Cell Reports
Article

Operator Sequence Alters Gene Expression Independently of Transcription Factor Occupancy in Bacteria

Hernan G. Garcia,[1,6,9] Alvaro Sanchez,[3,7,9] James Q. Boedicker,[2,9] Melisa Osborne,[4,8] Jeff Gelles,[4,*] Jane Kondev,[5,*] and Rob Phillips[2,*]
[1]Department of Physics
[2]Department of Applied Physics
California Institute of Technology, Pasadena, CA 91125, USA
[3]Graduate Program in Biophysics and Structural Biology
[4]Department of Biochemistry
[5]Department of Physics
Brandeis University, Waltham, MA 02453, USA
[6]Department of Physics, Princeton University, Princeton, NJ 08540, USA
[7]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]Harvard University School of Medicine and Children's Hospital, Boston, MA 02115, USA
[9]These authors contributed equally to this work
*Correspondence: gelles@brandeis.edu (J.G.), kondev@brandeis.edu (J.K.), phillips@pboc.caltech.edu (R.P.)
http://dx.doi.org/10.1016/j.celrep.2012.06.004

SUMMARY

A canonical quantitative view of transcriptional regulation holds that the only role of operator sequence is to set the probability of transcription factor binding, with operator occupancy determining the level of gene expression. In this work, we test this idea by characterizing repression in vivo and the binding of RNA polymerase in vitro in experiments where operators of various sequences were placed either upstream or downstream from the promoter in *Escherichia coli*. Surprisingly, we find that operators with a weaker binding affinity can yield higher repression levels than stronger operators. Repressor bound to upstream operators modulates promoter escape, and the magnitude of this modulation is not correlated with the repressor-operator binding affinity. This suggests that operator sequences may modulate transcription by altering the nature of the interaction of the bound transcription factor with the transcriptional machinery, implying a new layer of sequence dependence that must be confronted in the quantitative understanding of gene expression.

INTRODUCTION

Cells control how much, when and where to express a gene in response to changes in their intracellular and extracellular environments. A variety of mechanisms are employed to exert this control at each of the steps along the path from DNA to active protein (Alberts, 2008). An important mechanism of gene regulation in bacteria acts through transcription factors that bind to specific sites in the promoter region, the sequence of DNA immediately upstream of genes, where RNA polymerase binds. As a result of interactions or steric interference between transcription factors bound to these sites and RNA polymerase, activation or repression of transcription ensues (Bintu et al., 2005b; Ptashne and Gann, 2002). Indeed, an important activity of modern genome science is finding transcription factor binding sites and determining the rules by which promoter architecture, i.e., the position and sequence of these binding sites, dictates the level of the gene expression (Buchler et al., 2003; Segal and Widom, 2009).

It is often assumed that that the role of operators is simply to act as docking sites for transcription factors, recruiting them to the promoter region (Meijsing et al., 2009). In this view, which we here term the "occupancy hypothesis" the sequence of the operator simply determines its binding affinity for its target transcription factor. This binding affinity, together with the concentration of active transcription factors and the interactions with other DNA-binding proteins, determines the occupancy of the operator which, in turn, is thought to influence the level of transcriptional regulation exerted by the transcription factor (Alberts, 2008; Bintu et al., 2005b; Buchler et al., 2003; Davidson, 2006). For example, in Figure 1A we consider promoters presenting binding sites of different affinities for a repressor. Given these binding affinities the intracellular number of active repressors will determine the probability of finding the repressor bound to each one of the operators. As a result, the shape of the input-output function, that is, the level of output gene expression as a function of the input concentration of repressors, will reach the same level of repression at different repressor concentrations, which are determined by the binding affinities of the operators. A promoter that contains a strong operator, a site on the DNA that binds the transcription factor tightly, is expected to require a lower intracellular concentration of the transcription factor to reach the same level of repression (or activation) as

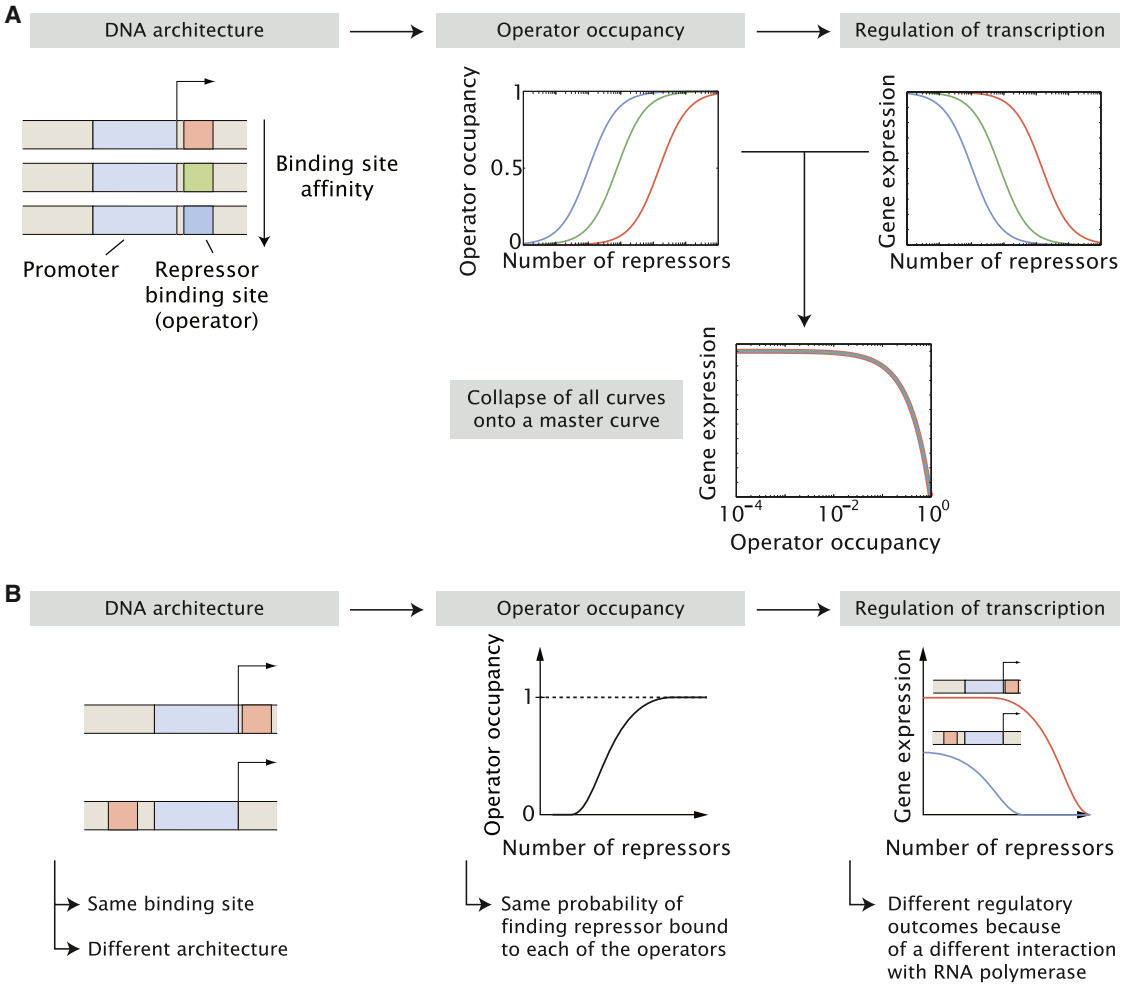

Figure 1. The Occupancy Hypothesis for the Action of Transcription Factors and Its Predictions for Gene Regulation

(A) Binding sites of different affinity lead to different occupancies as a function of the number of repressors with weaker operators requiring higher repressor number to achieve the same occupancy. The shapes of both the occupancy and gene expression curves as a function of the number of repressors depend on the operator binding affinity. However, all these curves collapse onto a single universal curve when the level of gene expression is plotted as a function of operator occupancy. Note that the use of repression to make this point is inconsequential and that the same point can be made using activation of gene expression.

(B) An alternative modulation of regulatory architecture can stem from moving the binding site with respect to the promoter by keeping the operator binding affinity constant. In this case, the probability of finding repressor bound to the operator (operator occupancy) should presumably be the same for both constructs. The positioning of the operator, however, is expected to modulate the nature of the interaction between the repressor and the transcriptional machinery leading to different input-output functions for each regulatory architecture.

a promoter that has a weaker operator (Bintu et al., 2005a, 2005b; Buchler et al., 2003; Vilar and Leibler, 2003). Quantitatively, this can be expressed by saying that the level of transcription is determined by the binding probabilities of these transcription factors to the DNA, as shown in Figure 1A. As a result, a key prediction of the occupancy hypothesis is that when plotting the level of gene expression as a function of the operator occupancy the curves corresponding to different operators, regardless of their affinity, will all collapse onto a master curve as shown in Figure 1A.

The effect of the modulation of the affinity of an operator as the surrounding sequence context is kept constant is not to be confused with the effect of moving a given operator with respect to the promoter while leaving its affinity constant, shown diagrammatically in Figure 1B. Because the affinity is kept constant one would, in principle, expect the probability of finding the transcription factor bound to the operator in each of the constructs exemplified in Figure 1B to be the same as a function of the transcription factor concentration. However, we expect the relative positioning of the operator and promoter to modulate the nature of the interaction between the transcription factor and the transcriptional machinery. As a result, the shape of the input-output function of each regulatory architecture will, in principle, differ.

From the point of view of transcriptional regulation described above it follows that once the occupancy of binding sites by

transcription factors within a regulatory region is determined and the effect of the bound transcription factor on the transcriptional machinery is known, the resulting level of gene expression can be calculated (Bintu et al., 2005b; Buchler et al., 2003; Segal et al., 2008; Vilar and Leibler, 2003). This view of transcriptional regulation has been challenged by recent results in eukaryotic cells that demonstrated that the affinity of transcription factors for different cofactors can be modulated by the sequence of the transcription factor binding site (Lefstin and Yamamoto, 1998; Meijsing et al., 2009). This suggests that in order to fully understand the function of a regulatory region, the effect of operators on the nature of the interaction between transcription factors and the transcriptional machinery may have to be considered in addition to finding their position in the genome and their affinity for transcription factors.

Despite this recent evidence, many quantitative studies both in the bacterial and eukaryotic context, make explicit or implicit use of the occupancy hypothesis in order to describe the action of transcription factors on the level of gene expression (Ackers et al., 1982; Amit et al., 2011; Davidson, 2006; Garcia and Phillips, 2011; Gertz et al., 2009; Ptashne and Gann, 2002; Segal et al., 2008; Zinzen et al., 2009). Indeed, every time that the rate of protein production is written in terms of Hill functions, for example, this occupancy hypothesis has been made implicitly (Cağatay et al., 2009; Elowitz and Leibler, 2000; Fowlkes et al., 2008; Gardner et al., 2000; Klumpp et al., 2009; Kuhlman et al., 2007; Novák and Tyson, 2008; Tsai et al., 2008). As a result, such quantitative descriptions of transcriptional regulation are at least potentially incomplete and not on par with our current qualitative knowledge of the nuanced role of operator sequence beyond that of determining binding affinity.

In the remainder of the study, we demonstrate a form of modulation in transcriptional regulation that is at odds with the traditional operator occupancy viewpoint, suggesting that the canonical picture is incomplete. We do this by adopting a synthetic biology approach, in which we deliberately tune operator position, operator strength and transcription factor copy number in order to systematically traverse the parameter space of the simple repression architecture (i.e., the case in which a repressor regulates a promoter through the presence of a single binding site in its vicinity) by Lac repressor. This repressor is one of the best understood transcription factors (Müller-Hill, 1996). Through systematic in vivo gene expression measurements, in vitro single molecule experiments, and theoretical modeling we show that when the repressor binds upstream from the promoter the choice of the sequence of its operator binding site influences the rate of synthesis of mRNA, but that the extent of repression does not respect the rank ordering of the strength of the different operators in the way predicted by the occupancy-based model of regulation (Figure 1A). As a result, we expand the quantitative view for the role of operator sequence in the context of the paradigmatic Lac repressor-operator interaction. In this context, operator sequence acts not only to determine the occupancy of DNA binding proteins, but also affects the nature of the interactions between the transcription factors and the transcriptional machinery.

RESULTS

Operator Position Leads to Different Repression Behaviors

In *Escherichia coli*, genome-wide studies have resulted in an atlas of binding sites for both repressors and activators that give a picture of the diversity of binding site arrangements even in the case of simple repression. For example, in Figure 2A we show a histogram of the positions of repressor binding sites that regulate promoters through simple repression in *E. coli* (Gama-Castro et al., 2008; Madan Babu and Teichmann, 2003). As can be seen from this histogram the simple repression motif may be able to act over a wide range of positions relative to the polymerase start site.

In order to investigate the effects of *lac* operator position relative to the polymerase binding site, we carried out systematic gene expression measurements for different operators as a function of their position relative to the transcription start site with single base pair resolution. Examples of the parameters varied in the construct library used to assay the effect of operator positioning on repression are shown schematically in Figure 2B, whereas a more detailed version including the sequences is shown in Figures S1A–S1C. We used the *lac*UV5 promoter, which is a mutant of the *lac* promoter that does not require activation by CRP (Müller-Hill, 1996). This promoter controls the expression of the YFP or *lacZ* gene, which we use to quantify the level of gene expression. We measure the regulatory effect of Lac repressor as repression, which is defined as

$$\text{Repression} = \frac{\text{Expression}(R=0)}{\text{Expression}(R\neq 0)}, \qquad \text{(Equation 1)}$$

where R is the intracellular number of repressors. When the operator is moved downstream from the transcriptional initiation site, the level of repression is relatively independent of position until the center of the operator reaches +16 as shown in Figure 2C. At these downstream positions, Lac repressor might be acting by the same mechanism as it does at +11, where it blocks open complex formation or an earlier step in initiation (Sanchez et al., 2011; Schlax et al., 1995). However, it is also possible that as the repressor is moved from its wild-type position repression might be realized through different mechanisms, as has already been shown for a variety of transcription factors (Hochschild and Dove, 1998; Pavco and Steege, 1990, 1991; Rojo, 1999), possibly affecting any of the various steps in transcription initiation or elongation as shown schematically in Figures S2A and S2B (Elledge and Davis, 1989; Lopez et al., 1998).

By way of contrast, when the operator is moved to positions upstream from the initiation site, the level of repression strongly depends on the location of the operator and the variation in repression is substantial, with at least a 15-fold effect between the peaks and valleys and with the valleys corresponding to no repression (Figure 2C). Interestingly, the repression profile shows two peaks, with a separation between them of ~10–11 bp, intriguingly close to the helical period of the double stranded DNA helix (Amit et al., 2011; Becker et al., 2005; Lee and Schleif, 1989; Müller et al., 1996). We find a maximum

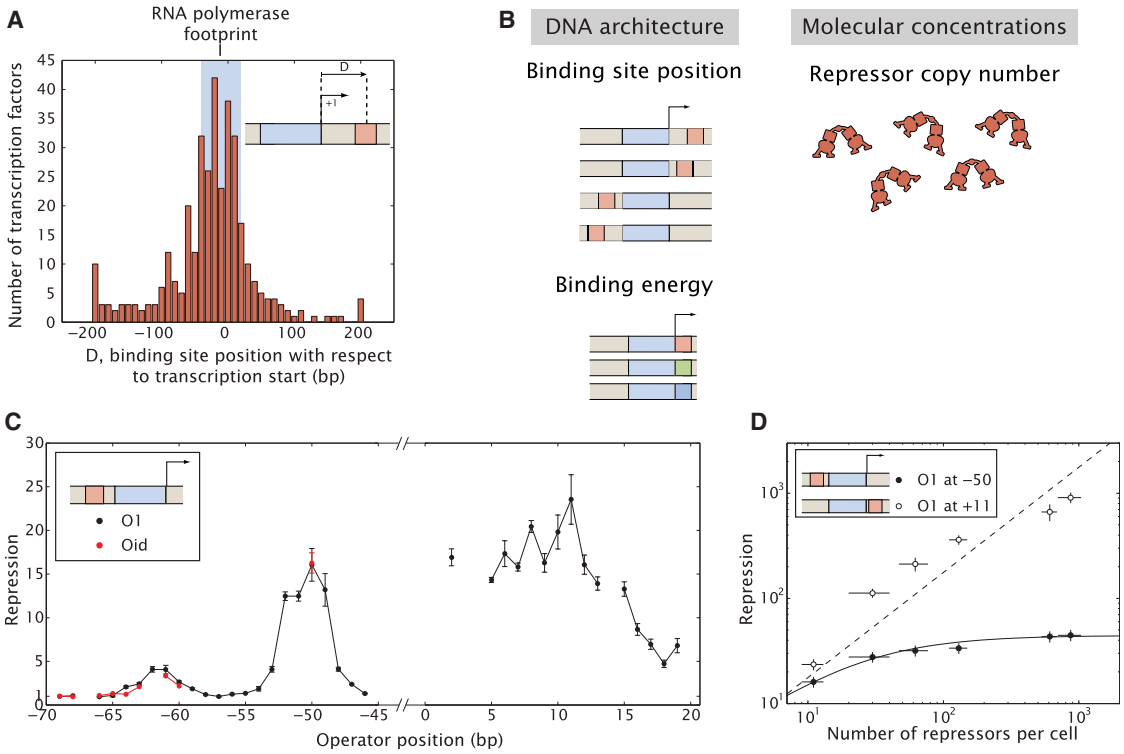

Figure 2. Operator Position Leads to Different Repression Behavior Upstream and Downstream from the Promoter

(A) Histogram of the positions of transcription factor binding with respect to the transcription start site of the regulated promoter (Gama-Castro et al., 2008). For reference, the DNase footprint of RNA polymerase in the open complex is overlaid in light blue (Straney and Crothers, 1987).

(B) Schematic of the parameters that may be tuned in vivo in order to elicit different levels of gene expression within the simple repression architecture.

(C) Repression as a function of operator position for an O1 and an Oid operator sequence.

(D) Unlike the case of repression for an operator located at +11, the input-output function of repression at −50 shows saturation. The lines are fits of the +11 data to Equation S6 and of the −50 data to Equation S7. The positions are defined as the distance between the center of the operator and the *lac*UV5 transcription start site. Error bars are SD over multiple repeats.

See also Figures S1 and S2.

repression level when the operator is centered at −50 base pairs upstream from the initiation site, with another smaller peak at −61. Our results are qualitatively consistent with previous studies on the effect of *lac* operator position on repression (Besse et al., 1986; Bond et al., 2010; Elledge and Davis, 1989). For a detailed comparison of previous results to our work please refer to the Extended Experimental Procedures and Figures S1D and S1E.

In order to better understand how operator location affects the input-output function, we measured repression as a function of the intracellular number of repressors over almost two orders of magnitude in the number of repressors (Garcia and Phillips, 2011) for two selected locations, one downstream at +11, the wild-type position of O1 in the *lac* promoter, and one at −50, where the peak of maximum upstream repression lies. As shown in Figure 2D, the two operator locations differ both qualitatively and quantitatively in the nature of their input-output function. For the +11 location, the repression factor grows linearly with the intracellular number of repressors. This behavior is expected for the repression mechanism based on blocking of open complex formation or of an earlier step in initiation, as discussed by Garcia and Phillips (2011); Sanchez et al. (2011); Vilar and Leibler (2003) and shown in Equation S6. A detailed description of this model in the context of simple repression for the constructs described above can be found in the Extended Experimental Procedures, Figure S3A, and Garcia and Phillips (2011).

In contrast, for the operator at −50 we found that the repression factor grows with repressor copy number only until it saturates. This result cannot be explained by a competition between Lac repressor and RNA polymerase, and suggests that RNA polymerase can bind to the promoter and initiate transcription even when Lac repressor is bound, although the overall transcription rate (i.e., the number of mRNA transcripts produced per unit of time) is reduced by the presence of the repressor to a low, basal level, about 40-fold less than the unregulated level. Thus, a direct prediction of this hypothesis is that Lac repressor does not completely inhibit the formation of stable RNA polymerase-promoter complexes when bound at −50.

It is important to note, however, that the data obtained with the operator at +11 could also accommodate a saturating behavior. In the Extended Experimental Procedures and

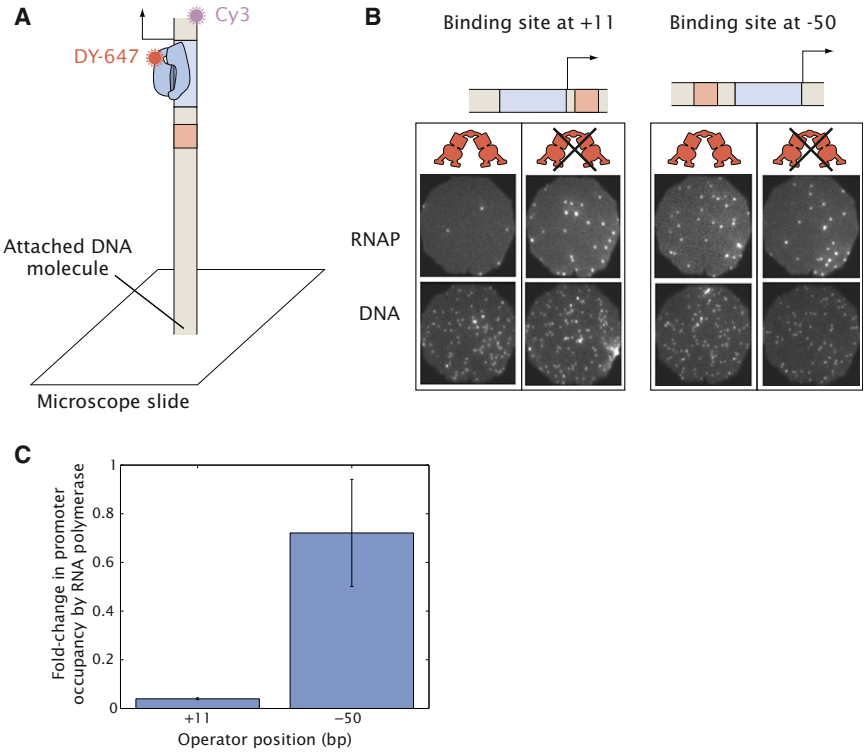

Figure 3. Upstream Repression through Lac Repressor-RNA Polymerase Ternary Complex Formation

(A) In vitro single molecule experiments in which the occupancy of RNA polymerase on the promoter DNA is quantified by fluorescently labeling each molecular species (Sanchez et al., 2011).

(B) Representative fields of view are shown where the occupancy of RNA polymerase on DNA is measured for the constructs with O1 located at +11 and −50 preincubated with and without Lac repressor.

(C) The analysis of several images such as those shown in (B) confirms the presence of a promoter bound RNA polymerase for the −50 construct because of the small change in RNA polymerase occupancy caused by the presence of repressor. In contrast, there is a significant fold-change in the promoter occupancy by RNA polymerase for the +11 construct, consistent with a model of inhibition of the open complex and previous work (Sanchez et al., 2011). More than 500 molecules were counted for each condition. Error bars are the SD over two replicates for each condition.

See also Figure S2.

Figure S2C we discuss this scenario in detail and conclude that, if that was the case, this would signal a violation of the occupancy hypothesis in repression at this well-studied operator location as well.

Single Molecule Imaging In Vitro Suggests that Lac Repressor Located at −50 and RNA Polymerase Are Bound Simultaneously

In order to test the hypothesis outlined above, and gain insight into the mechanism of repression when the *lac* operator is at −50, we performed single-molecule experiments where the occupancy of RNA polymerase on individual DNA molecules can be observed directly. Fluorescently labeled RNA polymerase and fluorescently labeled DNA were incubated together prior to adding heparin, which sequesters RNA polymerase molecules that have not formed an open complex. Finally, the reaction was introduced into a flow chamber yielding the arrangement shown in Figure 3A (Sanchez et al., 2011). We used multi-wavelength single molecule total internal reflection fluorescence (TIRF) microscopy to determine the fraction of DNA molecules tethered to the surface of the chamber that were occupied by RNA polymerase. A similar experiment was performed in which Lac repressor was preincubated with the DNA prior to the addition of RNA polymerase.

Representative fields of view for the experiment performed on both +11 and −50 constructs are shown in Figure 3B. We see that Lac repressor causes a significant change in DNA occupancy by RNA polymerase when the operator is located at +11, indicating that Lac repressor excludes the formation of stable RNA polymerase-DNA open complexes. In contrast, when the operator is located at −50 there is little change, which suggests that Lac repressor is not able to prevent formation of stable RNA polymerase-DNA open complexes when the repressor is bound at this location. At +11 the presence of stably bound polymerase on the DNA is not completely abolished by repressor due to the existence of nonpromoter polymerase binding site and presumably a similar effect occurs with our −50 constructs as described below (for details please see the Extended Experimental Procedures and Sanchez et al., 2011).

Many such fields of view were imaged for each construct. By counting the number of RNA polymerase-DNA complexes that form in the absence and the presence of Lac repressor (and correcting for the fraction of those events that correspond to RNA polymerase bound to a nonpromoter location) we can calculate the fold-change in promoter occupancy by polymerase induced by repressor. The results are shown in Figure 3C, and summarize the average occupancies obtained in different replicates of the experiment with different preparations of all the reagents. From Figure 3C we see again that Lac repressor bound at +11 largely inhibits RNA polymerase occupancy on the promoter DNA. In this construct repressor reduces the formation of RNA polymerase-promoter open complexes down to (4.0 ± 0.4)% of the number of complexes that form in the absence of Lac repressor. This reduction is consistent with recent measurements with the same promoter (Sanchez et al., 2011), which revealed that Lac repressor works by inhibiting open complex formation at the *lac*UV5 promoter, and indicates that under the conditions of our in vitro experiments, and for the concentrations of repressor (200 nM) and RNA polymerase (80 nM) we use, the O1 operator is almost saturated with repressor (~96%). By way of contrast, Lac repressor bound

at −50 reduces open complex formation only modestly, down to (72 ± 22)% of that in the absence of repressor.

These quantitative results indicate that RNA polymerase occupancy on the promoter is affected only slightly by repressor bound at −50. If Lac repressor at −50 reduces open complex formation by <2-fold in vitro, how can we observe a 40-fold reduction of gene expression in vivo? Because our results suggest that Lac repressor bound at −50 allows stable formation of open complexes by RNA polymerase at the promoter, they imply that the regulation of the level of gene expression comes from a substantial effect of repressor on steps occurring after open complex formation in the transcription initiation pathway.

In light of these results, we hypothesize that at −50 the repressor is directly affecting the overall rate of promoter escape, rather than just the occupancy of RNA polymerase on the promoter as an open complex. As a result we propose a thermodynamic model for in vivo upstream simple repression by Lac repressor that is schematized in Figure 4A and tested systematically in the following section.

The Input-Output Function of Simple Repression from Upstream Positions Involves Modulation of the Promoter Escape Rate by the Repressor

The general model for upstream repression proposed based on our experimental results and shown in Figure 4A covers three different mechanisms of regulation: (i) a direct, destabilizing interaction between RNA polymerase and Lac repressor that decreases occupancy of polymerase at the promoter when repressor is present, (ii) a direct, attractive interaction between RNA polymerase and Lac repressor in the closed and/or open complex that, by lowering the energy of the complex, effectively increases the amount of energy required for RNA polymerase to move forward on the pathway to transcription, and (iii) an increase in the activation energy for promoter escape, without any stabilization of RNA polymerase when Lac repressor is bound to the DNA. In the last case, Lac repressor does not affect the occupancy of the states, but only the kinetics of RNA polymerase escaping the promoter. These mechanisms are not mutually exclusive, but can act together to exert regulation depending on the values of the different parameters of the model. The different reaction diagrams corresponding to each one of these mechanisms are shown schematically in Figure 4B. In the following we explore these three cases through a quantitative comparison between theoretical predictions and expression data.

We start by considering mechanism (i). Qualitatively, this mechanism predicts a mutual destabilization between Lac repressor and RNA polymerase such that the occupancy of RNA polymerase on the promoter would be affected in the presence of repressor. However, our in vitro results shown in Figure 3C suggest that promoter occupancy is not affected significantly. We conclude that this effect, if present, will be of a small magnitude. As a result we do not consider this mechanism any further in this work. Further discussion of this point can be found in the Extended Experimental Procedures.

Next, we consider mechanism (ii), which leads to the following expression for the repression as a function of repressor copy number

$$\text{Repression}_{(ii)} = \frac{1+\frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}\,\xi}{1+\frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}} = 1+(\xi-1)\,\frac{\frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}}{1+\frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}},$$

(Equation 2)

where ξ is a function of the interaction parameter ε_{rp}, of the binding energy of polymerase to the promoter, and of the copy number of polymerases. Notice that the parameter ξ can only determine the maximum level of expression. However, it does not have an effect on the half-point, the repressor copy number at which the repression has reached half of its maximum value (this half-point is analogous to a dissociation constant, see Extended Experimental Procedures and Figures S3B and S3C for further details).

Finally, if we take mechanism (iii), where there is no stabilizing interaction between repressor and polymerase, but there is a change in the rate of promoter escape, we get the expression

$$\text{Repression}_{(iii)} = \frac{1+\frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}}{1+\frac{r_2}{r_1}\frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}} = 1+\left(1-\frac{r_2}{r_1}\right)\,\frac{\frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}}{1+\frac{r_2}{r_1}\frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}}.$$

(Equation 3)

This mechanism gives us a new parameter to consider: the ratio of the RNA polymerase escape rate in the presence of repressor to the rate in its absence, r_2/r_1, as shown in Figure 4A. However, unlike ξ in mechanism (ii), this parameter sets the value of both the half-point of the repression curve (notice the presence of r_2/r_1 in the denominator) as well as the maximum level of repression (see Extended Experimental Procedures and Figures S3B and S3C).

Continuing with the strategy employed in Figure 2D, we dissected simple repression upstream from the promoter in order to test the predictions of the different regulatory mechanisms posited by the model shown in Figure 4A. We created DNA constructs bearing all four *lac* operators (Oid, O1, O2, and O3, in order of high to low affinity) at −50 and we placed them in strains containing different intracellular numbers of Lac repressor that spanned nearly two orders of magnitude (Garcia and Phillips, 2011).

Figure 4C shows repression as a function of repressor number for O1 located at −50. As shown previously in Figure 2D, one of the surprising outcomes when comparing repression at −50 to repression at +11 is that repression at +11 grows with the number of repressors as called for by Equation S6 (see Figure S3A and Garcia and Phillips [2011]) whereas there is a saturation of repression at −50. This saturation is not consistent with the model embodied in Equation S6.

Given our previous knowledge of the in vivo binding energies of Lac repressor to the various operators (see Extended Experimental Procedures and Garcia and Phillips [2011]) the repression formulas for mechanisms (ii) and (iii) discussed above only have one free parameter each: ξ for mechanism (ii) and the ratio r_2/r_1 for mechanism (iii). In Figure 4C we show a fit of both mechanisms to our experimental data with O1 located at −50 shown in Figure 2D (for considerations on data fitting, please refer to

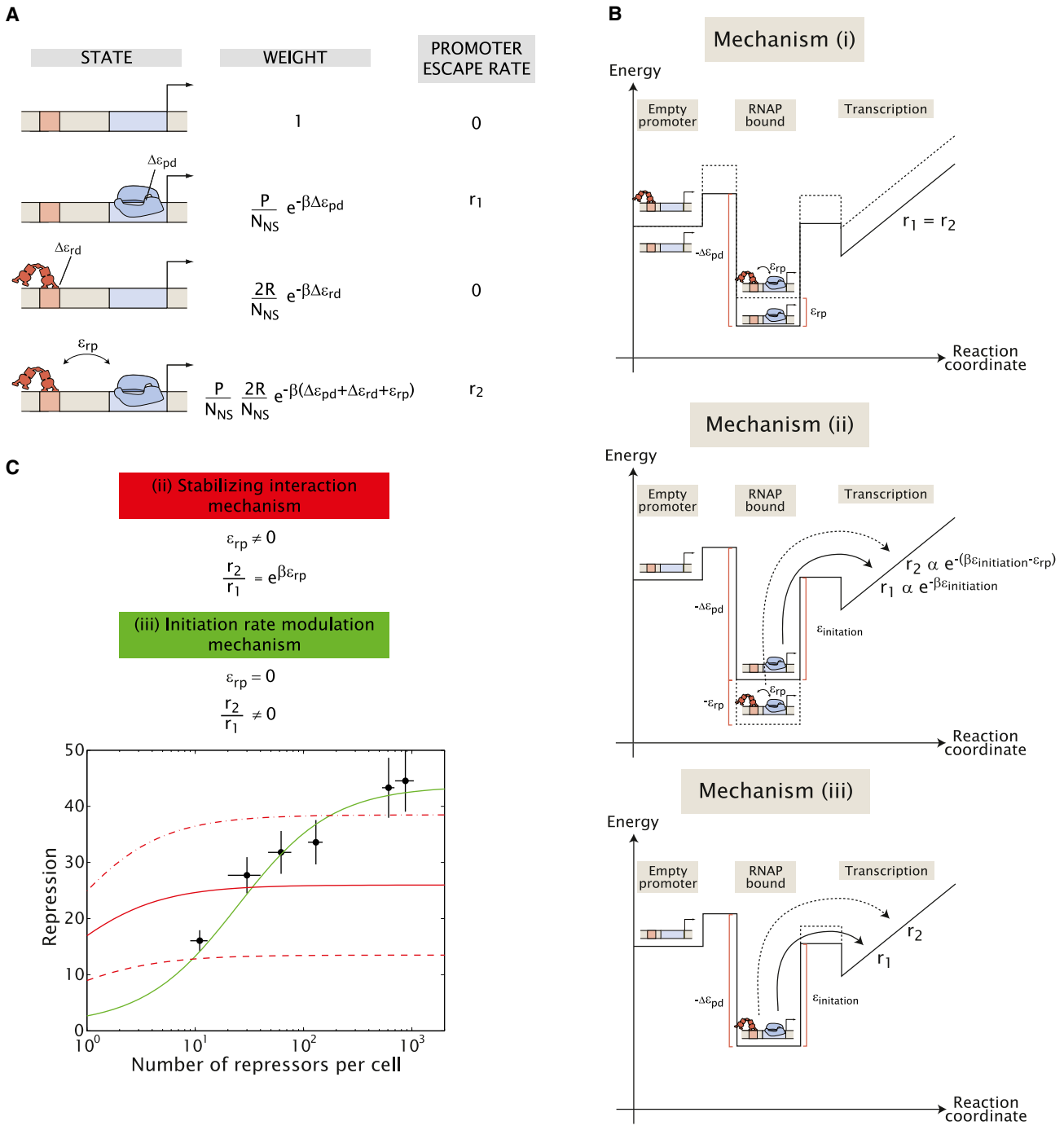

Figure 4. General Thermodynamic Model of Simple Repression for an Upstream Repressor Binding Site

(A) A thermodynamic model inspired by our in vitro results allows for an interaction between Lac repressor (shown in red) and RNA polymerase (shown in blue) with an energy ε_{rp}. Additionally, it allows for different rates of transcriptional initiation in the absence or presence of repressor. *P* and *R* are the number of RNA polymerase and Lac repressor molecules inside the cell, respectively, and $\Delta\varepsilon_{pd}$ and $\Delta\varepsilon_{rd}$ are their corresponding interaction energies with their specific sites measured with respect to their nonspecific binding energies. N_{NS} is the size of the bacterial genome in base pairs. $\beta = (k_B T)^{-1}$, with k_B the Boltzmann constant and T the absolute temperature. The rate of promoter escape is different in the absence (r_1) and presence (r_2) of Lac repressor. For a detailed description of this model, refer to the Extended Experimental Procedures.

(B) Simple energy diagrams for the possible mechanisms of transcriptional regulation: (i) destabilization of the RNA polymerase-promoter complex, (ii) stabilization of the RNA polymerase-promoter complex impeding promoter escape, and (iii) modulation of the rate of promoter escape.

(C) Repression as a function of intracellular number of repressors for O1 located at −50. A mechanism that exclusively considers stabilizing interactions [mechanism (ii)] cannot fit the data regardless of the value adopted by its only free parameter, ξ (see Equation 2). The solid red line is the best fit of such

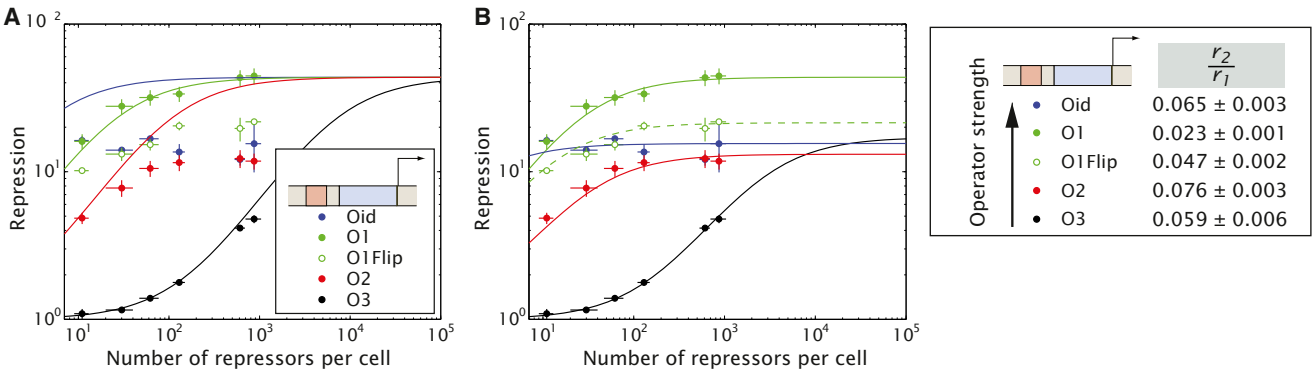

Figure 5. Repression at −50 by Lac Repressor as a Function of Repressor Copy Number and Operator Affinity Is Inconsistent with the Occupancy Hypothesis

(A) Given the modulation in transcription rate obtained from the fit to the O1 data shown in Figure 4C and our previous knowledge of the in vivo binding energies we predict the shape of the input-output function for constructs where the different operators are used (lines). The measured repression (circles) is systematically inconsistent with the model predictions.

(B) By letting the modulation in transcription rate (r_2/r_1) be different for each operator we can successfully fit the data. Notice that the modulation does not bear a correlation with operator strength.

O1Flip is the reverse complement of O1 and the dashed green line corresponds to a fit to its resulting experimental data. Error bars are SD over multiple repeats. See also Figures S4, S5, and S6.

the Extended Experimental Procedures). As indicated by the various red lines in that figure, mechanism (ii) shown in Equation 2 produces curves of the wrong shape and thus cannot fit the data regardless of the choice of parameter ξ. On the other hand, mechanism (iii), which leads to Equation 3, can fit the data as shown by the green line in Figure 4C.

Based on the analysis above we propose that the main mode of regulation by repressor is the modulation of promoter escape rate by RNA polymerase. This does not rule out a contribution from a stabilizing interaction between repressor and polymerase. In fact, a combination of both regulatory strategies can also fit the data as shown in Figure S3B. However, regulation of the escape rate constitutes a minimal mechanism that is sufficient to explain the data. We will assume this mechanism to further explore repression when the operator is located at −50.

Operator Strength Is Not Sufficient to Explain the Level of Repression

Given our knowledge of the modulation of the escape rate by Lac repressor obtained from the O1 data, we predict the shapes of the input-output functions for the remaining *lac* operators in Figure 5A. Under the occupancy hypothesis the model corresponding to this mechanism (mechanism (iii), shown in Equation 3) predicts that repression saturates at the same level regardless of the choice of operator because different operator sequences only change the affinity of Lac repressor to operator DNA. However, the operator choice determines the half-point of repression in a way that follows a clear rank-ordering based on the repressor binding affinity of the various operators considered.

In Figure 5A we also show the corresponding experimental data. It is clear from this plot that the model cannot describe the data. In particular, it is both intriguing and surprising that the data for different operators saturates at different levels and that this saturation does not follow the rank ordering of the in vivo and in vitro binding affinity of the operators. For example, Lac repressor binds to Oid ∼20 times more strongly than O2, with the K_d for Oid at ∼170 pM and the K_d for O2 at ∼4 nM. Yet, these two operators have a comparable level of repression at a high number of repressors of ∼900. On the other hand, Oid is also bound ∼5 times stronger than O1, with O1 having a K_d of ∼1 nM. Still, O1 presents a higher level of repression than Oid at the same intracellular number of repressors of 900. Perhaps even more interesting, if we replace the O1 binding site by its reverse complement, which should leave its binding affinity unaltered, we see a qualitatively different behavior from wild-type O1 suggesting that binding affinity alone is not sufficient to determine the different saturation levels.

If we abandon the view that the only role of the operator sequence is to set the binding affinity of Lac repressor to DNA, and adopt a view where it can modulate the transcription initiation rate in a sequence-dependent way (with a different choice of the parameter r_2/r_1) our model can account for all of the experimental data. For example, the choice of operator might modulate the nature of the interaction between repressor and RNA polymerase. Figure 5B shows that when we allow the parameter r_2/r_1 to change with operator sequence the model now accounts for the experimental data. Thus, the observed difference in modulation of initiation rate for the different constructs is at odds with the interpretation that the role of binding sites is exclusively to

a mechanism to the data whereas the dashed and dashed-dotted lines correspond to varying that parameter by a factor of two. On the other hand, a mechanism that just considers a modulation in the promoter escape rate [mechanism (iii)], also having only one free parameter, can fit the data. Details of these two mechanisms are given in the main text and the Extended Experimental Procedures. Error bars are SD over multiple repeats.
See also Figure S3.

determine the probability of finding the repressor bound to the DNA, but is consistent with models where operator sequence can alter the nature of the repressor-polymerase interaction in a way that modulates the polymerase escape rate.

An alternative hypothesis is that the modification of the operator sequence leads to a change in the unregulated level of gene expression. In this case the differences in the observed r_2/r_1 ratios could be purely due to a change in r_1 for each operator. In Figure S4 we show that there is no significant correlation between the unregulated levels of expression and the fitted r_2/r_1 values. We conclude that the observed effect of operator sequences cannot be explained by the change in the unregulated levels of expression.

An alternative way to examine the effect of operator sequence on the level of repression is to replot the data for repression as a function of operator occupancy. As described in the introduction, the occupancy hypothesis implies that all data should fall on the same curve, as shown in Figure 1A. In Figure S5 and the Extended Experimental Procedures we show that although the data for the +11 constructs collapses as expected from Figure 1A, the data corresponding to the −50 constructs does not, suggesting again that repressor occupancy is not sufficient to determine the level of repression.

The model used so far represents a simplified view of transcription initiation that combines both closed and open complexes into one effective complex. However, the exact same conclusions, without any loss of generality, can be reached when both complexes are considered independently (see Extended Experimental Procedures and Figure S6). Furthermore, the thermodynamic model used assumes quasi-equilibrium between states leading up to promoter escape. If we consider a full kinetic model in which no assumptions about equilibrium are made we nevertheless reach the same conclusions (see Extended Experimental Procedures): the different levels of repression observed for different operator sequences placed at the −50 location cannot be explained in the context of the occupancy hypothesis.

DISCUSSION

The occupancy hypothesis states that the role of operator sequence is to determine its occupancy by its target transcription factor. The nature of the interaction between the bound transcription factor and the transcriptional machinery is then determined by the spatial arrangements of binding sites and the DNA sequence context, i.e., the presence of DNA binding sites for other proteins in the vicinity, the particular mechanical properties of the surrounding DNA, etc. (Davidson, 2006; Ptashne and Gann, 2002). For example, the relative positioning between binding sites and the mechanical properties of the intervening DNA can have drastic effects on gene regulatory input-output functions (Aki et al., 1996; Amit et al., 2011; Belyaeva et al., 1998; Browning and Busby, 2004; Busby et al., 1994; Choy et al., 1995, 1997; Gaston et al., 1990; Hogan and Austin, 1987; Joung et al., 1994; Joung et al., 1993; Lilja et al., 2004; Mao et al., 1994; Ryu et al., 1998). Additionally, the nature of the promoter can modulate how a transcription factor will interact with its bound RNA polymerase (Monsalve et al., 1996, 1997). The majority of the current models of action of the diverse known interactions between transcription factors and the transcriptional machinery are based on assuming the applicability of the occupancy hypothesis (Bintu et al., 2005b; Buchler et al., 2003; Segal and Widom, 2009; Vilar and Leibler, 2003).

Several recent works have suggested that this canonical picture of transcriptional regulation is incomplete (see Haugen et al., 2008 and Voss et al., 2011 for two specific examples). More directly related to this work, the occupancy hypothesis has been suggested to be insufficient to describe regulation by MarA in bacteria (Martin et al., 2008; Wall et al., 2009) and examples where the occupancy hypothesis falls short have been found in the context of the regulation of cofactors by transcription factors in eukaryotes, as we describe below (Meijsing et al., 2009; Scully et al., 2000).

In this study, we quantitatively expanded our understanding of the paradigmatic Lac repressor and showed that the sequence of an operator located upstream from the promoter can dictate different gene regulatory input-output functions leading to different maximum repression values that cannot be explained by the occupancy of repressor on DNA. We used theoretical models of transcriptional regulation in order to qualitatively and quantitatively frame these conclusions. Whether thermodynamic models are used or a kinetic one in which no equilibrium assumption is invoked (see Figure S6), the conclusions are independent of the particular theoretical framework used to analyze the experimental results. As a result, in clear violation of the occupancy hypothesis, we conclude that the *lac* operator sequences encode more than just repressor binding affinity: they can also determine the nature of the "effective" interaction between repressor and RNA polymerase. We emphasize the word "effective" to make clear that our model cannot determine if the effect is due to a direct contact between repressor and RNA polymerase, due to information being transferred through the DNA in some "allosteric" way or due to some other, unknown mechanism.

What is the mechanistic nature of the effective interaction between Lac repressor and RNA polymerase? The −50 position, where we carried out our most detailed characterization of upstream repression, is within the footprint of the RNA polymerase alpha C-terminal domain (αCTD) subunit (Newlands et al., 1991). This suggests that αCTD might be involved in the repression mechanism through a direct contact with the repressor in a fashion analogous to class I activators (Busby and Ebright, 1999). A prediction of this direct contact hypothesis is that if we introduce mutations or deletions in αCTD the repression should be abolished. By way of contrast, in the allosteric hypothesis, such mutations should have little effect because the repression is mediated by binding to the DNA, not by protein-protein contacts. Previous experiments by Adhya and co-workers shed light on this issue (Choy et al., 1995, 1997; Roy et al., 2004). They found that Lac repressor (and also Gal repressor) bound at an operator at −60 (the position of the secondary peak of repression in Figure 2C) represses transcription of the *galP1* promoter, and that deletion of the αCTD completely alleviates repression at −60 (Choy et al., 1995). In addition, mutations in the αCTD also abolished repression (Choy et al., 1997). Both of these experiments support the direct

contact hypothesis. Furthermore, a mutant with a single point mutation in GalR was found to be able to bind to the operator at −60, but not to repress transcription (Roy et al., 2004). It is worth noting that all of these experiments were done for a different promoter than the one we have characterized here. However, their results support a mechanism based on direct contact between repressor and RNA polymerase.

None of the different mechanistic hypotheses discussed above can explain why different operator sequences determine the level of repression in a way that does not correlate with operator occupancy by repressor, which results in a violation of the occupancy hypothesis. One possible explanation is that these different regulatory outcomes result from subtle differences in the three-dimensional structures of the protein-DNA complexes or in the dynamics of these molecules. These differences could lead to altered interactions with RNA polymerase or the promoter region and result in the modulation of gene expression. In fact, differences in structure have been observed for the Lac repressor binding domain bound to its different operators as well as for the structural parameters of the intervening DNA such as twist, roll and base pair stacking, but their correlation with any phenotypic effects is unclear (Kalodimos et al., 2002, 2004; Romanuka et al., 2009). It is then also possible to speculate that the information about which operator is present is transferred through the DNA itself. However, because these studies resolved only the DNA binding-domain it remains unclear whether the conformation of the remaining protein was altered in any relevant way. Despite uncertainties about the detailed sequence-dependent molecular mechanism, the work reported here is a further step toward a more detailed understanding of the molecular interactions exerted by transcription factors.

A few studies in eukaryotic cells had previously found that DNA may act as more than simply a docking site for transcription factors; in addition, it may act as an allosteric ligand that conveys information about the mode of gene regulation (Geserick et al., 2005; Ma et al., 2010). These studies found that the specific sequence of a transcription factor binding site determined the affinity of the bound transcription factor for a different set of corepressors or coactivators. These changes in affinity may have profound physiological effects as has been suggested for the Pit-1 factor, the glucocorticoid receptor, and NF-κB (Lefstin and Yamamoto, 1998; Leung et al., 2004; Meijsing et al., 2009; Scully et al., 2000).

Our study demonstrates that modulation of transcription factor activity by the DNA sequence of its binding site may well be a much more general phenomenon, occurring as shown here in bacteria as well as in eukaryotes, despite the differences between transcriptional mechanisms in these two domains of life. Our study was performed in *E. coli*, where transcriptional regulation is thought to be much simpler than in eukaryotes, and we used a promoter that does not involve any cofactors. This simplicity has allowed us to find a direct mechanistic link between the DNA sequence of an operator and the transcriptional output. These results suggest that a similar effect of operator sequence on the modulation of promoter escape could arise in other bacterial transcription factors that either halt or enhance transcription at the same step as has recently been suggested for activation by MarA (Martin et al., 2008; Wall et al., 2009). Thus, MarA regulated promoters may be good candidate systems to further investigate the generality of our findings in bacterial gene regulation.

Much work that has focused on the dissection of gene regulatory regions has been based on assuming the applicability of the occupancy hypothesis (Ackers et al., 1982; Amit et al., 2011; Bintu et al., 2005a; Davidson, 2006; Gertz et al., 2009; Ptashne and Gann, 2002; Raveh-Sadka et al., 2012; Segal et al., 2008; Segal and Widom, 2009; Zinzen et al., 2009). This study gives further evidence for the existence of an additional layer of complexity to consider in transcriptional regulation given by the nature of the interaction between transcription factors and the transcriptional machinery imposed by transcription factor binding site sequence. Given the fact that a large number of repressors act on promoters by binding to a single site located upstream from the promoter region in *E. coli* (see Figure 2A and Gama-Castro et al., 2008) it is possible that this mechanism of repression might be widespread. Thus, knowing the list of operators and their strengths is not sufficient to predict the input-output function of a promoter. A detailed analysis of specific repressors will be necessary to determine how widespread the effects observed here may be.

EXPERIMENTAL PROCEDURES

Plasmids, Strains, and In Vivo Measurements

The construction of all plasmids and strains is described in the Extended Experimental Procedures. In short, we placed a YFP or *lacZ* reporter gene under the control of a *lac*UV5 promoter and the regulation of one of four *lac* operators at different positions with respect to the transcription start site. The different constructs used throughout the paper are shown schematically in Figures S1A–S1C. These constructs were integrated in the chromosome of *E. coli* strains bearing different intracellular numbers of Lac repressor (Garcia and Phillips, 2011).

Gene expression measurements were performed using a plate reader as described in the Extended Experimental Procedures and (Garcia et al., 2011).

Single Molecule Experiments

Single molecule experiments were performed as described in the Extended Experimental Procedures and (Sanchez et al., 2011). In short, fluorescently- and biotin-labeled DNA containing a promoter and a repressor binding site was incubated in the presence of RNA polymerase labeled with a second, spectrally distinct fluorophore. The DNA molecules were bound to a streptavidin coated glass slide and the fraction of RNA polymerase-bound DNA molecules was quantified. In order to assay the effect of repressor on the formation of RNA polymerase-bound complexes the DNA was pre-incubated with repressor before the addition of RNA polymerase and the resulting reaction was again imaged. As reported previously (Sanchez et al., 2011), not all RNA polymerase-bound DNA molecules correspond to stable, open complexes. This was taken into account in our analysis. Details pertaining to this point can be found in the Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and six figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.06.004.

LICENSING INFORMATION

ACKNOWLEDGMENTS

We are grateful to Jon Widom, Tom Kuhlman, Justin Kinney, Stephanie Johnson, Daniel Jones and Rob Brewster for helpful discussions. We dedicate this work to Prof. Widom who recently passed away unexpectedly. We would like to thank Larry Friedman for technical assistance with the single molecule experiments and for useful discussions. We thank Robert Landick, Rachel Mooney and Abbey Vangeloff for the generous gifts of purified SNAP-tagged core RNA polymerase and σ^{70}. This work was supported by National Institutes of Health Pioneer award DP1 OD000217 (H.G.G., R.P.) and grants R01 GM085286 and R01 GM085286-01S (H.G.G., J.Q.B., R.P.), GM81648 and GM43369 (J.G., A.S., M.L.O.), La Fondation Pierre Gilles de Gennes (R.P.), and National Science Foundation award DMR-0706458 (J.K.) and MRSEC-0820492 (J.K., J.G.).

Received: April 9, 2012
Revised: May 14, 2012
Accepted: June 6, 2012
Published online: July 12, 2012

REFERENCES

Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. Proc. Natl. Acad. Sci. USA *79*, 1129–1133.

Aki, T., Choy, H.E., and Adhya, S. (1996). Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by GAL repressor. Genes Cells *1*, 179–188.

Alberts, B. (2008). Molecular Biology of the Cell, Fifth Edition (New York: Garland Science).

Amit, R., Garcia, H.G., Phillips, R., and Fraser, S.E. (2011). Building enhancers from the ground up: a synthetic biology approach. Cell *146*, 105–118.

Becker, N.A., Kahn, J.D., and Maher, L.J., 3rd. (2005). Bacterial repression loops require enhanced DNA flexibility. J. Mol. Biol. *349*, 716–730.

Belyaeva, T.A., Rhodius, V.A., Webster, C.L., and Busby, S.J. (1998). Transcription activation at promoters carrying tandem DNA sites for the *Escherichia coli* cyclic AMP receptor protein: organisation of the RNA polymerase alpha subunits. J. Mol. Biol. *277*, 789–804.

Besse, M., von Wilcken-Bergmann, B., and Müller-Hill, B. (1986). Synthetic lac operator mediates repression through lac repressor when introduced upstream and downstream from lac promoter. EMBO J. *5*, 1377–1381.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005a). Transcriptional regulation by the numbers: applications. Curr. Opin. Genet. Dev. *15*, 125–135.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005b). Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. *15*, 116–124.

Bond, L.M., Peters, J.P., Becker, N.A., Kahn, J.D., and Maher, L.J., 3rd. (2010). Gene repression by minimal lac loops in vivo. Nucleic Acids Res. *38*, 8072–8082.

Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. *2*, 57–65.

Buchler, N.E., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. Proc. Natl. Acad. Sci. USA *100*, 5136–5141.

Busby, S., and Ebright, R.H. (1999). Transcription activation by catabolite activator protein (CAP). J. Mol. Biol. *293*, 199–213.

Busby, S., West, D., Lawes, M., Webster, C., Ishihama, A., and Kolb, A. (1994). Transcription activation by the Escherichia coli cyclic AMP receptor protein. Receptors bound in tandem at promoters can interact synergistically. J. Mol. Biol. *241*, 341–352.

Cağatay, T., Turcotte, M., Elowitz, M.B., Garcia-Ojalvo, J., and Süel, G.M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell *139*, 512–522.

Choy, H.E., Park, S.W., Aki, T., Parrack, P., Fujita, N., Ishihama, A., and Adhya, S. (1995). Repression and activation of transcription by Gal and Lac repressors: involvement of alpha subunit of RNA polymerase. EMBO J. *14*, 4523–4529.

Choy, H.E., Hanger, R.R., Aki, T., Mahoney, M., Murakami, K., Ishihama, A., and Adhya, S. (1997). Repression and activation of promoter-bound RNA polymerase activity by Gal repressor. J. Mol. Biol. *272*, 293–300.

Davidson, E.H. (2006). The Regulatory Genome: Gene Regulatory Networks in Development and Evolution (Burlington, MA; San Diego, CA: Academic).

Elledge, S.J., and Davis, R.W. (1989). Position and density effects on repression by stationary and mobile DNA-binding proteins. Genes Dev. *3*, 185–197.

Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335–338.

Fowlkes, C.C., Hendriks, C.L., Keränen, S.V., Weber, G.H., Rübel, O., Huang, M.Y., Chatoor, S., DePace, A.H., Simirenko, L., Henriquez, C., et al. (2008). A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell *133*, 364–374.

Gama-Castro, S., Jiménez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spinola, M.I., Contreras-Moreira, B., Segura-Salazar, J., Muñiz-Rascado, L., Martínez-Flores, I., Salgado, H., 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.

Garcia, H.G., and Phillips, R. (2011). Quantitative dissection of the simple repression input-output function. Proc. Natl. Acad. Sci. USA *108*, 12173–12178.

Garcia, H.G., Lee, H.J., Boedicker, J.Q., and Phillips, R. (2011). Comparison and calibration of different reporters for quantitative analysis of gene expression. Biophys. J. *101*, 535–544.

Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in Escherichia coli. Nature *403*, 339–342.

Gaston, K., Bell, A., Kolb, A., Buc, H., and Busby, S. (1990). Stringent spacing requirements for transcription activation by CRP. Cell *62*, 733–743.

Gertz, J., Siggia, E.D., and Cohen, B.A. (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature *457*, 215–218.

Geserick, C., Meyer, H.A., and Haendler, B. (2005). The role of DNA response elements as allosteric modulators of steroid receptor function. Mol. Cell. Endocrinol. *236*, 1–7.

Haugen, S.P., Ross, W., and Gourse, R.L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat. Rev. Microbiol. *6*, 507–519.

Hochschild, A., and Dove, S.L. (1998). Protein-protein contacts that activate and repress prokaryotic transcription. Cell *92*, 597–600.

Hogan, M.E., and Austin, R.H. (1987). Importance of DNA stiffness in protein-DNA binding specificity. Nature *329*, 263–266.

Joung, J.K., Le, L.U., and Hochschild, A. (1993). Synergistic activation of transcription by Escherichia coli cAMP receptor protein. Proc. Natl. Acad. Sci. USA *90*, 3083–3087.

Joung, J.K., Koepp, D.M., and Hochschild, A. (1994). Synergistic activation of transcription by bacteriophage lambda cI protein and E. coli cAMP receptor protein. Science *265*, 1863–1866.

Kalodimos, C.G., Bonvin, A.M., Salinas, R.K., Wechselberger, R., Boelens, R., and Kaptein, R. (2002). Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator 01 through alternative conformations of its DNA-binding domain. EMBO J. *21*, 2866–2876.

Kalodimos, C.G., Boelens, R., and Kaptein, R. (2004). Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the Lac repressor system. Chem. Rev. *104*, 3567–3586.

Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. Cell *139*, 1366–1375.

Kuhlman, T., Zhang, Z., Saier, M.H., Jr., and Hwa, T. (2007). Combinatorial transcriptional control of the lactose operon of Escherichia coli. Proc. Natl. Acad. Sci. USA *104*, 6043–6048.

Lee, D.H., and Schleif, R.F. (1989). *In vivo* DNA loops in araCBAD: size limits and helical repeat. Proc. Natl. Acad. Sci. USA *86*, 476–480.

Lefstin, J.A., and Yamamoto, K.R. (1998). Allosteric effects of DNA on transcriptional regulators. Nature *392*, 885–888.

Leung, T.H., Hoffmann, A., and Baltimore, D. (2004). One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. Cell *118*, 453–464.

Lilja, A.E., Jenssen, J.R., and Kahn, J.D. (2004). Geometric and dynamic requirements for DNA looping, wrapping and unwrapping in the activation of *E. coli* glnAp2 transcription by NtrC. J. Mol. Biol. *342*, 467–478.

Lopez, P.J., Guillerez, J., Sousa, R., and Dreyfus, M. (1998). On the mechanism of inhibition of phage T7 RNA polymerase by lac repressor. J. Mol. Biol. *276*, 861–875.

Ma, B., Tsai, C.J., Pan, Y., and Nussinov, R. (2010). Why does binding of proteins to DNA or proteins to proteins not necessarily spell function? ACS Chem. Biol. *5*, 265–272.

Madan Babu, M., and Teichmann, S.A. (2003). Functional determinants of transcription factors in Escherichia coli: protein families and binding sites. Trends Genet. *19*, 75–79.

Mao, C., Carlson, N.G., and Little, J.W. (1994). Cooperative DNA-protein interactions. Effects of changing the spacing between adjacent binding sites. J. Mol. Biol. *235*, 532–544.

Martin, R.G., Bartlett, E.S., Rosner, J.L., and Wall, M.E. (2008). Activation of the Escherichia coli marA/soxS/rob regulon in response to transcriptional activator concentration. J. Mol. Biol. *380*, 278–284.

Meijsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L., and Yamamoto, K.R. (2009). DNA binding site sequence directs glucocorticoid receptor structure and activity. Science *324*, 407–410.

Monsalve, M., Calles, B., Mencía, M., Salas, M., and Rojo, F. (1997). Transcription activation or repression by phage psi 29 protein p4 depends on the strength of the RNA polymerase-promoter interactions. Mol. Cell *1*, 99–107.

Monsalve, M., Mencía, M., Salas, M., and Rojo, F. (1996). Protein p4 represses phage phi 29 A2c promoter by interacting with the alpha subunit of Bacillus subtilis RNA polymerase. Proc. Natl. Acad. Sci. USA *93*, 8913–8918.

Müller-Hill, B. (1996). The *lac* Operon: A Short History of a Genetic Paradigm (Berlin, New York: Walter de Gruyter).

Müller, J., Oehler, S., and Müller-Hill, B. (1996). Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator. J. Mol. Biol. *257*, 21–29.

Newlands, J.T., Ross, W., Gosink, K.K., and Gourse, R.L. (1991). Factor-independent activation of Escherichia coli rRNA transcription. II. characterization of complexes of rrnB P1 promoters containing or lacking the upstream activator region with Escherichia coli RNA polymerase. J. Mol. Biol. *220*, 569–583.

Novák, B., and Tyson, J.J. (2008). Design principles of biochemical oscillators. Nat. Rev. Mol. Cell Biol. *9*, 981–991.

Pavco, P.A., and Steege, D.A. (1990). Elongation by Escherichia coli RNA polymerase is blocked in vitro by a site-specific DNA binding protein. J. Biol. Chem. *265*, 9960–9969.

Pavco, P.A., and Steege, D.A. (1991). Characterization of elongating T7 and SP6 RNA polymerases and their response to a roadblock generated by a site-specific DNA binding protein. Nucleic Acids Res. *19*, 4639–4646.

Ptashne, M., and Gann, A. (2002). Genes and Signals (New York: Cold Spring Harbor Laboratory Press).

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012). Tuning expression in a predictable manner by manipulating nucleosomes disfavoring sequences. Nat. Rev. Genet. http://dx.doi.org/10.1038/ng.2305

Rojo, F. (1999). Repression of transcription initiation in bacteria. J. Bacteriol. *181*, 2987–2991.

Romanuka, J., Folkers, G.E., Biris, N., Tishchenko, E., Wienk, H., Bonvin, A.M., Kaptein, R., and Boelens, R. (2009). Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein-DNA complexes. J. Mol. Biol. *390*, 478–489.

Roy, S., Semsey, S., Liu, M., Gussin, G.N., and Adhya, S. (2004). GalR represses galP1 by inhibiting the rate-determining open complex formation through RNA polymerase contact: a GalR negative control mutant. J. Mol. Biol. *344*, 609–618.

Ryu, S., Fujita, N., Ishihama, A., and Adhya, S. (1998). GalR-mediated repression and activation of hybrid lacUV5 promoter: differential contacts with RNA polymerase. Gene *223*, 235–245.

Sanchez, A., Osborne, M.L., Friedman, L.J., Kondev, J., and Gelles, J. (2011). Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. EMBO J. *30*, 3940–3946.

Schlax, P.J., Capp, M.W., and Record, M.T., Jr. (1995). Inhibition of transcription initiation by lac repressor. J. Mol. Biol. *245*, 331–350.

Scully, K.M., Jacobson, E.M., Jepsen, K., Lunyak, V., Viadiu, H., Carrière, C., Rose, D.W., Hooshmand, F., Aggarwal, A.K., and Rosenfeld, M.G. (2000). Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. Science *290*, 1127–1131.

Segal, E., and Widom, J. (2009). From DNA sequence to transcriptional behaviour: a quantitative approach. Nat. Rev. Genet. *10*, 443–456.

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature *451*, 535–540.

Straney, S.B., and Crothers, D.M. (1987). Lac repressor is a transient gene-activating protein. Cell *51*, 699–707.

Tsai, T.Y., Choi, Y.S., Ma, W., Pomerening, J.R., Tang, C., and Ferrell, J.E., Jr. (2008). Robust, tunable biological oscillations from interlinked positive and negative feedback loops. Science *321*, 126–129.

Vilar, J.M., and Leibler, S. (2003). DNA looping and physical constraints on transcription regulation. J. Mol. Biol. *331*, 981–989.

Voss, T.C., Schiltz, R.L., Sung, M.H., Yen, P.M., Stamatoyannopoulos, J.A., Biddie, S.C., Johnson, T.A., Miranda, T.B., John, S., and Hager, G.L. (2011). Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. Cell *146*, 544–554.

Wall, M.E., Markowitz, D.A., Rosner, J.L., and Martin, R.G. (2009). Model of transcriptional activation by MarA in Escherichia coli. PLoS Comput. Biol. *5*, e1000614.

Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature *462*, 65–70.

Cell PRESS
Open ACCESS

EXTENDED EXPERIMENTAL PROCEDURES

Plasmids

Plasmids pZS25O1+11-YFP, pZS25O1+11-lacZ and their equivalents with the different Lac repressor binding sites have been described previously (Garcia et al., 2011; Garcia and Phillips, 2011). In short, these plasmids are based on the pZ system (Lutz and Bujard, 1997). They contain the *lac*UV5 promoter cloned between unique restriction sites, a ribosomal binding site (RBS) and the YFP or LacZ reporters. The *lac*UV5 promoter has a unique binding site for Lac repressor centered at position +11 with respect to the transcription start site. These plasmids are shown schematically in Figure S1A.

In order to create constructs where the position of the main operator was varied downstream from the promoter, site directed mutagenesis was used. The sequences corresponding to each one of these constructs is shown in Figure S1B. The promoter corresponding to construct O1-45 was assembled from oligonucleotides. This construct was used as the starting point for the site directed mutagenesis used to move the binding site further upstream leading to the sequences shown in Figure S1C.

To move O1 upstream the corresponding upstream sequence of the *lac* promoter was used to generate the spacer up to the CRP binding site. From there the random sequence E8 was used (Cloutier and Widom, 2004, 2005). To move O1 downstream the sequence E8 was used as a spacer. For a given position of the operator its sequence was changed to O2, O3, Oid and O1Flip (see Figure S1A) using site directed mutagenesis.

A construct where YFP was deleted was used to measure the cell autofluorescence. All constructs were verified by sequencing. Constructs and sequences are available upon request.

Strains

As mentioned in the text, repression is calculated by measuring the level of gene expression in a strain bearing the construct of interest in the absence of Lac repressor and dividing this level by the gene expression of the same construct in a strain containing Lac repressor. The strain bearing no Lac repressor is TK140 (Kuhlman et al., 2007), which is MG1655 with a deletion of the *lacI* gene. For wild-type levels of Lac repressor the strain MG1655 was used. This strain bears the same amount of Lac repressor as strain HG104, which has a deletion of the *lacZYA* operon, but leaving the *lacI* gene intact (Garcia and Phillips, 2011).

Different strains with synthetic constructs expressing several levels of Lac repressor were used as described in Garcia and Phillips (2011). The reporter constructs were integrated into the *galK* gene of one of our strains and then moved into the other strains using P1 transduction (http://openwetware.org/wiki/Sauer:P1vir_phage_transduction).

Growth Conditions and Gene Expression Measurements

Samples were grown as described in Garcia and Phillips (2011) and Garcia et al. (2011). Cells were grown overnight in LB and with the appropriate antibiotics. They were then diluted in triplicate by about 1:40,000 into M9 + 0.5% glucose minimal media and grown for about 10 generations to an OD_{600} of 0.3 to 0.5.

Gene expression measurements were performed using a Tecan Saphire2 plate reader as described in (Garcia et al., 2011).

In Vitro Experiments

Single Molecule Experiments

A detailed description of the single molecule and transcription experiments can be found in (Sanchez et al., 2011b). Briefly, 289 bp linear DNA with biotin at one end and Cy3 dye at the other end was prepared by PCR (primers: 5′-CCT TCA CCC TCT CCA CTG ACA G-3′ and 5′-ATA GGC GTA TCA CGA GGC CC-3′) from plasmids pZS25O1-50 (for the O1-50 construct) and pzS25O1+11 (for the O1 centered at +11 construct). DY-647 labeled *E. coli* σ^{70} RNA polymerase holoenzyme was prepared from SNAP-surface 647 (New England Biolabs, Ipswich, MA), SNAP-tagged core $\alpha_2\beta\beta'\omega$ RNA polymerase, and σ^{70} (provided by Robert Landick, Rachel Mooney and Abbey Vangeloff, University of Wisconsin) as described in (Sanchez et al., 2011b). In order to form open complexes 8 nM DNA and 80 nM labeled holoenzyme were incubated together for 25 min at 37°C. Then, they were diluted 100-fold into buffer containing 16 μg/ml heparin and an O_2-scavenging system. The resulting mixture was flushed into a flow chamber that was pre-coated with biotinylated polyethyleneglycol and streptavidin. The resulting heparin- resistant complexes were imaged by single molecule total internal reflection fluorescence microscopy. In control samples that lacked DNA, binding of RNA polymerase to the flow chamber surface was negligible: an average of < 1 RNA polymerase per field of view (area = 315 μm^2) was detected. Cy3 and DY-647 were imaged separately, by alternating the excitation lasers and by optically separating the emission into images less than and greater than 635 nm wavelength (see (Friedman et al., 2006) for details). The fraction of DNA molecules bearing at least one bound RNA polymerase molecule was determined by directly counting the number of Cy3 (DNA) and DY-647 (RNA polymerase) spots on the microscope slide. A total of 5 to 8 fields of view were imaged for each experiment resulting in over 500 observed DNA molecules per sample; representative fields of view are shown in Figure 3B. Quantitative analysis of RNA polymerase-DNA co-localization accounted for incomplete labeling of RNA polymerase as described in Sanchez et al. (2011b). In order to measure the effect of LacI on the formation of competitor-resistant DNA-RNA polymerase complexes, the experiment described above was repeated using the same procedure, conditions and concentrations of all reactants, except that excess Lac repressor (200 nM) was pre-incubated with DNA for 15 min prior to addition of RNA polymerase.

Bulk In Vitro Transcription Assays

The procedure and conditions are described in more detail in (Sanchez et al., 2011b). Briefly, in order to test the transcriptional activity of the RNA polymerase-DNA complexes formed in the presence and the absence of LacI, we repeated the incubations described in the previous section, using the exact same DNA templates labeled with biotin at one end and Cy3 at the other end, and after 25 min we added rNTPs and heparin (to a final concentration of 1 mM ATP, 50 μM UTP, GTP, CTP, as well as 1μCi/μL [α^{32}P]UTP). For these experiments we used wild-type RNA polymerase holoenzyme purchased from Epicenter (Madison, WI). The resulting elongation reaction was allowed to proceed for 15 min at 37°C. At that time, the reaction was stopped and the reactants denatured by addition of an equal volume of stop buffer consisting of 7 M urea, 10 mM EDTA, and 1mg/ml bromophenol blue and xylene cyanol. Each sample was heated to 95°C for 5 min, and then loaded onto a 12% polyacrylamide denaturing gel, which was run at 1000 V for about 2 hr. The gel was scanned on a Typhoon 9410 phosphorimager (GE Health Sciences) and the amount of radiolabeled mRNA in the resulting bands was quantified using ImageQuant Software (Amersham, Piscataway NJ).

Determination of the Fraction of DNA in Open Complexes

We have previously shown that σ^{70} RNA polymerase holoenzyme can form competitor resistant complexes even with a DNA that does not bear any functional promoter (Sanchez et al., 2011b). We devised a statistical partitioning analysis to extract the fraction of RNA polymerase-DNA complexes formed at the promoter from the total number of complexes which includes RNA polymerase bound at both promoter and non-promoter locations. Using the same *lac*UV5 promoter used in this paper (but with altered flanking sequences), we showed that essentially all of the open complexes formed at the promoter had escaped into elongation 15 min after addition of rNTPs. The fraction of surviving RNA polymerase-DNA complexes corresponds to RNA polymerase bound to the non-promoter binding sites. For the DNA construct used in (Sanchez et al., 2011b), we also demonstrated that RNA polymerase-promoter and RNA polymerase-non-promoter complexes form independently and do not affect each other. Additionally, we showed that the binding of RNA polymerase to the non-promoter site is independent of the presence of Lac repressor.

To put this in quantitative terms, we can define N_T as the total number of DNA molecules in solution that are bound by at least one RNA polymerase. The population of complexes present includes three different species: (i) N_P, the number of complexes formed between a single RNA polymerase and the promoter harbored on the DNA construct, (ii) N_{NP}, the number of DNA molecules with an RNA polymerase bound at the non-promoter location and (iii) N_{PNP}, the number of DNA molecules where two polymerases are bound (one at the promoter and one at the non-promoter location). These different numbers are related such that

$$N_T = N_P + N_{PNP} + N_{NP}. \qquad \text{(Equation S1)}$$

Note that the complexes considered here are only those that are resistant to the competitor heparin. As a result, complexes such as RNA polymerase molecules bound weakly to non-specific DNA do not have to be taken into account. If we use N_{DNA} to denote the total number of DNA molecules in solution, the fraction of DNA molecules occupied by at least one RNA polymerase is given by

$$f_T = \frac{N_T}{N_{DNA}} = f_{open} + f_{NP}, \qquad \text{(Equation S2)}$$

where we have also defined the fraction of DNA molecules where RNA polymerase is only bound at the non-promoter location as $f_{NP} = N_{NP} / N_{DNA}$ and the fraction of DNA molecules bearing an open complex as $f_{open} = (N_P + N_{PNP}) / N_{DNA}$.

In our single molecule experiments, which were all done with the same concentration of DNA (8 nM), we measured only the total fraction of DNA molecules occupied by at least one polymerase, f_T. This quantity does not reveal whether RNA polymerase is bound to the promoter, the non-promoter site or to both.

In order to estimate the fraction of DNA with RNAP bound to the promoter, we resorted to bulk in vitro transcription assays. In earlier work, we showed for the *lac*UV5 promoter with O1 at +11 that the number of mRNA molecules produced in an in vitro single-round transcription reaction is identical to the number of open complexes present in solution at the time of rNTP addition (Sanchez et al., 2011b). As a result, we argue that the ratio of mRNA produced in the presence and absence of repressor, $f_{mRNA} = mRNA(R) / mRNA(R = 0)$, is a proxy for the ratio of open complexes formed in the presence and absence of repressors, namely,

$$f_{mRNA} = \frac{N_P(R) + N_{PNP}(R)}{N_P(R=0) + N_{PNP}(R=0)}. \qquad \text{(Equation S3)}$$

By simple algebraic substitutions, we can further relate f_{mRNA}, a quantity that we have measured here by bulk transcription reactions, with the fraction of DNA molecules that are only bound to RNA polymerase at non-promoter locations by writing

$$f_{mRNA} = \frac{N_P(R) + N_{PNP}(R)}{N_P(R=0) + N_{PNP}(R=0)} = \frac{N_T(R) - N_{NP}}{N_T(R=0) - N_{NP}}$$
$$= \frac{\frac{N_T(R) - N_{NP}}{N_{DNA}}}{\frac{N_T(R=0) - N_{NP}}{N_{DNA}}} = \frac{f_T(R) - f_{NP}}{f_T(R=0) - f_{NP}}, \qquad \text{(Equation S4)}$$

where it is important to remember that N_{NP} is not affected by the presence of repressor. This equation can be solved for f_{NP} which yields the fraction of RNA polymerases that are bound to non-promoter locations in terms of measurable quantities as

$$f_{NP} = \frac{f_T(R) - f_{mRNA} f_T(0)}{1 - f_{mRNA}}. \qquad \text{(Equation S5)}$$

Once we know the fraction f_{NP} we can get the fraction of molecules with RNA polymerase forming an open complex by plugging the previous equation into Equation S2 and solving for f_{open}.

Using the linear DNA template with O1 at +11, biotin at one end and a Cy3-dye at the other end, which construction we describe in the previous section, we have measured $f_T^{+11}(R=0) = 0.399 \pm 0.047$, and $f_T^{+11}(R) = 0.118 \pm 0.017$. For the same promoter, and under the exact same experimental conditions, the ratio of mRNA produced in the presence and absence of repressor has been measured to be $f_{mRNA} = 0.042 \pm 0.003$ [7]. This gives us an estimate of $f_{NP} = 0.106 \pm 0.002$. We further assume that the fraction of binding that corresponds to non-promoter locations is the same for this DNA template and for the one where O1 is at −50. The basis for this assumption is that, other than the position of O1 and the promoter region, both templates are similar in sequence and were amplified with the same PCR primers so they contain identical sequences at the ends of the DNA. We use the value of f_{NP} reported above to determine the fraction of RNA polymerase-promoter complexes formed for O1+11 and O1-50 templates, which are plotted in Figure 3C, where we also entered the measurements of $f_T(R = 0)$ and $f_T(R)$ for O1 at −50 (we found: $f_T^{-50}(R=0) = 0.40 \pm 0.09$ and $f_T^{-50}(R) = 0.32 \pm 0.04$).

In order to compute an estimate for the fraction of open complexes that form in the presence of Lac repressor bound at −50, we made the assumption that the value of f_{NP} is equal for both DNA constructs with O1 at −50 and +11, and we have based that assumption on sequence similarity. From this analysis we calculated that Lac repressor reduces open complex formation to 72 ± 22% of the number of complexes that form in the absence of repressor, as shown in Figure 3C. To test the validity of this estimate, we have in addition performed bulk transcription reactions and determined the amount of mRNA produced from the DNA with O1 at −50 in the presence and the absence of LacI, in a single-round transcription assay where we quantified the total amount of mRNA produced after the elongation reaction had proceeded to completion. We have found $f_{mRNA}^{-50} = 0.58 \pm 0.11$, which means that Lac repressor reduces open complex formation to at most ~60% of the open complexes that form in the absence of repressor. The reasonable agreement between the two measurements supports our methodology to estimate the fractional reduction in open complex formation by Lac repressor for the −50 construct.

Alternative Interpretation of In Vitro Single Molecule Results

In the text, we have proposed a mechanism consistent with our in vitro experimental results, in which Lac repressor and RNA polymerase can bind simultaneously to the promoter region when O1 is located at −50. Unfortunately, we were not able to confirm the presence of the proposed ternary complex by direct imaging because a fluorescently labeled tetrameric Lac repressor with suitable properties was not available. Still, we believe that formation of the ternary complex represents the most likely explanation for our data.

Here we consider a possible alternative explanation for our results: that when O1 was placed at −50, Lac repressor occupancy of O1 might be reduced by competition with RNA polymerase. In this scenario only a small fraction of the templates (note that the observed reduction in open complex formation in the presence of Lac repressor is very small, 28 ± 22%, as shown in Figure 3C) would actually be occupied by repressor, whereas the rest would be occupied by polymerase. This hypothesis can be ruled out on the basis of our control at +11. This control was performed with the exact same concentrations of DNA (8 nM), RNA polymerase (80 nM), and Lac repressor (200 nM), identical buffers and under the same conditions of temperature, ionic strength and pH, etc. as our measurement at −50. Given that open complexes with O1 located at +11 were reduced by as much as 96% by Lac repressor, and that repressor at +11 works by competitive inhibition, not allowing RNAP to stably bind the promoter in an open complex form (Sanchez et al., 2011b), we conclude that O1 had to be occupied at a level of at least 96%.

This fact, that RNA polymerase cannot outcompete Lac repressor is to be expected, given the known affinities of both RNA polymerase and Lac repressor for their binding sites on the *lac*UV5 promoter region. Previous measurements of the equilibrium dissociation constant K_P of RNAP at the *lacUV5* promoter placed this value in the vicinity of ~25-50 nM (Matlock and Heyduk, 1999; Straney and Crothers, 1987). The lowest estimate we have found (Matlock and Heyduk, 1999) places this value in the vicinity of 7 nM. In contrast, the equilibrium dissociation constant for the Lac repressor-O1 interaction in vitro has been measured by many groups,, and always found to be in the picomolar range ($K_R \ll$ 1nM) (Lin and Riggs, 1975; Rutkauskas et al., 2009; Wong et al., 2008). In addition, note that repressor is added in excess over polymerase (200 nM LacI for 80 nM RNAP), and very much in excess over DNA (8 nM). Given that the repressor is present at higher concentrations than RNAP, and it has a higher affinity for O1 than RNAP has for the promoter, it would not be possible for RNAP to outcompete LacI unless the affinities for their binding sites were dramatically changed by moving O1 from +11 to −50. Previous studies argue against this possibility (Lanzer and Bujard, 1988).

Due to all of these arguments, we believe that the interpretation that Lac repressor can and does bind to O1 at −50 in the presence of RNA polymerase is the simplest, most reasonable interpretation of our data, and based on this interpretation, we formulate the models that we have tested in vivo.

Theoretical Modeling and Data Fitting

Considering a Saturating Model for Repression at +11

In Figure 2D and in the main text we claim that repression at +11 is drastically different from repression from the −50 location in terms of its behavior at large number of repressors per cell. Whereas at the +11 location we claim that repression is linear with number of repressors, at the −50 location we show that the level of repression saturates. In this section we explore the consequences of allowing the model describing repression at +11 to also saturate.

Using thermodynamic models of transcriptional regulation it has been shown that such regulatory architecture can be described by

$$\text{Repression}_{+11} \approx 1 + \frac{2R}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}}, \qquad \text{(Equation S6)}$$

where R is the intracellular number of repressors, $\Delta\varepsilon_{rd}$ is the interaction energy between repressor and its operator and N_{NS} is the size of the *E. coli* genome and reflects the number of nonspecific sites where transcription factors can potentially bind (Bintu et al., 2005a; Garcia and Phillips, 2011; Vilar and Leibler, 2003).

In Figure S2C we show fits of the data of repression at +11 to the linear model given by Equation S6 and to a general saturation model given by

$$\text{Repression}_{saturation} = 1 + \frac{A\frac{R}{K_d}}{1 + \frac{R}{K_d}}, \qquad \text{(Equation S7)}$$

where R is the intracellular number of repressors, K_d is a dissociation constant and A is the saturating level of repression. We see that the data for all operators located at +11 can be fit by the saturating model as well as by the linear model. Of course, in order to really determine the saturating functional form of the data it would be desirable to measure repression for even higher intracellular numbers of repressor. The hint of saturation of the data points at higher number of repressors was already pointed out previously as a systematic shift in the repression (Garcia and Phillips, 2011). There, it was noted that the measured level of repression was systematically lower for those strains than expected on the basis of the theoretical model for repression at +11 and on the measurement of the absolute number of repressors in those strains performed using immunoblots.

One plausible explanation for the systematic shift is that there is a fraction of non-active repressors in those strains such that the actual number of active repressors is lower than thought. An alternative explanation is, however, that the theoretical model breaks down at those high numbers of repressors such that repression cannot increase as the number of repressors grows leading to a saturation described qualitatively by Equation S7. The parameters determined from fitting the model in Equation S7 to the data are shown in Figure S2C. In particular, adopting the saturating model would lead to conclude that each construct bearing a different operator saturates at a different level, as shown by the parameter A in Figure S2C. If that was the case we would then be also observing a violation of the occupancy hypothesis at +11, as for fully occupied operators the level of repression would still depend on operator sequence.

In this work we favor the explanation that the systematic shift is due to an inactive fraction of repressors in some of our strains and choose to then explore the consequence of the occupancy hypothesis in the context of the model for repression at +11 leading to Equation S6. We use these theoretical considerations to point out that the occupancy hypothesis is violated when the operator is located at −50. On the other hand, if we were to adopt the perfectly valid alternative view that the model of repression at +11 does break down for high numbers of repressors this would signal a violation of the occupancy hypothesis for the well-studied motif of simple repression of Lac repression at +11 (Bintu et al., 2005a; Garcia and Phillips, 2011; Vilar and Leibler, 2003).

Model that Combines Closed and Open Complexes into a Single State

We start by developing the model introduced in Figure 4. This model combines both the closed and open complex species into one state. This assumption is a simplification that is aligned with previous thermodynamic models of transcriptional regulation that have mainly focused on accounting for this effective complex when describing in vivo data (Ackers et al., 1982; Bintu et al., 2005b; Buchler et al., 2003; Garcia and Phillips, 2011; Vilar and Leibler, 2003).

The states, weights and rates of transcription initiation are shown in Figure 4A. The partition function for this model is

$$Z_{-50} = 1 + \frac{P}{N_{NS}} e^{-\beta \Delta\varepsilon_{pd}} + \frac{2R}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}} + \frac{P}{N_{NS}} \frac{2R}{N_{NS}} e^{-\beta\left(\Delta\varepsilon_{pd} + \Delta\varepsilon_{rd} + \varepsilon_{rp}\right)}. \qquad \text{(Equation S8)}$$

In this model we consider that RNA polymerase and repressor are found inside the cell at copy numbers of P and R, respectively and that they bind to DNA with energies $\Delta\varepsilon_{pd}$ and $\Delta\varepsilon_{rd}$. $\beta = (k_B T)^{-1}$ and N_{NS} is the number of non-specific sites, which we take to be the size of the genome and represents the non-specific background. Lac repressor can interact with RNA polymerase through an interaction energy ε_{rp}. Finally, we allow for different rates of transcription initiation off of each one of the states bearing RNA polymerase.

In order to calculate the level of gene expression we have to take the different rates of transcription initiation into account. The total rate of transcription initiation is the sum of the different initiation rates times the probability of the system being in each one of the corresponding states that allow transcription. As a result we get that the level of gene expression is proportional to

$$\text{Expression} \propto \frac{1}{Z_{-50}}\left(r_1\frac{P}{N_{NS}}e^{-\beta\Delta\varepsilon_{pd}} + r_2\frac{P}{N_{NS}}\frac{2R}{N_{NS}}e^{-\beta\left(\Delta\varepsilon_{pd}+\Delta\varepsilon_{rd}+\varepsilon_{rp}\right)}\right), \quad \text{(Equation S9)}$$

which leads to the following expression for the repression

$$\text{Repression} = \frac{1 + r[1 + p_0(e^{-\beta\varepsilon_{rp}} - 1)]}{1 + r\frac{r_2}{r_1}e^{-\beta\varepsilon_{rp}}}. \quad \text{(Equation S10)}$$

Here we have we have defined the probability of finding RNA polymerase bound to the promoter in the absence of repressor as

$$p_0 = \frac{p}{1+p}. \quad \text{(Equation S11)}$$

Additionally, we used the shorthand notation

$$p = \frac{P}{N_{NS}}e^{-\beta\varepsilon_{pd}}, \quad \text{(Equation S12)}$$

and

$$r = \frac{2R}{N_{NS}}e^{-\beta\varepsilon_{rd}}. \quad \text{(Equation S13)}$$

We are now ready to consider the different mechanisms discussed in the text (see Figure 4B). Each of these mechanisms corresponds to a different choice for the parameters of the model. As a result, these mechanisms are not mutually exclusive, but can act together in order to exert regulation. The energy diagrams for each of these mechanisms are shown schematically in Figure 4B.

First, we consider mechanism (i), where a destabilizing interaction $\varepsilon_{rp} > 0$ makes the formation of the RNA polymerase-bound complex less likely when the repressor is bound. The in vitro data shown in Figure 3C suggests that the change in RNA polymerase occupancy due to the presence of Lac repressor is at the most reduced by less than 2-fold. On the other hand, we see an in vivo repression of 40 for O1 at saturating concentrations of repressor, which would imply a change in polymerase occupancy of about 98%. As a result, the slight change in promoter occupancy by RNA polymerase due to the presence of repressor observed in vitro is not enough to explain the measured in vivo repression values. Hence, we view this model as being inconsistent with the data.

We now turn our attention to mechanism (ii). This mechanism incorporates a stabilizing interaction that prevents RNA polymerase from leaving the promoter. This interaction implies that there is a change in the height of the barrier between the polymerase-bound complex and transcription from $\varepsilon_{initiation}$ to $\varepsilon_{initiation} - \varepsilon_{rp}$ as shown in Figure 4B. As a result, according to the Arrhenius law, the rate of promoter escape has to change from r_1 to $r_2 = r_1 e^{\beta\varepsilon_{rp}}$ (Dill and Bromberg, 2011). This type of reaction scheme, of a stabilizing interaction between RNA polymerase and Lac repressor which effectively lowers the promoter escape rate, was previously suggested in such a regulatory context in (Roy et al., 2004).

If we now substitute the relation between r_2 and r_1 into Equation S10 we find

$$\text{Repression}_{(ii)} = \frac{1 + r[1 + p_0(e^{-\beta\varepsilon_{rp}} - 1)]}{1 + r} \quad \text{(Equation S14)}$$

$$= \frac{1 + \frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}\xi}{1 + \frac{2R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}},$$

where we have defined $\xi = [1 + p_0(e^{-\beta\varepsilon_{rp}} - 1)]$ as in the main text. In the in vivo experiments shown throughout this paper we have direct control over the binding energy and the intracellular number of repressors (see Figure 5, for example). As a result for mechanism (ii) we can fit the repression for the parameter $\xi = [1 + p_0(e^{-\beta\varepsilon_{rp}} - 1)]$. In Figure 4C we show a fit of Equation S14, which is the same as a fit of Equation 3, to the experimental data in red (solid line). Given the known binding energy and numbers of repressors, which determine r, we cannot fit the data to the model. This is illustrated by adopting different values for the parameter in the model leading to the red dashed curves shown in Figure 4C.

An alternative way to reveal the failure of mechanism (ii) in describing our experimental data is to examine how the functional form for repression as a function of the number of repressors determines the different features of the curve. These features are shown schematically in Figure S3C. For example, the maximum repression or saturation value is given by

$$\text{Repression}_{max,(ii)} = \xi. \quad \text{(Equation S15)}$$

Additionally, we can determine the predicted mid-point (the number of repressors at which half of the maximum repression is reached) as a function of the parameters of the model. We find this mid-point, $R_{1/2,(ii)}$, by solving the equation

$$\frac{\text{Repression}_{max,(ii)} - 1}{2} = \text{Repression}_{(ii)}\left(R_{1/2,(ii)}\right) - 1. \quad \text{(Equation S16)}$$

Notice the fact that we subtract one from the expressions on both sides of the equation. This is related to the fact that repression has a value of one when the number of repressors is zero as shown schematically in Figure S3C. Solving for $R_{1/2,(ii)}$ in Equation S16 results in

$$R_{1/2,(ii)} = \frac{N_{NS}}{2} e^{\beta \Delta \varepsilon_{rd}}. \quad \text{(Equation S17)}$$

We see that the mid-point in mechanism (ii) is solely determined by the binding energy of repressor to the operator. Since these energies are already known, the mid-point is already predetermined by our previous knowledge of the binding energy of repressor to the DNA. This results in the failure of the model to fit the experimental data shown in Figure 4C.

Finally, we consider mechanism (iii), where we allow for a different rate of promoter escape in the presence of repressor, $r_2 \neq r_1$. The corresponding reaction diagram is shown in Figure 4B. The idea behind the mechanism is that repressor is not stabilizing the RNA polymerase-DNA complex, but interacting with the promoter escape transition state. As a result, the energy of the polymerase-bound complex does not change in the presence of repressor, but the height of the barrier to transcription does. In this mechanism $\varepsilon_{rp} = 0$ and the expression we get for the repression is

$$\text{Repression}_{(iii)} = \frac{1 + r}{1 + \frac{r_2}{r_1} r}. \quad \text{(Equation S18)}$$

In this mechanism the repression saturation value is given by

$$\text{Repression}_{max,(iii)} = \frac{r_1}{r_2}. \quad \text{(Equation S19)}$$

We can also calculate the mid-point by solving the equation analogous to Equation S16 for this mechanism. Notice that in this case the mid-point is not only determined by the binding energy, but also by the modulation in the promoter escape rate, namely,

$$R_{1/2,(iii)} = \frac{N_{NS}}{2} e^{\beta \Delta \varepsilon_{rd}} \frac{r_1}{r_2}. \quad \text{(Equation S20)}$$

This is unlike mechanism (ii), where the mid-point was only determined by the binding energy (Equation S17). As a result the data is fitted well by the prediction, which leads to a value for r_2/r_1, the modulation in promoter escape rate due to the presence of repressor as shown in Figure 4C in green.

From the arguments shown above we find that the most satisfactory description of the experimental data corresponds to mechanism (iii), a direct modulation of the promoter escape rate. A combination of both mechanisms (ii) and (iii) will also describe the experimental data in a satisfactory fashion as shown in Figure S3B. In this figure we fit the data to the general formula posed in Equation S10. However, mechanism (ii) will not be able to describe the data by itself.

General Model Considering Closed and Open Complexes

The model shown in the previous section does not include the existence of both a closed and an open complex. However, experiments have typically identified at least two different complexes between RNA polymerase and the promoter (Buc and McClure, 1985; Saecker et al., 2011). In this section we show that the conclusions put forth in the previous section still apply when we make our theoretical model more realistic by considering independent closed and open complexes.

The more complete model considering both closed and open complexes is shown in in Figure S3D. In this model Lac repressor can interact with RNA polymerase in the closed complex through an interaction energy ε^c_{rp}. In order to form the open complex, an energy ε_{open} must be paid and in this context Lac repressor and RNA polymerase can still interact, but now with an energy ε^o_{rp}. Finally, we allow for different rates of transcription initiation off of each one of the open complex states bearing RNA polymerase.

The partition function for this model is given by

$$Z_{full} = 1 + \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}} + \frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}} + \frac{P}{N_{NS}} \frac{2R}{N_{NS}} e^{-\beta\left(\Delta \varepsilon_{pd} + \Delta \varepsilon_{rd} + \varepsilon^c_{rp}\right)} + \frac{P}{N_{NS}} e^{-\beta\left(\Delta \varepsilon_{pd} + \varepsilon_{open}\right)} + \frac{P}{N_{NS}} \frac{2R}{N_{NS}} e^{-\beta\left(\Delta \varepsilon_{pd} + \varepsilon_{open} + \Delta \varepsilon_{rd} + \varepsilon^o_{rp}\right)}, \quad \text{(Equation S21)}$$

where the subscript *full* indicated that this corresponds to a full model where closed and open complexes are considered explicitly. Using this partition function and the rates of transcription initiation off of each complex we get the following expression

$$\text{Expression}_{full} \propto \frac{1}{Z}\left(r_1\frac{P}{N_{NS}}e^{-\beta\left(\Delta\varepsilon_{pd}+\varepsilon_{open}\right)}+r_2\frac{P}{N_{NS}}\frac{2R}{N_{NS}}e^{-\beta\left(\Delta\varepsilon_{pd}+\varepsilon_{open}+\Delta\varepsilon_{rd}+\varepsilon_{rp}^{o}\right)}\right). \quad \text{(Equation S22)}$$

We are interested in the repression, the ratio of the expression in the absence of Lac repressor to the level of expression in the presence of repression. The resulting formula is

$$\text{Repression}_{full} = \frac{1+r[1+p_0(-1+e^{\beta\varepsilon_{open}}(e^{-\beta\varepsilon_{rp}^{c}}-1)+e^{-\beta\varepsilon_{rp}^{0}})]}{1+r\frac{r_2}{r_1}e^{-\beta\varepsilon_{rp}^{o}}}, \quad \text{(Equation S23)}$$

where p_0 is the probability of finding RNA polymerase bound to the promoter given by

$$p_0 = \frac{pe^{-\beta\varepsilon_{open}}}{1+p+pe^{-\beta\varepsilon_{open}}}. \quad \text{(Equation S24)}$$

Here, we have also defined r and p according to Equations S12 and S13.

We start by considering mechanism (ii) as defined in the previous section. In this case we will have to once again account for the change in the promoter escape rate due to the stabilization of the open complex. As a result we can make the substitution $r_2 = r_1 e^{\beta\varepsilon_{rp}^{o}}$ in Equation S23 which leads to

$$\begin{aligned}\text{Repression}_{full,(ii)} &= \frac{1+r[1+p_0(e^{\beta\varepsilon_{open}}(e^{-\beta\varepsilon_{rp}^{c}}-1)+e^{-\beta\varepsilon_{rp}^{o}}-1)]}{1+r}\\ &= \frac{1+r\xi_{full}}{1+r}.\end{aligned} \quad \text{(Equation S25)}$$

This equation and the expression for the repression in the simpler model shown in Equation 3 in the main text and Equation S14 in the previous section are basically identical. We conclude that for mechanism (ii) the full model predicts the exact same functional form for the repression as a function of repressor copy number as the simpler model where only an effective RNA polymerase-promoter complex is considered.

When considering mechanism (iii) for this full model we just need to set $\varepsilon_{rp}^{c} = \varepsilon_{rp}^{o} = 0$. The resulting repression is

$$\text{Repression}_{full,(iii)} = \frac{1+r}{1+\frac{r_2}{r_1}r}, \quad \text{(Equation S26)}$$

which is exactly the same expression obtained for this mechanism in the case of the simpler model shown in Equation S18 and Equation 3.

To conclude, we have shown that the predicted expressions for both mechanisms have the exact same functional form for a model that considers the closed and open complexes explicitly and one that considers one effective complex. As a result, none of the conclusions put forth throughout the text are affected by the simplification of using the effective complex.

Data Fitting

In order to fit the data corresponding to repression as a function of repressor copy number for the different operators located at −50 shown in Figures 4 and 5 we used our previous knowledge of the in vivo binding energy of Lac repressor to each of its operators (Garcia and Phillips, 2011). These energies are shown in Figure S3A. The errors on both axes were taken as weights for the least-squares minimization. We performed three separate fits for a given operator. First, we fixed the energy to the value reported in Figure S3A. Additionally, we performed two extra fits setting the energy to the upper and lower bounds given by the error bars. The fitted parameters and standard errors were then averaged using the errors as weights in order to obtain the values that are reported throughout the text.

Operator Occupancy

As described in the introduction, the occupancy hypothesis implies that all data should fall on the same curve, as shown in Figure 1A. Here we show in detail how we can use our theoretical models to calculate the occupancy, the probability of finding repressor bound to the different operators, and the assumptions involved in such calculations.

One of the key predictions of the occupancy hypothesis is that when plotting repression as a function of operator occupancy for different operator sequences all curves should collapse onto a single master curve. This was shown diagrammatically in Figure 1. In order to generate such a plot, however, we need to have knowledge of the occupancy of the operator at the different repressor numbers used throughout this work. An alternative strategy, which we employ here, is based on estimating the repressor occupancy using our thermodynamic model and a few assumptions. In this section we show how we can compute this probability of finding repressor bound to its operator, p_{Rbound}, in the different models of repression considered throughout the text.

For a model of steric exclusion such as the one developed in detail in (Garcia and Phillips, 2011) it can be shown that the partition function is

$$Z_{+11} = 1 + \frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}} + \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}, \quad \text{(Equation S27)}$$

where we have used the subscript +11 to indicate that we are modeling simple repression downstream from the promoter (Bintu et al., 2005a; Bintu et al., 2005b; Garcia and Phillips, 2011; Vilar and Leibler, 2003). If we calculate the probability of finding the repressor bound to the operator we get

$$p_{Rbound,+11} = \frac{1}{Z_{+11}} \frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}}. \quad \text{(Equation S28)}$$

In order to make progress we invoke the "weak promoter approximation." The key concept behind this approximation is that the binding of RNA polymerase to the promoter does not occur often enough to modulate the probability of finding repressor bound to its operator appreciably. The available in vitro measurements as well as simple estimates based upon the relevant parameters suggest that this approximation holds for the *lac* and *lac*UV5 promoters (Bintu et al., 2005b; Garcia and Phillips, 2011). Mathematically, the weak promoter approximation can be formulated as $\frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}} \ll 1$ (Bintu et al., 2005b; Garcia and Phillips, 2011), we then get

$$p_{Rbound,+11} \approx \frac{\frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}}}{1 + \frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}}}. \quad \text{(Equation S29)}$$

As a result, under the weak promoter approximation, the occupancy of Lac repressor on its operator is not affected by the binding parameters of RNA polymerase to the nearby promoter.

Next, we calculate the probability of finding repressor bound to its operator in the general thermodynamic model for upstream simple repression shown in Equation S8. In this case the probability is

$$p_{Rbound,-50} = \frac{1}{Z_{-50}} \left(\frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}} + \frac{P}{N_{NS}} \frac{2R}{N_{NS}} e^{-\beta \left(\Delta \varepsilon_{pd} + \Delta \varepsilon_{rd} + \varepsilon_{rp} \right)} \right). \quad \text{(Equation S30)}$$

Under the weak promoter approximation this probability is the same expression as $p_{Rbound,+11}$, namely

$$p_{Rbound,-50} \approx \frac{\frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}}}{1 + \frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}}} \approx p_{Rbound,+11}. \quad \text{(Equation S31)}$$

As a result, when the promoter is weak the probability of finding Lac repressor bound to its operator, p_{Rbound}, is the same regardless of the particular model of repressor action.

Throughout this work the *lac*UV5 promoter was used exclusively. Previous works suggest that the weak promoter approximation is valid for this particular promoter (Bintu et al., 2005b; Garcia and Phillips, 2011). As a result, we use Equation S31 to estimate the occupancy of Lac repressor on each operator for the different constructs and experimental conditions explored here.

In order to generate the curve of repression as a function of operator occupancy for the operators located at +11 we combined Equation S6 with Equation S31 in order to obtain

$$\text{Repression}_{+11} \approx \frac{1}{1 - p_{Rbound}}. \quad \text{(Equation S32)}$$

On the other hand, for the curves corresponding to the −50 constructs we combined Equation 4 and Equation S31, which results in

$$\text{Repression}_{-50} \approx \frac{1}{1 + p_{Rbound} \left(\frac{r_2}{r_1} - 1 \right)}. \quad \text{(Equation S33)}$$

In Figure S5A we plot our repression data as a function of operator occupancy for constructs bearing operators at +11 and −50 together with the theoretically predicted curves of the repression as a function of operator occupancy based on Equations S6 and S31 (dashed black line).

To be more specific, each data point shown in Figures S5A and S5B corresponds to a construct with an operator located at +11 or −50 in the presence of a given total number of repressor molecules. The choice of operator determines the binding free energy, $\Delta \varepsilon_{rd}$, whereas R is the intracellular number of repressors. These values are plugged into Equation S31 in order to obtain the operator occupancy, which is then plotted together with its corresponding measured repression value. Errors in the operator occupancy are

estimated using Equation S31 and the errors in R and $\Delta\varepsilon_{rd}$ (Garcia and Phillips, 2011). Each of the curves for the −50 location shown in Figures S5A and S5B corresponds to a different choice of the ratio r_2/r_1 stemming from the fits shown in Figure 5B.

It is clear that the level of occupancy is sufficient to explain repression at +11 within experimental error. This data collapsed is further explored in Figures S5C and S5D and in the next section. Our results suggest that the occupancy hypothesis holds for this regulatory architecture. In contrast, we see a different behavior for the data corresponding to the −50 constructs. Each operator has associated with it a curve for the repression versus operator occupancy which is significantly different from the other operators.

One strategy to show how different these curves are is explored in Figure S5B. Here we plot some of the data shown in Figure S5A together with the theoretical expectation. Additionally, we show the range of the theoretical expectation given by the error bars in the parameters used to generate them. We can see from these ranges that the curves are indeed different by several fold from the region delimited by their corresponding error bars. From these figures it follows that there is a qualitative difference between data sets corresponding to different operators. As a result we conclude that in the case of repression when the operator is at −50, repressor occupancy is not sufficient to determine the level of repression.

Data Collapse for Repression at +11

In Figure S5A we show that all the repression data for the operator located at +11 is consistent with Equation S32. This is done by plotting Repression as a function of $1 - p_{Rbound}$ (one minus the probability of finding Lac repressor bound to its operator) on a log-log plot, such that the data should fall on a line of slope −1. In this section we explore this claim further by applying some more rigorous statistical arguments to the claim.

One way to test the prediction of the model is to fit the data to Equation S32 leaving the slope as a free parameter. This means that we plot Repression as a function of $1/(1 - p_{Rbound})$ and fit for α in the following equation

$$\text{Repression}_{+11} \approx \alpha \frac{1}{1 - p_{Rbound}}. \qquad \text{(Equation S34)}$$

If our data is consistent with the model advocated by Equation S32 we would expect α to be close to 1. Indeed, when we perform such fit we obtain a value of $\alpha = 1.10 \pm 0.07$. In Figure S5C we show how this fit is clearly comparable to a line of slope 1.

Despite the good agreement between out prediction and the slope of the fit, the error obtained from the least-squares minimization might not be representative. In order to explore an alternative strategy to estimate the error associated in determining α we bootstrapped the data 500 times and for each new data set performed the corresponding fit. In Figure S5D we show a histogram of the values that are obtained for the slope, where we can see that the expected value of 1 is within two standard deviations of the mean.

General Kinetic Model of Repression

One way to reveal the violation of the occupancy hypothesis is to look at the repression value for a high number of repressors. Different operator sequences lead to different plateau levels in a way that our thermodynamic models can only explain if the operator sequence is determining the nature of the interaction between RNA polymerase and repressor. In this section we show that these thermodynamic models were a useful way of revealing the surprising outcome of our measurements, but that the assumptions implicit in them are by no means necessary. We consider the most general model of transcriptional repression, where we allow Lac repressor to inhibit transcription by an arbitrary mechanism which may involve inhibition of closed complex formation, inhibition of open complex formation, inhibition of promoter clearance, or any combination thereof. We do not constrain this model by making any assumptions, other than the rate of binding of Lac repressor to the operator is linear with the concentration of repressor. This mechanism is depicted in Figure S6. The promoter dynamics are described by the following master equation,

$$\frac{d\vec{P}}{dt} = \hat{K} \cdot \vec{P}. \qquad \text{(Equation S35)}$$

This equation describes the time evolution of the probability to find the promoter in each one of the six promoter states depicted in Figure S6. These six probabilities can be described by a vector $\vec{P} = (P_1, P_2, P_3, P_4, P_5, P_6)$, where we have defined P_i as the probability to find the promoter in state i. The matrix $\hat{K}$ contains the rates for all the possible promoter transitions. In order to compute the steady state probabilities, we need to solve the equation $\hat{K} \cdot \vec{P} = 0$. The mean rate of transcription in steady state is then given by $k_{txn} = k_{esc}P_3 + k^R_{esc}P_6$. The reader is referred to reference (Sanchez et al., 2011) for a careful explanation and derivation of this equation. For the general model described by Figure S6, the matrix $\hat{K}$ takes the form

$$\hat{K} = \begin{bmatrix} -(u_c + k^R) & v_c & k_{esc} & a^R & 0 & 0 \\ u_c & -(v_c + k_o + k^R_c) & v_o & 0 & a^R_c & 0 \\ 0 & k_o & -(k_{esc} + v_o + k^R_o) & 0 & 0 & a^R_o \\ k^R & 0 & 0 & -(a^R + u^R_c) & v^R_c & k^R_{esc} \\ 0 & k^R_c & 0 & u^R_c & -(a^R_c + v^R_c + u^R_o) & v^R_o \\ 0 & 0 & k^R_o & 0 & u^R_o & -(a^R_o + k^R_{esc} + v^R_o) \end{bmatrix}. \qquad \text{(Equation S36)}$$

The general expression for the mean transcription rate can be easily obtained analytically using software such as Mathematica. However, this expression is rather long and uninformative. As a result, we will not reproduce it here, but we will reproduce the repression under certain limits. We define the repression through the expression $\text{Repression} = k_{txn}(0)/k_{tnx}(R)$. We take the limit of a large intracellular number of repressors, which corresponds to setting the rate constants for repressor binding k_r, k_c^R and k_o^R to infinity. In this limit we get the repression value corresponding to the operator being occupied 100% of the time with a corresponding expression given by

$$\text{Repression}(R \rightarrow +\infty) \approx \frac{k_o u_c k_{esc}}{k_{esc}^R u_c^R u_o^R} \frac{k_{esc}^R \left(u_c^R + u_o^R + v_c^R\right) + v_c^R v_o^R + u_c^R \left(u_o^R + v_o^R\right)}{k_o u_c + k_{esc}\left(k_o + u_c + v_c\right) + v_o\left(u_c + v_c\right)}. \quad \text{(Equation S37)}$$

Notice that the repression is independent of the rate of dissociation of the repressor for any kinetic mechanism of repression by LacI given by the rates a_R, a_c^R, and a_o^R in Figure S6. As a result, the repression at saturating copy number of repressor does not depend on the operator sequence. Therefore, from this general model, we expect that the value of the plateau in our experiments will be the same regardless of the operator sequence in clear contradiction with our experimental observations.

Comparison to Previous Data

Several previous experiments have addressed the question of simple repression by Lac repressor as a function of the position of the operator. In the following section we try to compare these previous data sets to our own. Some of these experiments were done in slightly different experimental contexts such as using a different promoter geometry, different intracellular number of repressors, etc. Nevertheless, we view this comparison as useful and, in some cases, reassuring that the effects we observe are not artifacts.

Two previous works addressed the dependence of repression on the position of the main operator downstream from the promoter. Besse et al. (1986) measured the repression as a function of the position of an Oid operator. In order to do so they measured repression using a strain lacking Lac repressor and a strain overexpressing repressor through the mutation *lacI*^{q}. The number of LacI molecules per cell stemming from this mutation has been estimated to be 100 times that of the wild-type gene (Müller-Hill, 1996). This experiment was done in the presence of the auxiliary operators O2 and O3. This allows additional states of the promoter through looping of the DNA by Lac repressor, which may complicate the comparison with our data.

Elledge and Davis (1989) performed a similar experiment to those of Besse et al. However, there were some key differences. First, they used a strong, constitutive promoter with a *lac* operator located downstream. They also used an Oid operator and the *lacI*^{q} mutation. They measured repression in a way consistent with Besse et al. and our work. The construct used again contained the O2 site located within the *lacZ* coding region. As a result the data can also be potentially affected by DNA looping.

If repression is working through steric exclusion all three data sets (Elledge and Davis, Besse et al. and ours) should be described by

$$\text{Repression} = 1 + \frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}}. \quad \text{(Equation S38)}$$

In fact, we should be able to relate repression under our conditions (an O1 operator and wild-type levels of LacI) by computing

$$\text{Repression}_{O1,WT} = \left(\text{Repression}_{Oid,lacI^q} - 1\right) \frac{R_{WT}}{R_{lacI^q}} \frac{e^{-\beta \Delta \varepsilon_{rd,O1}}}{e^{-\beta \Delta \varepsilon_{rd,Oid}}} + 1, \quad \text{(Equation S39)}$$

where the subscripts correspond to the choice of operator and number of repressors. Given our knowledge of the number of Lac repressor molecules per wild-type cell (Garcia and Phillips, 2011) we estimate

$$\frac{R_{WT}}{R_{lacI^q}} \approx 10^{-2}. \quad \text{(Equation S40)}$$

We also know the difference in binding energy between Oid and O1 (Figure S3A) such that

$$\frac{e^{-\beta \Delta \varepsilon_{rd,O1}}}{e^{-\beta \Delta \varepsilon_{rd,Oid}}} = e^{-\beta\left(\Delta \varepsilon_{rd,O1} - \Delta \varepsilon_{rd,Oid}\right)} \approx 0.18. \quad \text{(Equation S41)}$$

With these two quantities in hand we can now relate the repression measured under both conditions through

$$\text{Repression}_{O1,WT} \approx \left(\text{Repression}_{Oid,lacI^q} - 1\right) \times 0.18 \times 10^{-2} + 1.$$

In Figure S1D we show a comparison of our data to that of Besse et al. and Elledge and Davis once it has been converted to our approximate experimental conditions using Equation S42. We see what seems to be a relatively good agreement in terms of the general trends of all data sets. However, the magnitude of the effects seems to be different. It is not clear whether this is due to Equation S40 not successfully converting all the data or to some other differences in the experiments that we have not been able to account for.

Bond et al. (2010) measured repression as a function of an Oid operator located upstream from the promoter in the absence of a main operator. They expressed Lac repressor from an F-plasmid at wild-type levels. One very important difference between their and our approaches is the fact that they measure a "repression ratio" by quantifying the level of expression in the presence of the inducer IPTG and dividing it by the level of gene expression in the absence of IPTG. Figure S1E shows a comparison of our repression and the repression ratio measured by Bond et al. Both data sets seem to show the same features, although ours shows a quantitatively stronger modulation of repression. This difference could be attributable to, for example, differences in the actual number of repressors in both experiments. Another possible source of the discrepancy could be incomplete derepression by IPTG. Some previous measurements suggest that Lac repressor can still loop and lead to repression in vivo in the presence of IPTG (Becker et al., 2007; Becker et al., 2005; Bond et al., 2010). As a result, even though IPTG can abolish some of the binding of LacI to the DNA it may not always abolish it completely. Additionally, it is not clear if IPTG could cause changes in the structure of the protein and the protein-DNA complex such that the contacts with the transcriptional machinery could be of a different nature.

SUPPLEMENTAL REFERENCES

Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. Proc. Natl. Acad. Sci. USA *79*, 1129–1133.

Becker, N.A., Kahn, J.D., and Maher, L.J., 3rd. (2007). Effects of nucleoid proteins on DNA repression loop formation in *Escherichia coli.* Nucleic Acids Res. *35*, 3988–4000.

Becker, N.A., Kahn, J.D., and Maher, L.J., 3rd. (2005). Bacterial repression loops require enhanced DNA flexibility. J. Mol. Biol. *349*, 716–730.

Besse, M., von Wilcken-Bergmann, B., and Müller-Hill, B. (1986). Synthetic lac operator mediates repression through lac repressor when introduced upstream and downstream from lac promoter. EMBO J. *5*, 1377–1381.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005a). Transcriptional regulation by the numbers: applications. Curr. Opin. Genet. Dev. *15*, 125–135.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005b). Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. *15*, 116–124.

Bond, L.M., Peters, J.P., Becker, N.A., Kahn, J.D., and Maher, L.J., 3rd. (2010). Gene repression by minimal lac loops in vivo. Nucleic Acids Res. *38*, 8072–8082.

Buc, H., and McClure, W.R. (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.

Buchler, N.E., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. Proc. Natl. Acad. Sci. USA *100*, 5136–5141.

Cloutier, T.E., and Widom, J. (2004). Spontaneous sharp bending of double-stranded DNA. Mol. Cell *14*, 355–362.

Cloutier, T.E., and Widom, J. (2005). DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. Proc. Natl. Acad. Sci. USA *102*, 3645–3650.

Dill, K.A., and Bromberg, S. (2011). Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience, Second Edition (London, New York: Garland Science).

Elledge, S.J., and Davis, R.W. (1989). Position and density effects on repression by stationary and mobile DNA-binding proteins. Genes Dev. *3*, 185–197.

Friedman, L.J., Chung, J., and Gelles, J. (2006). Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. Biophys. J. *91*, 1023–1031.

Garcia, H.G., Lee, H.J., Boedicker, J.Q., and Phillips, R. (2011). Comparison and calibration of different reporters for quantitative analysis of gene expression. Biophys. J. *101*, 535–544.

Garcia, H.G., and Phillips, R. (2011). Quantitative dissection of the simple repression input-output function. Proc. Natl. Acad. Sci. USA *108*, 12173–12178.

Hochschild, A., and Dove, S.L. (1998). Protein-protein contacts that activate and repress prokaryotic transcription. Cell *92*, 597–600.

Kuhlman, T., Zhang, Z., Saier, M.H., Jr., and Hwa, T. (2007). Combinatorial transcriptional control of the lactose operon of *Escherichia coli.* Proc. Natl. Acad. Sci. USA *104*, 6043–6048.

Lanzer, M., and Bujard, H. (1988). Promoters largely determine the efficiency of repressor action. Proc. Natl. Acad. Sci. USA *85*, 8973–8977.

Lin, S., and Riggs, A.D. (1975). The general affinity of lac repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eucaryotes. Cell *4*, 107–111.

Lutz, R., and 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.

Matlock, D.L., and 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.

Müller-Hill, B. (1996). The *lac* Operon: a short history of a genetic paradigm (Berlin, New York: Walter de Gruyter).

Pavco, P.A., and Steege, D.A. (1990). Elongation by *Escherichia coli* RNA polymerase is blocked in vitro by a site-specific DNA binding protein. J. Biol. Chem. *265*, 9960–9969.

Pavco, P.A., and Steege, D.A. (1991). Characterization of elongating T7 and SP6 RNA polymerases and their response to a roadblock generated by a site-specific DNA binding protein. Nucleic Acids Res. *19*, 4639–4646.

Record, J. M. T., Reznikoff, W., Craig, M., McQuade, K., and Schlax, P. (1996). *Escherichia coli* RNA polymerase (sigma70) promoters and the kinetics of the steps of transcription initiation. In In *Escherichia coli* and Salmonella Cellular and Molecular Biology, F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K. Brooks Low, B. Magasanik, W.S. Reznikopp, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington, DC: ASM Press), pp. 792-821.

Rojo, F. (1999). Repression of transcription initiation in bacteria. J. Bacteriol. *181*, 2987–2991.

Roy, S., Semsey, S., Liu, M., Gussin, G.N., and Adhya, S. (2004). GalR represses galP1 by inhibiting the rate-determining open complex formation through RNA polymerase contact: a GalR negative control mutant. J. Mol. Biol. *344*, 609–618.

Rutkauskas, D., Zhan, H., Matthews, K.S., Pavone, F.S., and Vanzi, F. (2009). Tetramer opening in LacI-mediated DNA looping. Proc. Natl. Acad. Sci. USA *106*, 16627–16632.

Saecker, R.M., Record, M.T., Jr., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J. Mol. Biol. *412*, 754–771.

Sanchez, A., Garcia, H.G., Jones, D., Phillips, R., and Kondev, J. (2011a). Effect of promoter architecture on the cell-to-cell variability in gene expression. PLoS Comput. Biol. *7*, e1001100.

Sanchez, A., Osborne, M.L., Friedman, L.J., Kondev, J., and Gelles, J. (2011b). Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. EMBO J. *30*, 3940–3946.

Straney, S.B., and Crothers, D.M. (1987). Kinetics of the stages of transcription initiation at the *Escherichia coli* lac UV5 promoter. Biochemistry *26*, 5063–5070.

Vilar, J.M., and Leibler, S. (2003). DNA looping and physical constraints on transcription regulation. J. Mol. Biol. *331*, 981–989.

Wong, O.K., Guthold, M., Erie, D.A., and Gelles, J. (2008). Interconvertible lac repressor-DNA loops revealed by single-molecule experiments. PLoS Biol. *6*, e232.

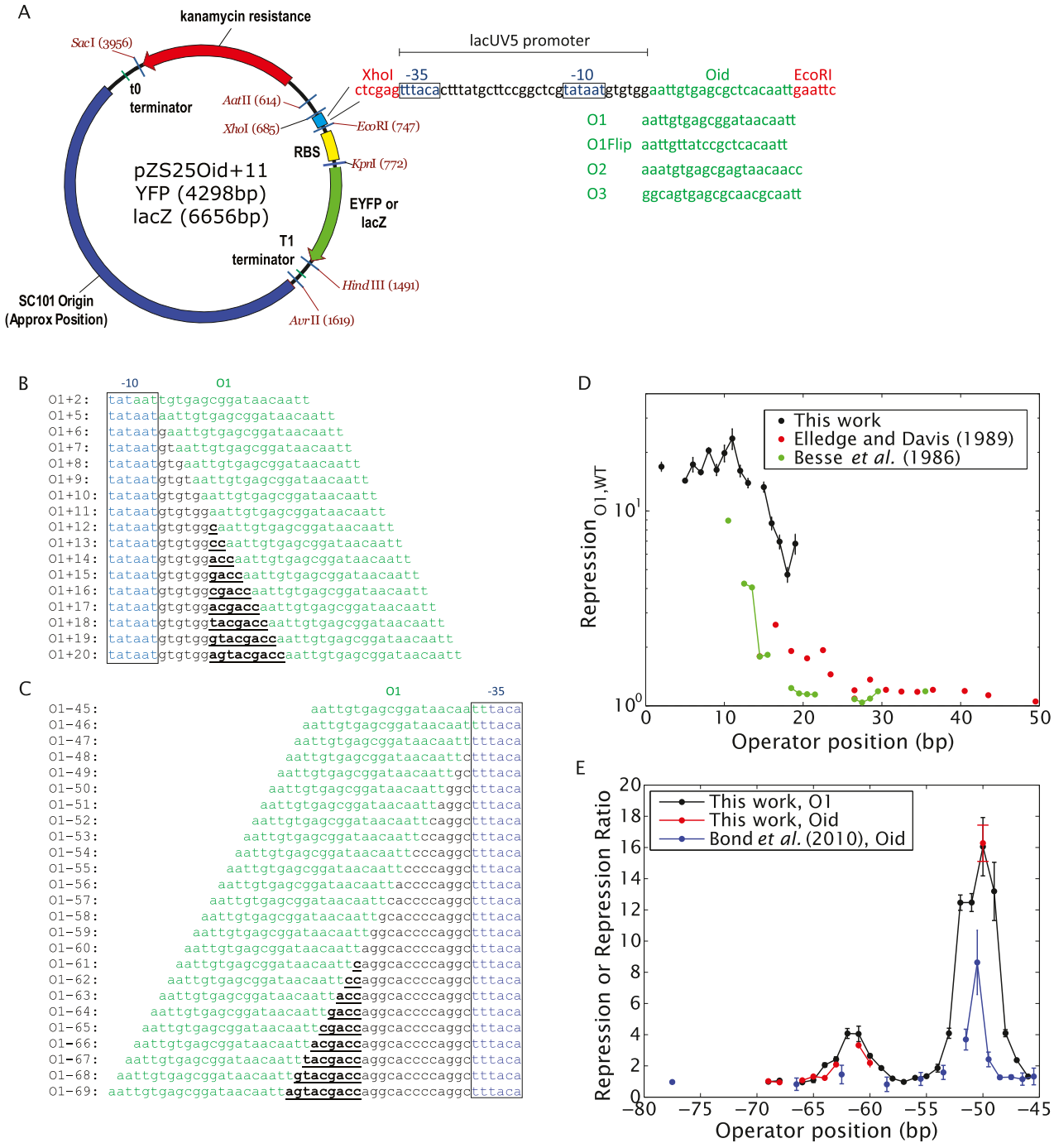

Figure S1. Plasmid Diagram, Promoter Sequences and Comparison of Our Experimental Data to Previous Similar Experiments, Related to Figure 2

(A) The main features of the plasmids pZS25O1+11-YFP and pZS25O1+11-lacZ are shown flanked by unique restriction sites. The particular promoter sequence based on the *lac*UV5 promoter is shown together with the sequences of the different Lac repressor binding sites used.

(B) Sequences of downstream repression constructs. When possible the sequences of the wild-type *lac* promoter were used.

(C) Sequences of upstream repression constructs. We used the upstream sequence of the *lac* promoter until the CRP binding site (Müller-Hill, 1996). In (B) and (C) the underlined and bold letters correspond to extra sequence obtained from the random sequence E8 (Cloutier and Widom, 2005).

(D) Our data for operator positions located downstream from the promoter is compared to the data sets of Besse et al. (Besse et al., 1986) and Elledge and Davis (Elledge and Davis, 1989) via the conversion shown in Equation S40.

(E) Our repression data is compared to the repression ratio data (defined with respect to the presence and absence of IPTG in the media) from Bond et al. (Bond et al., 2010) which shows qualitative trends similar to our data. Error bars are standard deviations over multiple repeats.

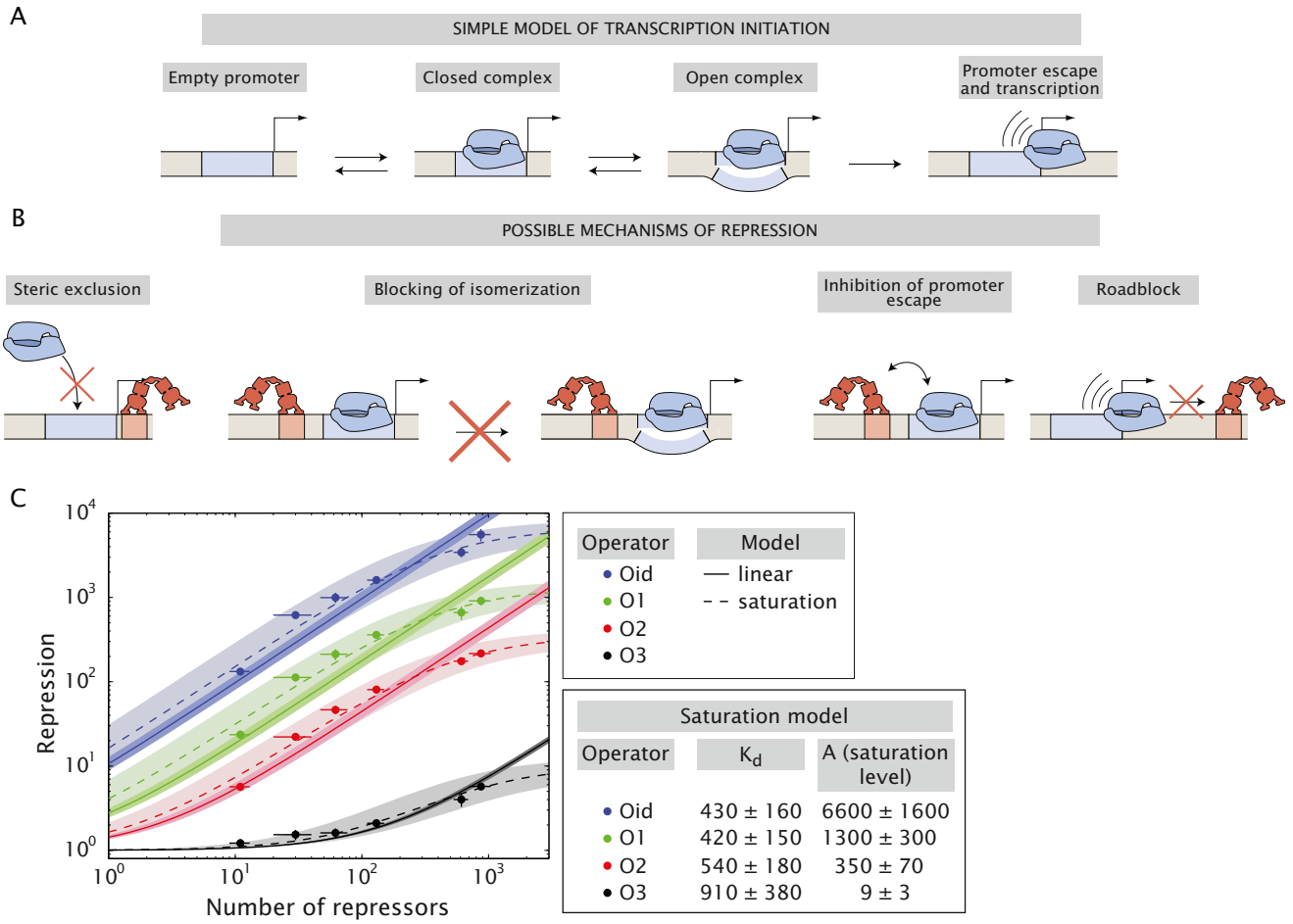

Figure S2. Transcriptional Regulation and the Mechanisms of Simple Repression, Related to Figures 2 and 3

(A) Simplified model of transcription initiation (Record et al., 1996). In this schematic, the DNA in the open-complex state is not meant to reflect the actual structure of the complex with any accuracy. The key point of the cartoons is to highlight the different allowed states.

(B) Different mechanisms of repression where transcription can either be impaired by excluding RNA polymerase from one of its DNA complexes, by the inhibition of isomerization into the open complex, by inhibition of promoter escape, or by creating a roadblock for RNA polymerase once it has escaped the promoter (Hochschild and Dove, 1998; Pavco and Steege, 1990, 1991; Rojo, 1999).

(C) Fits of the repression data for the operator located at +11 to the linear model given by Equation S6 and a saturation model given by Equation S7. K_d is measured in units of absolute number of repressor molecules. Error bars are standard deviations over multiple repeats and shaded regions represent the errors from the fits.

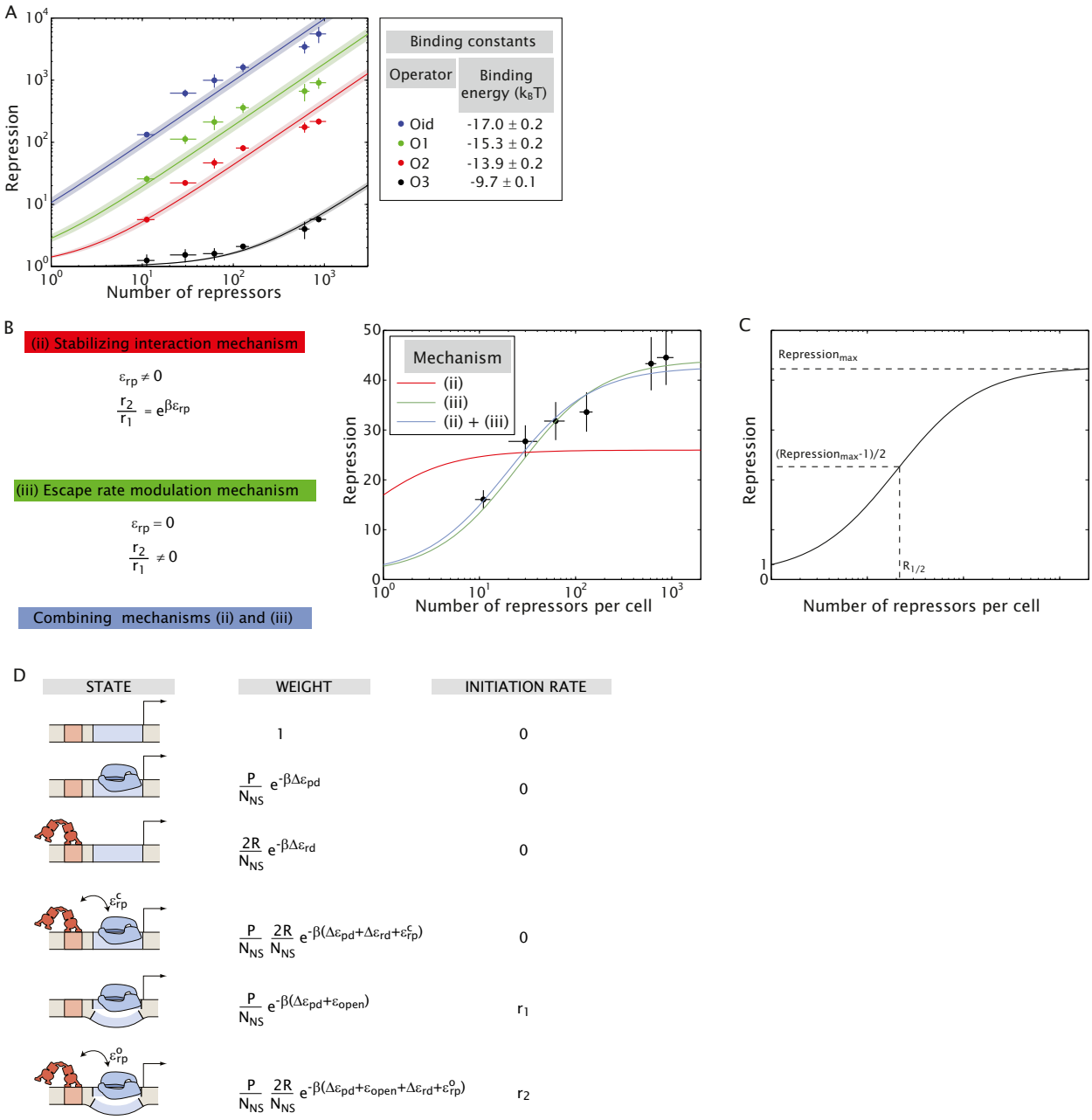

Figure S3. Repression for Operators Located at +11 and Various Proposed Models and Mechanisms to Fit Our Experimental Data for Repression at −50, Related to Figure 4

(A) Repression measured for the four different operators for Lac repressor located at +11 as a function of the intracellular number of repressors. By fitting Equation S6 to the experimental data we can obtain the in vivo binding energy of Lac repressor to each operator. Notice that repression increases steadily with the number of repressors. This behavior is to be contrasted with the saturation in repression for operators located at −50 shown in Figures 4 and 5. Data replotted from (Garcia and Phillips, 2011).

(B) Fit of the O1 −50 data to the various versions of our thermodynamic model. A stabilizing interaction between RNA polymerase and LacI only cannot account for the observed trends in the data (see also Figure 4) as shown by the best fit to mechanism (ii) (red line). A modulation in promoter escape rate by bound repressor is necessary to account for the observed trends in the data as shown by the best fit to mechanism (iii) (green line). Finally, a combination of both models can also account for the data as shown by a best fit to Equation S10 (blue curve).

(C) Graphical definition of the maximum level of repression, $Repression_{max}$, and the repressor copy number corresponding to the mid-point of the repression curve, $R_{1/2}$.

(D) Full model of upstream simple repression. In this model we differentiate between the open and closed complexes. We allow for an interaction between repressor and RNA polymerase in the context of both complexes. Additionally, we allow for a different rate of transcriptional initiation in the presence of repressor. Error bars are standard deviations over multiple repeats.

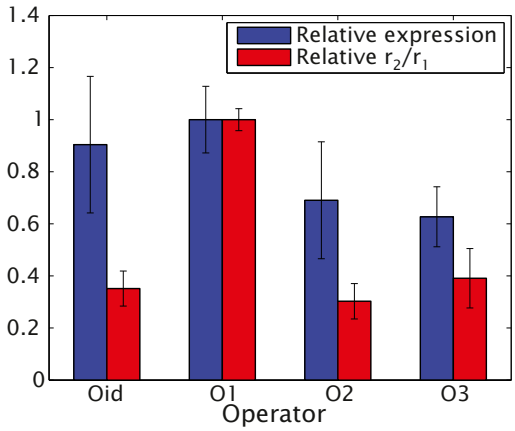

Figure S4. Comparison between Relative Levels of Expression and the Modulation in Ratio of Promoter Escape Rates r_2/r_1, Related to Figure 5

The level of expression in the absence of Lac repressor is compared to the result stemming from the different values of r_2/r_1 following Equation 4. Values are expressed relative to the O1 values, which are set to one. We see no significant correlation between the unregulated level of gene expression and the ratio of escape rates suggesting that the rate of promoter escape in the absence of repressor, r_1 cannot account for the observed differences in r_2/r_1 for the different operators. Error bars are standard deviations over multiple repeats.

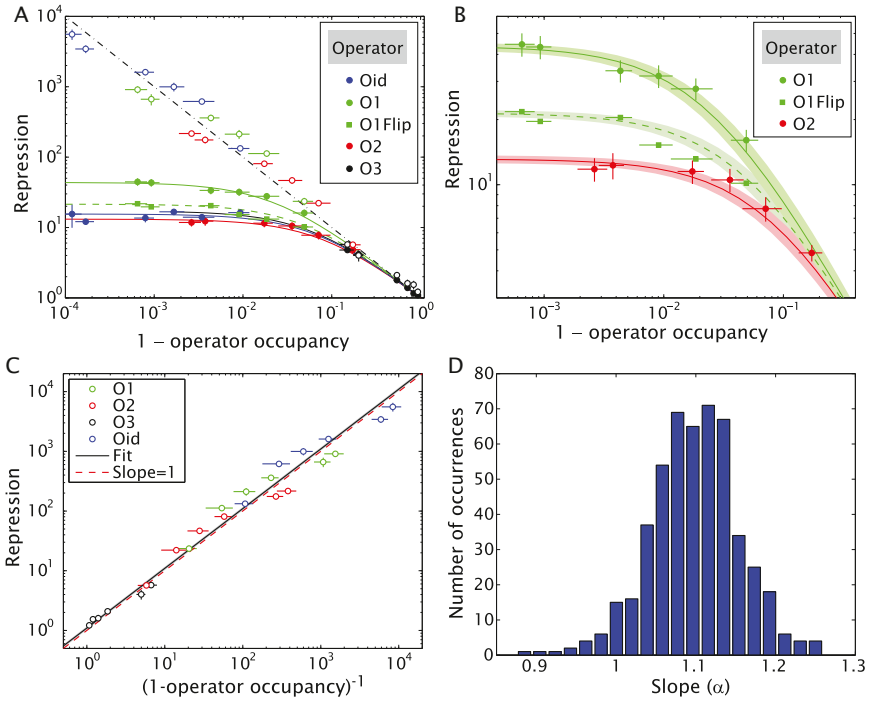

Figure S5. Repression at +11 Yields a Master Curve as Predicted by the Occupancy Hypothesis while Repression at −50 Does Not, Related to Figure 5

(A) Repression as a function of the probability of finding the operator empty (one minus operator occupancy) for operators at −50 (filled circles and squares) and at +11 (open circles). Note that for the +11 architecture, all of the data points collapse onto a single master curve (dash-dot black line) as predicted by the occupancy hypothesis and shown in detail in the Extended Experimental Procedures. For the −50 architecture, there is no data collapse onto a master curve as revealed by the fact that the different operator sequences are associated with different modulations of the rate of transcription initiation (solid and dashed lines).

(B) A subset of the data shown in (A) is reproduced here on a different scale to illustrate the lack of data collapse for the −50 architecture. The shaded regions represent the error in our prediction given the errors associated with the various parameters used to generate them. Error bars are standard deviations over multiple repeats.

(C) We fit our repression data for the operator located at +11 to Equation S35 and obtain the corresponding slope. The value obtained is comparable to the value predicted by the occupancy hypothesis of one.

(D) Bootstrapping our data 500 times and performing the same fit as in (C) for each one of these iterations allows us to obtain a histogram of slopes. Although there is a systematic shift in the obtained slope from the expected value of one the obtained data is consistent with the prediction of the occupancy hypothesis.

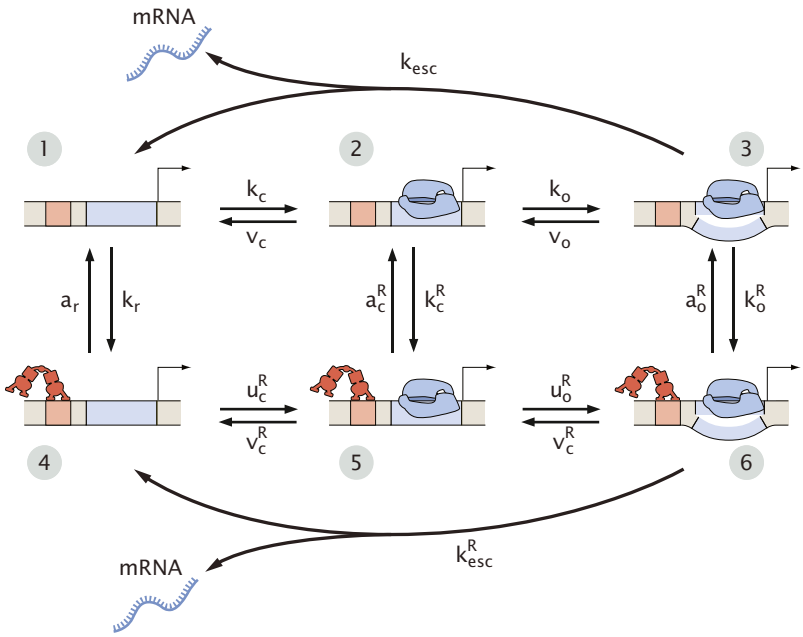

Figure S6. Full Kinetic Model of Repression, Related to Figure 5

This model allows for Lac repressor to inhibit transcription by inhibition of closed complex formation, inhibition of open complex formation, inhibition of promoter escape, or any combination thereof. RNA polymerase states are indicated using the same symbols as in Figure S2A.