SUPPLEMENTAL MATERIAL for "Theoretical and experimental dissection of DNA loop-mediated repression"

James Q. Boedicker and Rob Phillips
Department of Applied Physics, California Institute of Technology,
1200 East California Boulevard, Pasadena, CA 91125

Hernan G. Garcia
Department of Physics, Princeton University, Jadwin Hall, Princeton, NJ 08544

Experimental Procedures:

Strain Construction:

We created a library of strains containing *lac* operon constructs with a main operator centered at +11 bp relative to the transcription start site and an upstream auxiliary operator, as seen in Fig. 1(A). We first created these constructs in the promoter region of plasmid pZS25O1+11-YFP (as described in Garcia and Phillips [1]). Site directed mutagenesis was used to extend the looping sequence between the operator and the promoter in single base pair increments, resulting in a library with operator distances (the center to center distance between the operators) between 61.5 and 161.5 bp. These constructs did not contain a binding site for cAMP receptor protein (CRP), which has been shown to be involved in loop formation in the wild type operon [2]. Mutagenesis was used to change or remove the main operator from these constructs. Constructs were verified by sequencing, and are available upon request. The sequences corresponding to these constructs and their variable looping sequences can be found in Tables S3-4.

Constructs were integrated into the *galK* locus of *E. coli* strain HG104 with the wild-type *lacI* background through recombineering, as described in Garcia and Phillips [1]. To measure the unregulated level of expression for each construct, DNA looping constructs were transduced into strain HG105, containing a deletion of the *lacI* gene, by P1 transduction (http://openwetware.org/wiki/Sauer:P1vir_phage_transduction). Looping constructs were selected using kanamycin. In order to titrate the number of Lac repressor molecules per cell, previously characterized constructs expressing several different average intracellular numbers of repressor molecules, measured in a bulk population of cells, were transferred into looping constructs in strain HG105 by P1 transduction using chloramphenicol resistance as a selection marker [1].

Measuring Gene Expression:

Overnight cultures were grown in Luria Broth (EMD, Gibbstown, NJ) containing kanamycin at 37^{o}C with shaking at 250 rpm, and were used to inoculate 3 mL scale cultures containing M9 buffer (2 mM $MgSO_4$, 0.10 mM $CaCl_2$, 48 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 19 mM NH_4Cl) with 0.5% glucose as a carbon source. Cells were grown in 14 mL Falcon BD tubes with the cap placed loosely at 37^{o}C with shaking at 250 rpm for approximately 10 generations. Cells were harvested at OD_{600} between 0.25 and 0.65.

Repression is the relative change in gene expression due to the presence of Lac repressor, as shown in Eq. 1. Experimentally, this is the ratio of expression for the constructs integrated into HG104 and HG105, with and without Lac repressor respectively. Cells not containing YFP were grown to determine the background autofluorescence. Fluorescence measurements were obtained using a Tecan Safire 2 by pipetting 200 μL of culture into the wells of a 96 well plate (Costar, #3631, Corning, NY). Measurements were taken from the bottom of each well with excitation and emission of 505 and 535 nm respectively, both with a 12 nm bandwidth. Here we quantified gene expression using a fluorescent reporter and a plate reader. We have previously demonstrated that comparable results can be obtained with single-cell fluorescence microscopy or a *lacZ* gene reporter for the range of expression we observed [3].

To calculate the repression at a given operator distance, first the background from the media was subtracted from all measurements. Then fluorescence measurements were normalized by dividing by the optical density of each culture at 600 nm (OD_{600}) and corrected for cell autofluorescence. On each day, all strains were measured in triplicate and a mean and standard deviation for each day was calculated for each strain. Measurements were repeated on multiple days, and the mean and standard error for each construct was calculated from the means of each day weighted by the standard deviation, see below. It is important to note that some particular strains showed a much higher variability between replicates and between days than others. The strains corresponding to 62 repressors per cell in Fig. 2(A) are an example of such behavior. We associate this to the fact that the overall fluorescence of this construct (both in the presence and absence of Lac repressor) is very close to the autofluorescent background, making a reliable quantification of the level of gene expression challenging and increasing our uncertainty in the reported value for repression.

We determined if the final optical density of the culture introduced a bias on the normalized fluorescence intensity. For every set of experiments, a positive control strain was measured in triplicate, and we use this standard to examine how the fluorescence intensity normalized to optical density correlated with the optical density. The control strain contained a YFP reporter with a single operator O2 centered at +11 in host strain HG105. Compiling all of the positive control data from each experiment, we observe a weak dependence of fluorescence per OD_{600} on the final optical density of the culture, as shown in Fig. S1(E). All measurements were taken between OD_{600} of 0.25 to 0.65 and over this range the average ratio of fluorescence intensity to OD_{600} decreased approximately 20%, which is similar to the day to day variation of replicate measurements at a similar OD_{600}. Because the variation between replicate measurements was similar to variation caused by the final optical density of the culture, it should not introduce significant noise or bias in gene expression measurements.

The weak promoter approximation

The weak promoter approximation has been previously implemented in similar models [1, 4]. The experimentally determined values for the number of RNA polymerase molecules per cell, *P*, is roughly 2000 and the binding energy of RNA polymerase to the *lac*UV5 promoter is approximately -5 k_BT, which makes $\frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}$ less than 10^{-1} [1, 5, 6]. This is much smaller than the weights of the other states

listed in Fig. 1(B). As shown in Fig. S6, the probability of states with RNA polymerase bound are at most approximately 10^{-3}.

Connecting the statistical mechanical theory to the language of dissociation constants

The repression equation found in Fig. 1(C) can also be converted to an equation which uses concentrations of repressor, binding constants of repressor to the operator, and J-factors. As has been shown previously [4, 7, 8], the terms in the *repression* equation can be converted into expression involving K_d using $K_{d,Oid} = \frac{[R][Oid]}{[R-Oid]}$, (S1)

in which [R] is the concentration of repressor tetramers, [Oid] is the concentration of the auxiliary operator Oid, and [R-Oid] is the concentration of repressors bound to the auxiliary operator Oid (all three terms in units of M), and $K_{d,Oid}$ is the dissociation constant of Lac repressor from the auxiliary operator Oid. An analogous term can be derived for the main operator O2. The dissociation constant is related to the free energy change by,

$$\Delta G = -k_B T \ln(\frac{K_d}{C^o}), \quad \text{(S2)}$$

in which ΔG is the free energy change of binding, k_B is Boltzmann's constant, T is temperature, and C^o is the standard reference concentration of 1 M. Note that the reference concentration is needed to cancel the units of K_d. This conversion leads to,

$$\frac{2R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rad}} = \frac{[R]}{K_{d,Oid}}, \quad \text{(S3)}$$

in which $\Delta\varepsilon_{rad}$ is the binding energy of repressor to the auxiliary operator. The factor of two in the left side of Eq. S3 accounts for two binding heads on the repressor. Similarly the term involving ΔF_{loop} can be rewritten as,

$$\frac{2R}{N_{NS}} e^{-\beta(\Delta \varepsilon_{rmd} + \Delta \varepsilon_{rad} + \Delta F_{loop})} = \frac{[R]J}{2K_{d,O2}K_{d,Oid}}, \quad \text{(S4)}$$

in which J is the J-factor or the effective local concentration of the repressor as a result of loop formation with a factor of ½ to reflect the symmetry of the Lac protein and the binding sites [4, 7], and $\Delta\varepsilon_{rmd}$ is the binding energy of repressor to the main operator. In these terms repression is,

$$repression = \frac{1 + \frac{[R]}{K_{d,O2}} + \frac{[R]}{K_{d,Oid}} + \frac{[R]}{K_{d,O2}}\frac{[R]}{K_{d,Oid}} + \frac{[R]J}{2K_{d,O2}K_{d,Oid}}}{1 + \frac{[R]}{K_{d,O2}}}. \quad \text{(S5)}$$

It can be seen that upon converting from the statistical mechanical expression to the expression in terms of dissociation constants that reference to binding energies and the nonspecific background through the factors of N_{NS} no longer appear. However, hidden in these dissociation constants is a reference state of 1 M activity as noted above. N_{Ns} serves a similar purpose of normalizing the number

of repressors per cell to a reference state, however it also has the microscopic interpretation as the number of non-specific binding sites for the repressor in the cell.

In terms of how the value of N_{NS} influences calculations with the model, let us consider what happens when N_{NS} is changed by a factor of α. Such an adjustment of N_{NS} leads to changes in the weighting terms in Fig. 1(B). For example, the weight of state 6 becomes,

$$weight(state=6)=\frac{4R(R-1)}{\alpha^2 N_{NS}{}^2}e^{-\beta(\Delta\varepsilon_{rmd}+\Delta\varepsilon_{rad})} \quad . \quad \text{(S6)}$$

Letting $\frac{1}{\alpha}=e^{-\varepsilon_\alpha}$, Eq. S6 can be rewritten as,

$$weight(state=6)=\frac{4R(R-1)}{N_{NS}{}^2}e^{-\beta(\Delta\varepsilon_{rmd}+\varepsilon_\alpha+\Delta\varepsilon_{rad}+\varepsilon_\alpha)}, \text{ (S7)}$$

in which the correction to N_{NS} now appears as a correction to the binding energies. Similar transformations can be made to the weights of the remaining states in Fig. 1(B). Adjustments to the value of N_{NS} do change the weight of state 1, which is normalized to a value of 1 regardless of the definition of N_{NS}. Since changes to N_{NS} result in a rescaling of the background energy it will be important to keep this factor constant when measuring binding energies, which we do when inheriting the operator binding energies from a previous publication [1].

Projecting previous data sets onto our experimental conditions

One question of interest is the extent to which different studies on similar genetic architectures yield the "same" results. In the context of looping, for example, we can ask whether different experiments performed in different labs yield the same looping free energies when analyzed through the thermodynamic model. We compare our results to two previous studies on similar genetic constructs [9, 10] . Making comparisons between different studies is not straightforward since key parameters such as the number of repressors per cell and the operator strengths were different in each case. Our equilibrium statistical mechanical model predicts that these parameters strongly influence the extent of repression. Hence, in order to make a direct comparison, our model was used to "project" results from earlier studies onto the results of this study. This was done by calculating the looping free energy of each construct measured in the previous studies, and then using the looping free energy to predict the fold repression for the conditions used in our experiment, Oid-O2 loops with 11 repressors per cell, as reported in Fig. S1(A). Even using this method, which corrects for differences in repressor number and choice of operators, we still are not adjusting for other experimental parameters which differed between the studies including culturing conditions, strain backgrounds, whether or not IPTG was used to turn off repression, and the sequence of the looped DNA. Nonetheless, this comparison should prove fruitful in determining whether parameters such as repressor number and operator binding energies are the major factors that set the overall level of repression.

The study by Becker *et al*. was performed with constructs containing the operators Oid and O2, as in our constructs [9]. In this study repression was defined as the ratio of the expression level in the presence

of IPTG to the expression in the absence of IPTG. IPTG was used to "turn off" the Lac repressor whereas in our case, repression is "turned off" by eliminating Lac repressor altogether, a difference that could impact differences observed between the two studies. The authors also measured the expression level for constructs not containing any operators in the absence of IPTG, data which we obtained through a personal communication. No operator constructs cells still contain Lac repressor, but since specific binding sites for repressor in the vicinity of the gene reporter are absent, we assume their level of expression to be similar to our Δ*lacI* strains. We use the no operator construct to calculate repression,

$$\text{repression}_{Becker} = \frac{\text{normalized gene expression no operator}(R \neq 0, \text{no IPTG})}{\text{normalized gene expression looping construct } (R \neq 0, \text{no IPTG})}, \quad \text{(S8)}$$

in which the data from different measurements is normalized by dividing by gene expression of the positive control strain, O2 alone at +11, from each data set.

In these measurements, the *lacI* gene was placed on a single copy episome and reported to produce wild-type levels of repressor. Based on previous measurements, the wild-type LacI levels are approximately 11 copies of repressor per cell (Table S5).

Data from Muller *et al*. was obtained using constructs containing Oid as the auxiliary operator, O1 as the main operator, and approximately 50 repressors per cell [10]. This data set is measured using the definition of repression introduced in Eq. 1. The equation in Fig. 1(C) was used to calculate $\Delta F_{loop}(L)$ from the reported repression data. The looping energy was then used to calculate repression for Oid-O2 with 11 repressors per cell using the equation in Fig. 1(C). The projected repression from both sets of data is shown against the data reported here in Fig. S1(C). The looping energies extracted from all three data sets are shown in Fig. S1(D).

We find that some features of each data set, such as the exact positions of the peaks and troughs and the detailed shapes of the curves, did vary between the data sets. Such subtle differences in the shapes of the looping curves have been attributed to various loops being able to form either through conformational changes in the protein or because of different DNA loop topologies [11-13]. However, it is not clear why these protein and DNA loop topologies would differ in the different experiments. Still, there is an overall agreement on the oscillatory pattern of repression with distance and the approximate range of looping energies observed. Similar conclusions were already reached when comparing the data sets of Muller *et al.* and Becker *et al.* [14].

Model for regulation from an upstream operator

To quantify the direct contribution of the upstream operator on repression, we measured gene regulation from constructs containing only an upstream auxiliary operator, as shown schematically in Fig. 4(A). The data was analyzed using a thermodynamic model described previously [15]. Briefly, the states and weights for this model are shown in Fig. S5(A). Analogous to Eq. 2 derived for the looping case, in the case of a single auxiliary operator, the effective transcription rate is given by,

$$\textit{gene expression}_{\textit{auxiliary}}(L) = \frac{\frac{r_2 P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}} + \frac{r_3 P}{N_{NS}} \frac{2R}{N_{NS}} e^{-\beta(\Delta\varepsilon_{pd}+\Delta\varepsilon_{rad})}}{1+\frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}} (1+\frac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rad}}) + \frac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rad}}}, \quad \text{(S9)}$$

Here we assumed that the effective transcriptional rate constants for states not containing RNAP (1 and 4) are 0 and for states 2 and 3 are r_2 and r_3, respectively. We assume that a repressor bound upstream from the promoter can interact with RNA polymerase resulting in a modulation of the transcription rate of that state. After making the weak promoter approximation, the repression can be expressed as,

$$\textit{repression}_{\textit{auxiliary}}(L) = \frac{1+\frac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rad}}}{1+\frac{r_3(L)}{r_2(L)} \frac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rad}}}, \quad \text{(S10)}$$

where we have explicitly included the dependence on the position of the operator in the rates r_2 and r_3. $\textit{repression}_{\text{auxiliary}}(L)$ is reported in Fig. S5(B). The nature of the interaction of the upstream-bound repressor and RNA polymerase is, presumably, a function of the relative distance between the two. As a result we can calculate the ratio of transcription rates, r_3/r_2, for each position of the auxiliary operator using

$$\frac{r_3}{r_2}(L) = \frac{1+\frac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rad}} - \textit{repression}_{\textit{auxiliary}}(L)}{\textit{repression}_{\textit{auxiliary}}(L) \frac{2R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rad}}}. \quad \text{(S11)}$$

The ratio of transcription rates shown in Fig. 4(B) was calculated from the data in Fig. S5(B). Using this ratio, we can account for looping independent regulation by the upstream operator in the calculation of the looping free energy.

Accounting for the effect of the auxiliary operator in looping

Direct auxiliary gene regulation can be incorporated into the model by no longer assuming that the rate of transcription is equivalent for all states in which the RNA polymerase is bound. In other words, the same equilibrium thermodynamic model determines the probability of each state, but we no longer assume that all states with RNAP bound initiate transcription at the same rate, i.e. r_2 and r_3 in Fig. 1(B) can have different values. It has been previously suggested that this adjustment to the transcription rate is due to interference with promoter escape [15].

We adjust our calculation of ΔF_{loop} to account for this new mode of repression using the value of $r_3/r_2(L)$ extracted using Eq. S11. The model allows us to separate the contributions of loop formation and direct upstream gene regulation in the total amount of repression measured for two operator constructs.

Fourier analysis was used to quantify the oscillation frequency in our data. We analyzed the data between 73.5 and 121.5 bp shown in Fig. S1(A), the region which exhibited the strongest oscillatory

pattern. Linear interpolation was used to estimate repression values for the few operator distances which were not measured (see Table S4 for strain list). The fast Fourier transform algorithm of Matlab (version 2010a, The MathWorks, Inc.) was used to identify the dominant frequencies. The strongest peak corresponded to an approximately 11.3 bp period, as shown in Fig. S5(C). It has been shown previously that the period of oscillation for similar constructs was approximately 10-12 bp *in vivo* [9, 10, 12, 14, 16], in reasonable agreement with our measurement and for the number of base pairs in a single helical turn of relaxed double stranded DNA reported for other looping systems [17, 18].

We compare this period to the position of the peaks in our uncorrected looping energies, as shown in Fig. S5(D). The blue circles indicate the position of the peaks in the looping energy, and the numbers above indicate the number of base pairs between peaks. Peaks are spaced 12 bp apart at long distances, but at shorter distances the period becomes first 11 bp and then 5 bp. Similar trends have been previously attributed to multiple looped states corresponding to different orientations of the operators in the loop [11-13]. Independent of the particular interpretation, it is clear that there are peculiarities in the looping energies between 60 and 72 bp. First, the expected peak should fall around 61.5 bp, or 12 bp behind the peak at 73.5 bp, but is not found in the data. Second, the lowest looping energy is found at 70.5 bp, where the looping energy is much lower than the local minima of looping energies found at longer distances (as approximated by the red dotted line in Fig. S5(D)).

As a result of correcting for the extra effect of the auxiliary operator, the pattern of looping energy with length at short operator distances is dramatically changed, as shown in Fig. 4(C). In Fig. S6 we show this correction significantly changes the probability of looped and unlooped states at short operator distances. At some loop lengths accounting for the extra role of the auxiliary operator increased the inferred looping free energy by more than 2 k_BT with respect to the simpler model ($r_3/r_2(L)= 1$). Also correcting for direct repression from the upstream operator removes the anomalous trends in the looping energies observed at short distances prior to the correction.

Statistics

The reported means were weighted by the standard deviation of the measurement each day. Each day, 3 independent cultures of cells were grown. The average fluorescence intensity per cell was calculated by normalizing each measurement to OD_{600} and then averaging over the triplicate measurements using

$$\left(\frac{FI}{cell}\right)_{mean} = \frac{1}{N}\sum_{i=1}^{N}(FI_i - FI_{media})/(OD_{600,i} - OD_{600,media}). \quad \text{(S12)}$$

in which FI_i and $OD_{600,i}$ are the fluorescence intensity and optical density at 600 nm for sample i, and FI_{media} and $OD_{600,media}$ are the fluorescence intensity and optical density at 600 nm for the media, and N is the total number of replicate samples.

The average fluorescence intensity per cell was used to calculate the repression

$$repression_{mean} = \frac{(FI/cell)_{mean,\Delta lacI} - (FI/cell)_{mean,\Delta lacI,\Delta YFP}}{(FI/cell)_{mean,WTlacI} - (FI/cell)_{mean,WTlacI,\Delta YFP}}, \quad \text{(S13)}$$

in which ΔYFP strains were control cells not containing the YFP gene reporter used to account for the background fluorescence of the cells.

To calculate the standard deviation of a measurement we used

$$(\frac{FI}{cell})_{stdev} = [\frac{1}{N-1}\sum_{i=1}^{N}((\frac{FI}{cell})_i - (\frac{FI}{cell})_{mean})^2]^{0.5}. \quad \text{(S14)}$$

The standard deviation of *repression* for each set of measurements was calculated using

$$repression_{stdev} = repression_{mean} [[\frac{((FI/cell)^2_{stdev,\Delta lacI} + (FI/cell)^2_{stdev,\Delta lacI,\Delta YFP})^{0.5}}{(FI/cell)_{mean,\Delta lacI} - (FI/cell)_{mean,\Delta lacI,\Delta YFP}}]^2 + [\frac{((FI/cell)^2_{stdev,WTlacI} + (FI/cell)^2_{stdev,WTlacI,\Delta YFP})^{0.5}}{(FI/cell)_{mean,WTlacI} - (FI/cell)_{mean,WTlacI,\Delta YFP}}]^2]^{0.5}.$$

(S15)

Repression was averaged over multiple days' readings using a weighted mean and standard deviation. The weighted mean was calculated using

$$repression_{mean,weighted} = \frac{\sum_{j=1}^{M}(\frac{repression_{mean,j}}{repression^2_{stdev,j}})}{\sum_{j=1}^{M}(\frac{1}{repression^2_{stdev,j}})}, \quad \text{(S16)}$$

in which *j* denotes the repeat of the experiment on different days, a total and *M* repeats, and the mean and standard deviation of repression for each repeat are calculated using Eqs. S13 and S15. The mean for each day was weighted by its standard deviation.

The weighted standard deviation is calculated using,

$$repression_{stdev,weighted} = [\frac{\sum_{j=1}^{M}(\frac{(repression_{mean,j} - repression_{mean,weighted})^2}{repression^2_{stdev,j}})\sum_{j=1}^{M}(\frac{1}{repression^2_{stdev,j}})}{\sum_{j=1}^{M}(\frac{1}{repression^2_{stdev,j}})^2 - \sum_{i=1}^{M}(\frac{1}{repression^4_{stdev,j}})}]^{0.5}. \quad \text{(S17)}$$

For propagation of errors in other calculations, partial derivatives were used. For example, to calculate the contribution of the error in binding energy to the error in looping energy calculations we use,

$$\Delta F_{loop,error}(x_1, x_2, \ldots x_K) = \left[\sum_{k=1}^{K} ((\frac{\partial \Delta F_{loop}}{\partial x_k}) x_{stdev})^2 \right]^{0.5} \quad \text{, (S18)}$$

in which ΔF_{loop} is a function of parameters (x_1, x_2,…) and $\Delta F_{loop,error}$ is the total error from all *N* parameters. Partial derivatives were calculated using Mathematica (Wolfram Research, Champaign, IL).

Table S1: Host strains used in this study.

Host Strains	Genotype	Source or reference
E. coli		
HG104	MG1655 Δ*lacZYA*	[1]
HG105	MG1655 Δ*lacIZYA*	[1]

Table S2: Primers used in this study. Homology regions for integration primers are shown in bold.

Primers	Sequence	Comments
6.1	**TTCATATTGTTCAGCGACAGCTTGCTGTACGGCAGG CACCAGCTCTTCCG**GGCTAATGCACCCAGTAAGG	Integrate *lac* constructs into *galK*
6.3	**GTTTGCGCGCAGTCAGCGATATCCATTTTCGCGAAT CCGGAGTGTAAGAA**ACTAGCAACACCAGAACAGCC	Integrate *lac* constructs into *galK*
3.1	GTGCAATCCATCTTGTTCAATCAT	Sequence *lac* constructs in *galK*
3.2	CCTTCACCCTCTCCACTGACAG	Sequence *lac* constructs in *galK*
RemoveO2+11Fw	CTTCCGGCTCGTATAATGTGTGGGAATTCATTAAAGA GGAGAAAGGTACC	Remove operator O2 from +11
RemoveO2+11Rv	GGTACCTTTCTCCTCTTTAATGAATTCCCACACATTAT ACGAGCCGGAAG	Remove operator O2 from +11

Table S3: Operator and promoter sequences.

Region	Sequence
Upstream of auxiliary operator	AGCCATCCAGTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCCCAGCTGGCAATTCCGACGTC
Auxiliary operator (Oid)	AATTGTGAGCGCTCACAATT
Promoter	TTTACAATTAATGCTTCCGGCTCGTATAATGTGTGG
Main operator (O2)	AAATGTGAGCGAGTAACAACC
Main operator (O1)	AATTGTGAGCGGATAACAATT
Downstream of main operator	AATTCATTAAAGAGGAGAAAGGTACCGCATGCGTAAAGGAGAAGAACTTT

Underlined portion indicates the beginning of the YFP coding region.

Table S5: Parameters used in calculations.

Parameter	Value	Error	Units	Reference	Parameter value depends upon thermodynamic model?
N_{NS} (estimated number of non-specific binding sites, number of base pairs in genome of *E. coli* K12)	4.6×10^6	-	base pairs	GenBank: U00096.2	no
operator binding energies					
Oid	-17	0.2	k_BT	[1]	yes
O1	-15.3	0.2	k_BT	[1]	yes
O2	-13.9	0.2	k_BT	[1]	yes
promoter binding energy					
$\Delta\varepsilon_{pd}$	-5	-	k_BT	[1]	yes
number of Lac repressor molecules per cell					
R for WT	11	2	-	[1]	no
R for RBS1147	30	10	-	[1]	no
R for RBS446	62	15	-	[1]	no
R for RBS1027	130	20	-	[1]	no
R for RBS1	610	80	-	[1]	no
R for 1l	870	170	-	[1]	no
number of RNA polymerase molecules per cell					
P	2000	-	-	[5]	no

Table S6: Fits to Lac repressor titration curves in Fig. 2(A) and Fig. S2.

	weighted fits (in units k_BT)				unweighted fits(in units k_BT)			
Sequence	O2 binding energy	SE	ΔFloop	SE	O2 binding energy	SE	ΔFloop	SE
E73.5	-13.8	0.1	11.3	0.2	-13.8	0.8	11.2	2.3
E81.5	-14.0	0.1	8.6	0.1	-14.3	0.3	8.2	0.5
E135.5	-13.9	0.1	8.8	0.1	-14.3	0.7	8.0	0.7

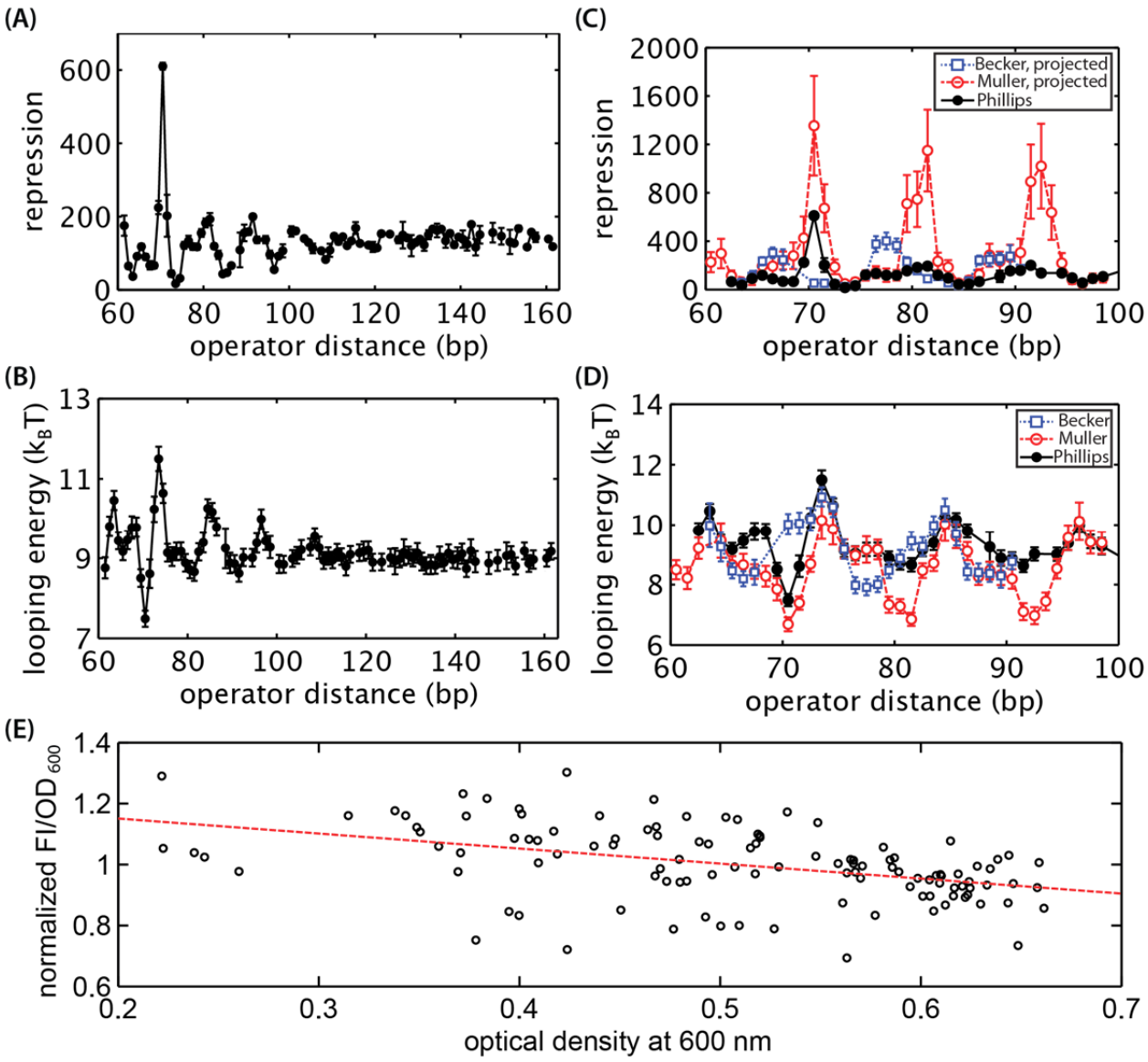

Figure S1: Repression and the free energy change of loop formation for many loop lengths and comparison to two previous studies. A) Repression for operator distances between 60.5 and 161.5 bp is shown for constructs containing the main operator O2 and the auxiliary operator Oid, and 11 repressors per cell. Subsets of this data set are used throughout the main text. (B) By applying the equation in Fig. 1(C), the looping free energy was extracted from the repression data for each operator distance. Error bars in A and B represent standard error. (C) Comparison of our repression data to similar DNA looping constructs measured in previous works by Muller *et al.* and Becker *et al.* [9, 10]. Although these experiments were measured under different conditions (operator strengths and number of repressors) we "project" their repression values onto our experimental conditions as described in the supplemental materials. (D) Direct comparison of the looping free energies from the data sets shown in (C). Error bars for Becker and Muller data represent standard deviation. (E) The ratio of fluorescence intensity (FI) to OD_{600} (OD) measured 114 times on 39 different days for the control strain HG105 O2+11. A best fit line to all the data reveals a trend of decreasing FI/OD ratio with optical density. The slope of

the best fit line (red dashed line) is -0.49 FI/OD_{600} per OD_{600} (95% confidence bounds -0.68 to -0.30), which predicts that over the range of OD_{600} used in experiments the FI/OD should change by ~20%, similar to the coefficient of variation for replicate measurements harvested at similar optical densities. Data is normalized to the mean of the data set.

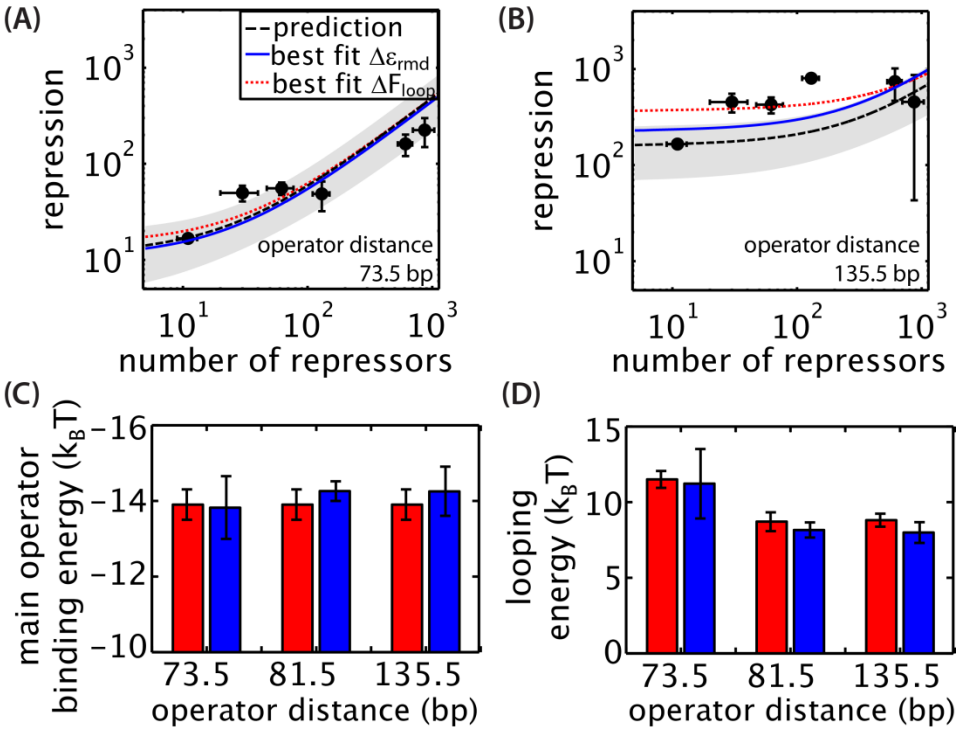

Figure S2: Titrating the number of repressors per cell for operator distances 73.5 and 135.5 bp and global fits for model parameters. (A-B) Predicted and measured repression for Oid-O2 constructs as a function of number of repressors per cell. The dashed black lines represent the zero parameter prediction, and shaded regions define 95% confidence interval for these predictions. Global fits to all data points are shown for main operator binding energy (blue solid line) and looping energy (red dotted line). Points are experimental data with error bars representing the standard error. (C) Global fits result in values for the main operator binding energy, blue bars, consistent with previously reported values, red bars using values from Table S5. (D) Global fits result in values for the looping energies, blue bars, similar to values calculated from repression measurements with the wild type number of repressors, red bars. Error bars are 95% confidence interval. Data for operator distance 81.5 bp is found in Fig. 2(A). See Table S6 for fit values, including fits when weighting each data point by its error.

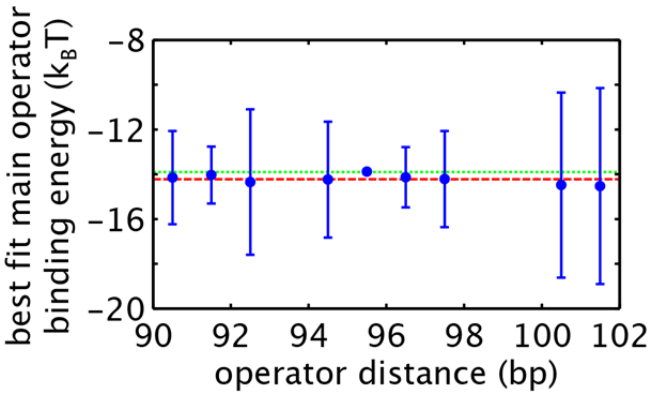

Figure S3: Calculating the best fit main operator binding energy from the data from Figure 2(B). At every operator distance in Fig. 2(B), the best fit main operator binding energy was calculated using the pair of data points corresponding to 11±2 and 130±20 repressors per cell. Data was fit using the equation in Fig. 1(C) and the parameters listed in Table S5. The red dashed line represents the mean best fit main operator binding energy, -14.2 k_BT. This differs slightly from the value of -15.1 k_BT reported in the Fig. 2(B), which was fit using only the points measured at 130 repressors per cell. The green dotted line represents the previously reported binding energy from Table S5. Error bars are 95% confidence interval of the fit.

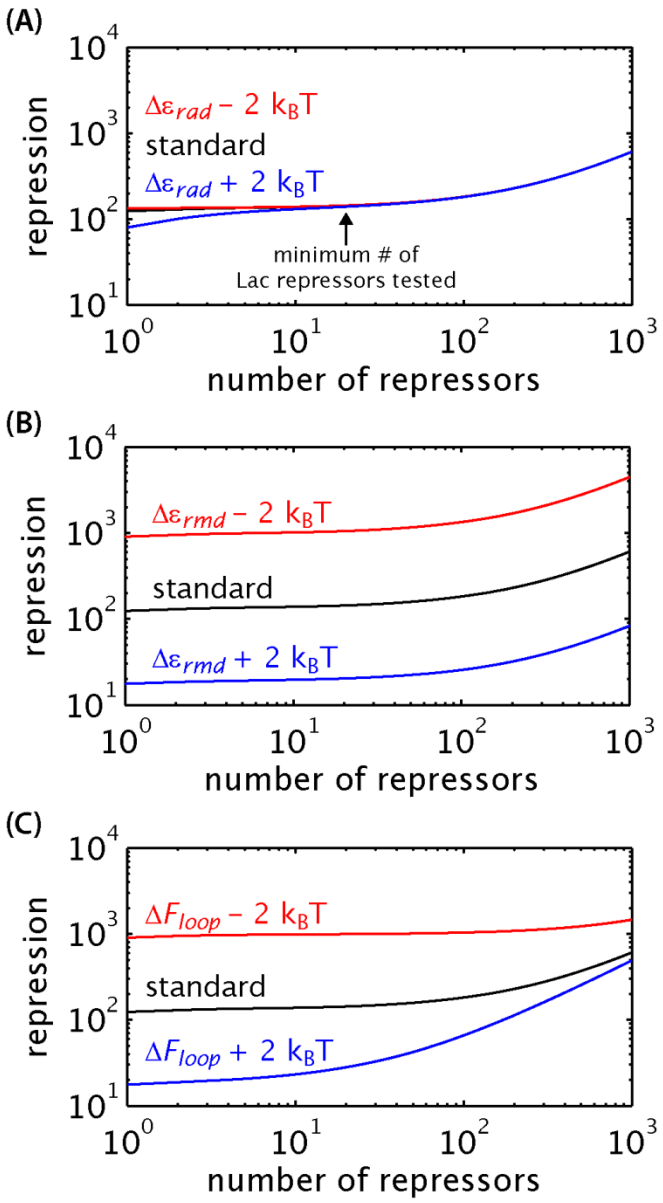

Figure S4: Predicted sensitivity of repression to the parameters $\Delta\varepsilon_{rad}$, $\Delta\varepsilon_{rmd}$, and ΔF_{loop} as the number of repressors per cell is increased from 1 to 1000. Predictions were made using the equation in Fig. 1(C), with Oid as the auxiliary operator, O2 as the main operator, and $r_3/r_2 = 1$. Standard (black lines) refers to calculations using the parameters found in Table S5 and ΔF_{loop} = 9 k_BT. Plots show the result of adding (blue lines) or subtracting (red lines) 2 k_BT to one of the energies in the standard condition: (A) Changing the auxiliary operator binding energy ($\Delta\varepsilon_{rad}$), (B) changing the main operator binding energy ($\Delta\varepsilon_{rmd}$), and (C) changing the looping energy (ΔF_{loop}). Repression is strongly dependent on the main operator binding energy over the range of repressor numbers measured in experiments. Repression is also sensitive to changes in the looping energy in this regime, especially decreases in the looping energy, but becomes less sensitive in the limit of high repressor number. It was found that repression is not sensitive to the auxiliary operator binding energy over the range of repressors per cell measured in Fig. 2(A) and Fig. S2, hence global fits to the auxiliary operator binding energy are not reported.

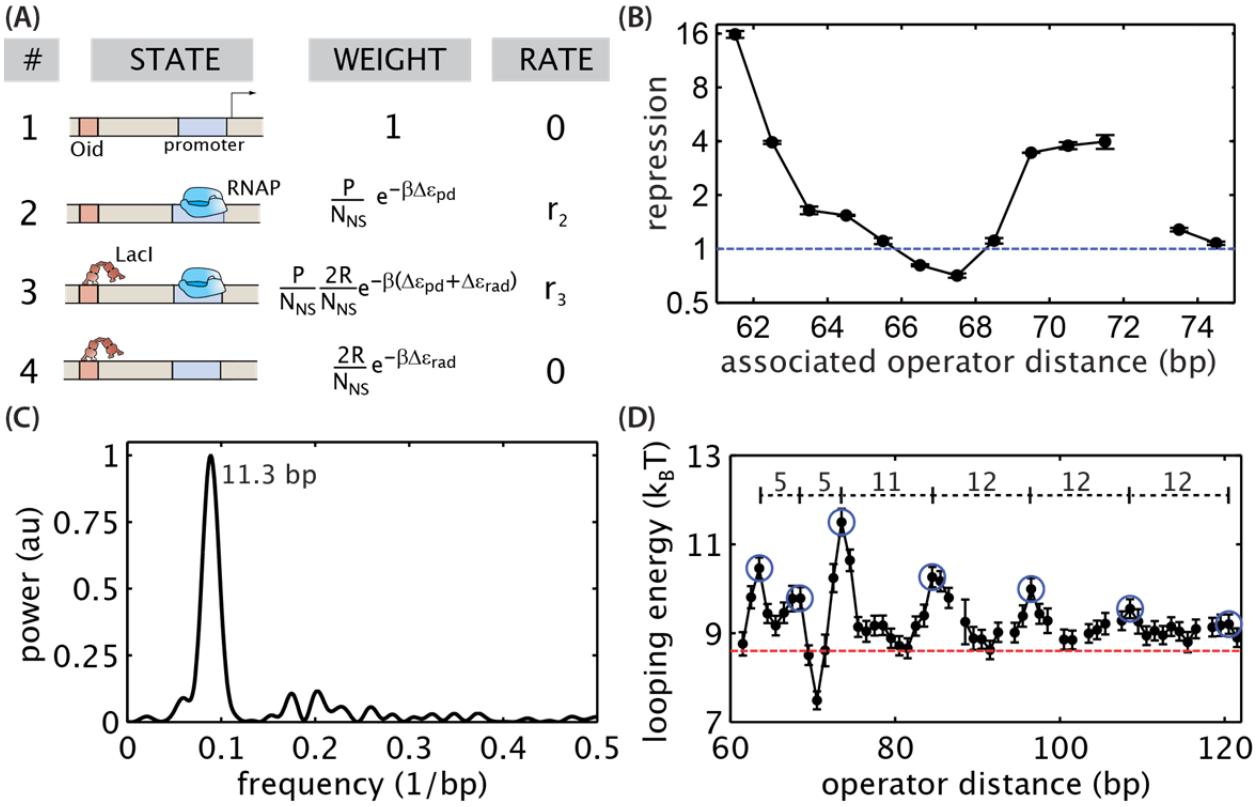

Figure S5: Correcting for direct upstream repression when calculating looping energies. (A) List of states and weights for repression by Lac repressor binding to an upstream auxiliary operator. (B) An oscillatory pattern of repression over operator distance is observed for constructs containing only the auxiliary operator, plotted on a semilog scale. The associated operator distance is the operator distance corresponding to a looping construct with the same length of DNA between the auxiliary operator and the promoter. The blue dashed line shows a repression of 1. A value of less than 1 indicates activation. (C) Fourier analysis of the repression vs. operator distance data shown in Fig. S1(A) revealed a peak in the oscillatory frequencies corresponding to 11.3 bp. (D) The position of the peaks in looping energy, shown in blue circles, are 11-12 bp apart at long operator distances, but the spacing is reduced at short operator distances. Numbers indicate the number of base pairs between peaks. The red dotted line approximates the positions of the local minima for operator distances between 72 and 120 bp. Error bars represent standard error.

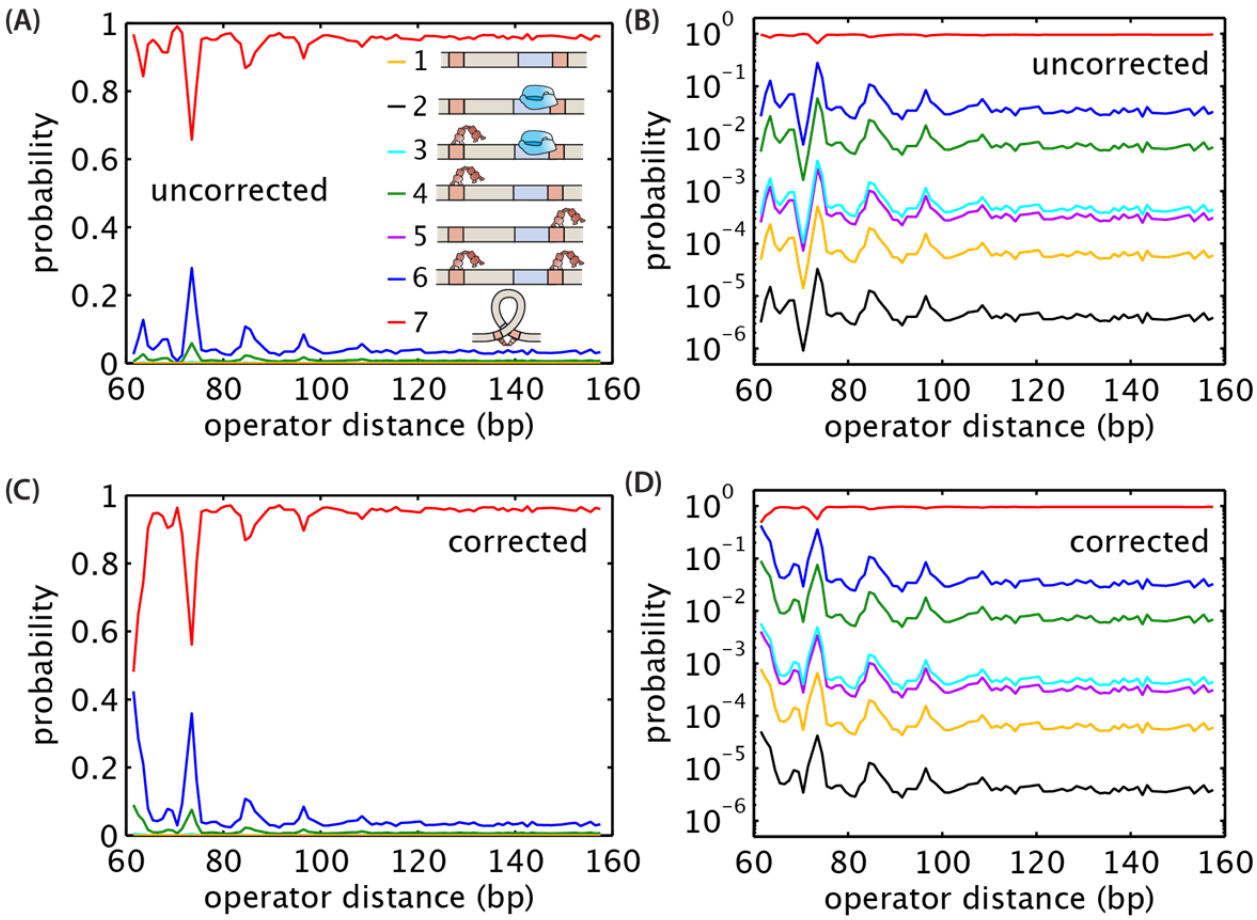

Figure S6: Changes to the probabilities of states for looping constructs after correcting for direct gene regulation by the upstream operator. Predicted distribution of the probabilities for each state of the promoter shown in Fig. 1(B) before (A,B) and after (C,D) correcting for the influence of the upstream operator on gene regulation. B and D are plotted on semilog axes in order to show the low probability states. All 7 states are shown in both plots, although in A and C only the three most probable states can be discerned on the graph. Plots were made using the weights listed in Fig. 1(B), the wild type number of repressors, a main operator of O2, an auxiliary operator of Oid, and parameters listed in Table S5. The numbers in the legend of (A) correspond to the state numbers listed in Fig. 1(B).

Supplemental References:

[1] H. G. Garcia, and R. Phillips, Proc Natl Acad Sci U S A **108**, 12173 (2011).
[2] T. Kuhlman *et al.*, Proc Natl Acad Sci U S A **104** (2007).
[3] H. G. Garcia *et al.*, Biophys J **101**, 535 (2011).
[4] L. Bintu *et al.*, Current Opinion in Genetics & Development **15**, 125 (2005).
[5] M. Jishage, and A. Ishihama, J Bactiol **177**, 6832 (1995).
[6] I. L. Grigorova *et al.*, Proc Natl Acad Sci U S A **103**, 5332 (2006).
[7] S. Johnson, M. Lindén, and R. Phillips, Nucleic Acids Res (2012).
[8] L. Han *et al.*, PLoS ONE **4**, e5621 (2009).
[9] N. A. Becker, J. D. Kahn, and L. J. Maher, Journal of Molecular Biology **349**, 716 (2005).
[10] J. Muller, S. Oehler, and B. Muller-Hill, J Mol Biol **257**, 21 (1996).
[11] L. Saiz, and J. M. G. Vilar, PLoS ONE **2**, e355 (2007).
[12] Y. Zhang *et al.*, PLoS ONE **1**, e136 (2006).
[13] D. Swigon, B. D. Coleman, and W. K. Olson, P Natl Acad Sci USA **103**, 9879 (2006).
[14] L. Saiz, J. M. Rubi, and J. M. Vilar, Proc Natl Acad Sci U S A **102**, 17642 (2005).
[15] Hernan G. Garcia *et al.*, Cell Reports **2**, 150 (2012).
[16] S. M. Law *et al.*, J Mol Biol **230**, 161 (1993).
[17] D. H. Lee, and R. F. Schleif, Proc Natl Acad Sci U S A **86**, 476 (1989).
[18] R. Amit *et al.*, Cell **146**, 105 (2011).