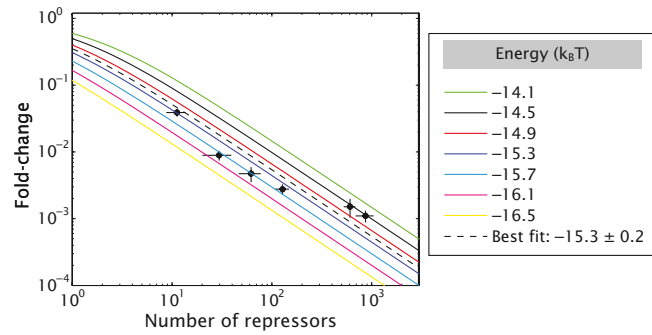
Fig. S8. Sensitivity in the determination of the binding energies. The data for binding site O1 are shown with its best fit along with several other choices of the binding energy parameter, which reveal how the positions of the curves depend upon this choice. Visual inspection of the curves constrains the value of the binding energy to within $<1 k_B T$ of the fit value.
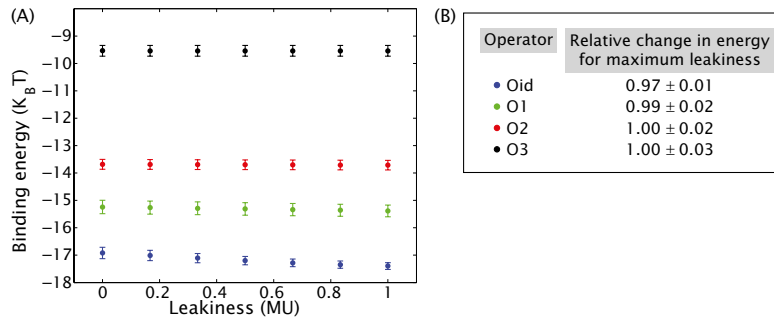

Fig. S9. Potential effects of leakiness on the calculation of binding energies. (*A*) A variable leakiness in the level of gene expression was assumed and the fold change in gene expression was reanalyzed using Eq. **S27**. The resulting binding energies are shown as a function of the assumed leakiness. (*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 ($k_B T$)	Tetramers ($k_B T$)
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	ΔlacZY A	Deletion in MG1655 from 360,483 to 365,579.
HG105	ΔlacZY A, ΔlacI	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.