Cell

Building Enhancers from the Ground Up: A Synthetic Biology Approach

Roee Amit,[1,3,4,*] Hernan G. Garcia,[2] Rob Phillips,[1,3] and Scott E. Fraser[1,3]
[1]Division of Biology
[2]Department of Physics
[3]Division of Engineering and Applied Science
California Institute of Technology, Pasadena, CA 91125, USA
[4]Present address: Department of Biotechnology and Food Engineering, Technion, Haifa 32000, Israel
*Correspondence: roeeamit@technion.ac.il
DOI 10.1016/j.cell.2011.06.024

SUMMARY

A challenge of the synthetic biology approach is to use our understanding of a system to recreate a biological function with specific properties. We have applied this framework to bacterial enhancers, combining a driver, transcription factor binding sites, and a poised polymerase to create synthetic modular enhancers. Our findings suggest that enhancer-based transcriptional control depends critically and quantitatively on DNA looping, leading to complex regulatory effects when the enhancer cassettes contain additional transcription factor binding sites for TetR, a bacterial transcription factor. We show through a systematic interplay of experiment and thermodynamic modeling that the level of gene expression can be modulated to convert a variable inducer concentration input into discrete or step-like output expression levels. Finally, using a different DNA-binding protein (TraR), we show that the regulatory output is not a particular feature of the specific DNA-binding protein used for the enhancer but a general property of synthetic bacterial enhancers.

INTRODUCTION

A classic view of transcriptional regulation in bacteria is built around the idea of regulated recruitment of RNA polymerase and the dissociable sigma factor σ^{70}. In this picture, the presence or absence of RNA polymerase at a promoter of interest is dictated by the corresponding presence or absence of batteries of transcription factors that either increase (activators) or decrease (repressors) the probability of polymerase binding. An increasingly sophisticated understanding of this kind of regulatory response has resulted in an explosion of efforts in synthetic and systems biology built using a broad palette of different activators and repressors for a range of different promoters of this kind (Bintu et al., 2005b; Elowitz and Leibler, 2000; Gardner et al., 2000; Joung et al., 1993; Müller et al., 1996; Mukherji and van Oudenaarden, 2009 and references therein).

Another whole set of bacterial promoters utilize an alternative sigma factor (σ^{54}) that, together with RNAP, forms a stable closed promoter complex that, unlike its σ^{70} counterpart, is unable to initiate transcription by itself (Buck et al., 2000; Rappas et al., 2007). This effectively causes the polymerase to be poised at the gene of interest, awaiting the arrival of a transcription factor partner that we term the "driver," which releases the polymerase. Consequently, these promoters are regulated in a different fashion than their recruitment counterparts. The activating or transcription driving complex is typically widely separated from the promoter (100–1000 bp) (Ninfa et al., 1987), precluding it from forming direct contact with the poised polymerase. It has been asserted (Huo et al., 2006; Schulz et al., 2000; Su et al., 1990) that DNA looping and ATP hydrolysis are required to induce open complex formation and transcription initiation (Rappas et al., 2007). These regulatory regions belong to a different class of regulatory elements called enhancers, which are more commonly associated with eukaryotic organisms. On its own, a poised promoter has the capability to execute little or no transcriptional regulation, but together with enhancers, they can express their full regulatory potential (Davidson, 2001; Magasanik, 1993).

Enhancer elements are ubiquitous in genomes from all domains of life (Buck et al., 2000; Ninfa and Atkinson, 2000; Rappas et al., 2007). It is hypothesized that enhancers execute their regulatory program by making direct contact with the basal promoter via DNA or chromatin looping. In general, they are made up of contiguous genomic regions that stretch from tens to thousands of base pairs and contain several binding sites for a variety of transcription factors (TF); often, their regulatory output is independent of their location or orientation relative to the basal promoter (Driever et al., 1989; Huo et al., 2006; Ninfa et al., 1987). As a result, enhancers, like gene regulatory networks themselves, can be viewed qualitatively (Davidson, 2006) as modular genomic entities made of three connected irreducible parts: the driver-binding sites responsible for initiation of transcription, protein-binding sites within the enhancer that are responsible for the regulation or modulation of expression levels, and the poised promoter. Whereas other aspects of gene regulation are becoming better defined (e.g., the input/output relationship between different genes in gene regulatory networks) (Bintu et al., 2005a; Kuhlman et al., 2007; Rosenfeld et al., 2005; Garcia

and Phillips, 2011), the underlying mechanisms of regulatory "action at a distance" that are responsible for integrating the various inputs in enhancers remain poorly understood.

To explore the kinds of action at a distance mechanisms that can yield complex regulatory behavior associated with enhancers, we opted to construct synthetic enhancers de novo. In this case, the synthetic approach permits us to systematically construct enhancers in a modular fashion, starting with a minimal enhancer made of driver-binding sites and the poised promoter region and progressively increasing the synthetic enhancer's complexity with the addition of discrete sets of defined enhancer-binding protein-binding sites (TetR or TraR in our case) that are not thought to interact directly with either the driver protein or the poised RNA polymerase. The synthetic approach provides us with an experimental foundation that can be utilized to develop thermodynamic models in which the various states of occupancy of the promoter and their associated statistical weights can be computed and used to explore the enhancer's regulatory output.

We hypothesized that a rich interplay between experiment and theory would not only allow us to increase our predictive capability with respect to enhancer regulatory output, but also tease out the underlying mechanisms for regulatory action at a distance by ensuring that the model and experiment be consistent at every stage of the cascade. At each experimental stage, when an increasingly complex set of regulatory architectures was characterized, the starting point for the theoretical description was the model utilized to describe the more simplified constructs explored during the previous step. Thus, throughout the paper, we will repeatedly resort to thermodynamic models, which exploit equilibrium statistical mechanics to serve as a conceptual framework for all of the experiments.

RESULTS

Expression Levels Are Controlled by DNA Looping

We selected the bacterial NRI/NRII (NtrC/NtrB) two-component system (Magasanik, 1993), controlling nitrogen assimilation in many prokaryotes, to test our methodology. We constructed minimal enhancers using driver-binding sites for the phosphorylated DNA-binding isoform of NRI (NRI∼P) and coupled them to a poised σ^{54} promoter with a DNA linker of varying length. The dimeric NRI∼P proteins assemble on the DNA to form a hexameric complex, which in turn functions as the transcriptional driver in our system. An mCherry reporter was used to measure the transcriptional activity of this promoter (for circuit details, see "Theory: Model for Looping Initiated Transcription" in the Supplemental Information available online).

We reasoned that systematically varying the length of the DNA sequence between the driver-binding sites and σ^{54} promoter will yield an expression pattern that depends on the length of the looped DNA and on the phasing of the complex (the orientation of the driver with respect to the polymerase bound to the promoter that depends on the DNA helical periodicity) in much the same way that phasing impacts expression levels in different looping regulatory contexts (Law et al., 1993; Lee and Schleif, 1989; Müller et al., 1996). In order to check the validity of this assumption, we cloned into the spacer region of the synthetic enhancer 65 distinct DNA sequences (Table S1 and Table S2) of variable length (28–315 bp; Figure 1A and Figure S1A). We carried out fluorescence measurements in bulk while the strains were growing in midlog phase and subsequently normalized the fluorescence levels obtained for each strain to the value measured for the maximally expressing strain (L = 70 bp).

At first glance, the results shown in Figure 1B seem to exhibit a strongly fluctuating behavior with a nontrivial dependence on looping length (*L*). However, a useful framework for considering this complex data is provided by the thermodynamic model schematized in Figure 1C, which invokes a model inspired by the underlying DNA biophysics of looping, transcriptional mechanics, and equilibrium thermodynamics (see "Theory: Model for Looping Initiated Transcription" in the Supplemental Information, Figure S2, and Figure S3). The essence of the model depicted in the figure is that there are two states of interest, both of which have the $(NRI\sim P)^6$ hexamer and RNA polymerase (RNAP) bound but only one of which is looped and transcriptionally active. The looped state is weighed by a looping J factor (a measure for the local concentration of the hexamer in the vicinity of RNAP) and a dissociation constant between the $(NRI\sim P)^6$ hexamer and RNA polymerase. To simplify the interpretation of the results, we collapse the looping J factor and the dissociation constant by defining the ratio J/K_{nr} as the looping capacity $\chi(L)$.

The model generates a fit that rises rapidly for L < 70, slowly declines for L > 70 (light blue dashed lined), and is modulated by a characteristic periodicity of 11.0 ± 0.1 bp. This value for the periodicity likely corresponds to the helical period of the DNA itself and is in good agreement with previous measurements (Becker et al., 2005; Law et al., 1993; Lee and Schleif, 1989; Müller et al., 1996). It is worth noting that, whereas the error to the fit of the periodicity exhibited by our data is low, the rest of the parameters, which characterize the looping capacity function, cannot be determined to a high level of certainty. As shown in Figures S2A and S2B, various candidates for the looping capacity function can generate plausible envelope functions for the data, as shown by the red line in Figure 1B and Figure S2B. Discriminating between alternative looping capacity functions would require data from larger loop lengths than those obtained here.

Enhancer Repression Is a Bimodal Function of Spacer Length

Given that the level of transcription depends critically upon DNA looping, we reasoned that, by installing binding sites for other transcription factors within the looped region, we might tune the propensity for loop formation and hence the level of expression by controlling the concentration of the active transcription factors. We suspected that one possible way of generating this effect was by making the intervening DNA more rigid though the binding of a common repressor TetR, whose binding to DNA is thought not to induce long-range deformations (Ramos et al., 2005 and references therein). This, in turn, would lead to an inhibition of the looping process, which would result in the repression of the synthetic enhancer circuit, yielding a reduction in the quantity of the fluorescent reporter.

In order to test this assertion, we added cassettes to the synthetic enhancer containing one, two, three, or six binding sites for TetR. The cassettes were cloned 28 bp downstream of the

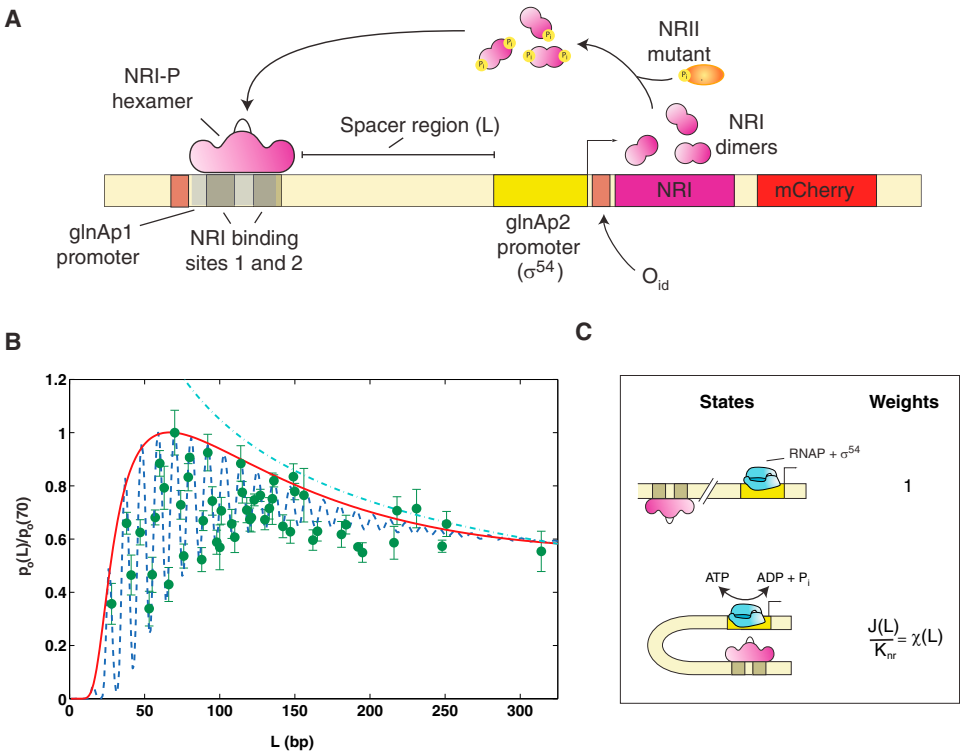

Figure 1. Enhancer Activation Depends Strongly on Looping

(A) Schematic for synthetic enhancer circuit. In short, the circuit expresses via a σ^{54} promoter the *glnG* (*ntrC*) gene, whose protein product (NRI) remains phosphorylated at all times via the action of the phosphatase-deficient mutant NRII2302 (Atkinson et al., 2003), which also serves to decouple the NRI/NRII system from the nitrogen assimilation pathway. The synthetic enhancer circuit was transformed into a *ΔGlnL:ΔGlnG:3.300 E. coli* strain (3.300LG) on a low-copy plasmid (≅10/cell).

(B) Relative fluorescence level $p_o(L)/p_o(70)$ versus looping length L data (green circles). For each looping length, $p_o(L)/p_o(70)$ is defined as the ratio between the measured fluorescence level of the synthetic enhancer strain to the fluorescence level of the brightest strain (L = 70 bp, the natural glnAp2 enhancer looping length). The fits correspond to our expression model with (blue dashed line) and without (red line) the periodic modulation (see "Theory: Model for Looping Initiated Transcription" in the Supplemental Information for more details). The light-blue dashed line corresponds to a fit by an empirical power-law decay curve of power −1/2. Error bars correspond to the standard deviation from multiple measurements.

(C) Schematic Model for enhancer-activated transcription for our constructs, which requires ATP hydrolysis and DNA looping to bring the driver/activator protein complex into physical contact with the "poised" σ^{54}-RNAP complex.

See also Figure S1, Figure S2, and Figure S3.

NRI#2 binding site (Figure 2A) to ensure that the first TetR (Hillen et al., 1984) does not interfere with the binding of the NRI∼P complex (Hervás et al., 2009). This isolates the repression effects to a modification of the looping capacity function when TetR is present, the description of which is developed in "Theory: Model for Looping Initiated Transcription" in the Supplemental Information. The extent of repression for each cassette was quantified by measuring the fluorescence of the reporter both in the presence of a high number of TetR proteins and in their absence. In Figure 2B, we plot repression values as a ratio of the repressed to the unrepressed fluorescence levels for each synthetic enhancer circuit as a function of the looping DNA length (as defined in Figure 2A). The figure shows the experimental data for the 1-Tet (one TetR-binding site), 2-Tet, and 3-Tet cassettes. For all cassettes used in the experiment, the data show a signature for bimodality with either strong repression for synthetic enhancer lengths $L < L_t$ or weak repression for lengths $L > L_t$. The length L_t, which serves as a DNA length scale setting a sharp transition between the two repression regimes, varies for each cassette type (labeled as L_{t1}, L_{t2}, and L_{t3} on the plots) and seems to depend systematically on the number of binding sites and the size of the binding region of TetR (Hillen et al., 1984).

In order to understand the bimodal behavior, it is instructive to consider the short and long loop length limits. For short loop lengths, one simple interpretation is that the DNA-TetR complex behaves like a "rigid" nucleoprotein complex with an effective persistence length longer than that of bare DNA. Alternatively, the heightened repression at short looping lengths could be due to some other biophysical mechanism that promotes TetR-induced interference with the ability of the NRI∼P complex to loop. Either way, for $L < L_t$, looping is far less likely to take place, and the RNAP will remain poised.

For long loop lengths, wherein the weak repression regime is observed, the rigidification effect hypothesized for lower lengths is diminished. In this regime, the data indicate that repression levels are weakly dependent on the loop length and the synthetic

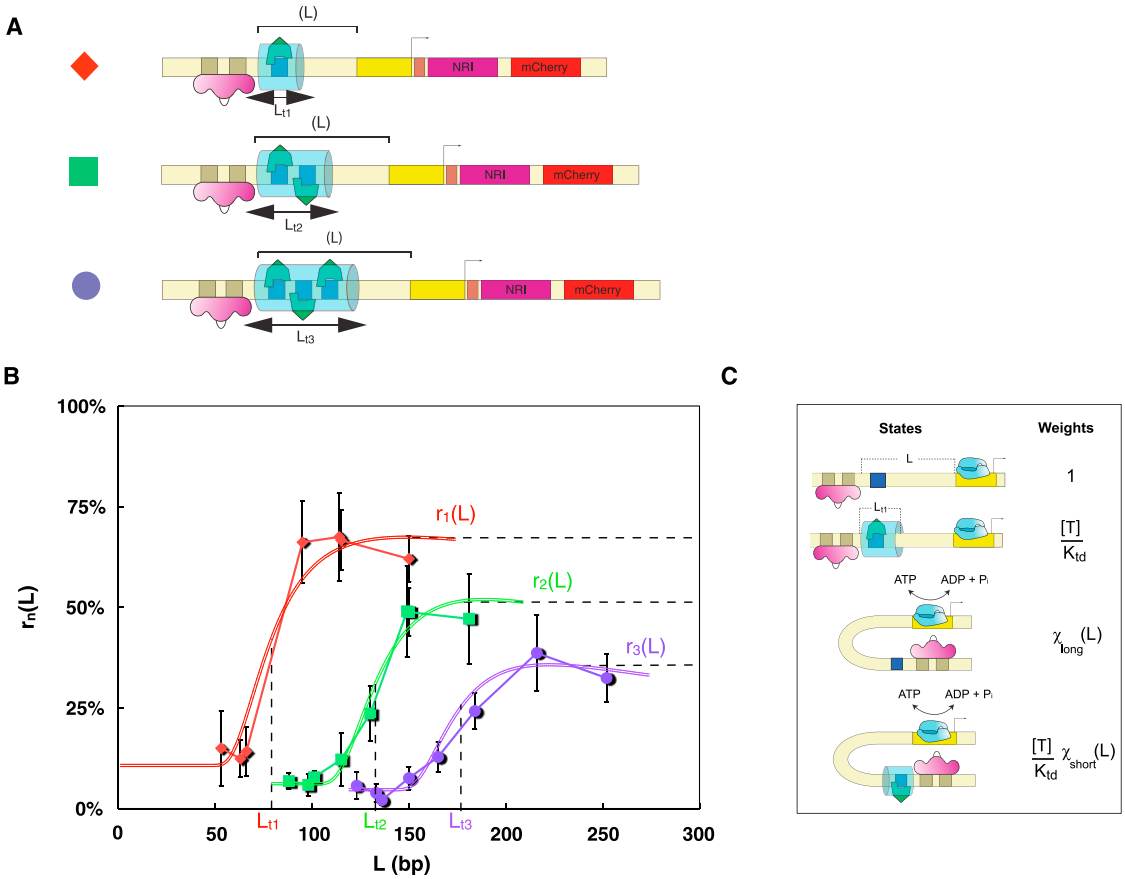

Figure 2. Bimodal Repression of Enhancer-Based Transcription

(A) Schematic showing the constructs used to study enhancer repression containing 1-, 2-, and 3-TetR-binding sites, respectively. The binding sites for TetR are positioned 28 bp upstream of the NRI#2 site and with 16 bp spacing for the 2- and 3-Tet cassettes. The TetR-rigidified region of the spacer DNA (denoted by light-blue shade and L_{t1}, L_{t2}, and L_{t3}) is hypothesized to be the mechanism responsible for repression.

(B) Expression data exhibiting bimodal behavior for the 1-Tet (red diamonds), 2-Tet (green squares), and 3-Tet (purple circles) cassettes. The data are depicted as percent relative to the unrepressed expression levels for the 1-, 2-, and 3-Tet cassettes, respectively. r_1(L), r_2(L), and r_3(L) levels correspond to the repression functions as defined in "Theory: Model for Looping Initiated Transcription" in the Supplemental Information (Equations S27–S29). The values of these functions at particular lengths are used as input to the model and fits (for the data shown in Figure 3, Figure 4, Figure 6, and Figure S5). The colored curved double lines for each data set correspond to model fits (see Figure S4 for additional detail), and the dashed lines correspond to the length-independent repression value that each cassette seems to approach. Error bars correspond to the standard deviation from multiple measurements.

(C) States and weights schematic for the model used to describe the 1-Tet repression data. The two additional states correspond to the looped and unlooped configurations of the DNA with TetR bound to the enhancer.

See also Figure S4.

enhancer's orientation relative to the promoter. Moreover, repression levels observed for weakly repressed synthetic enhancer circuits reflect the number of TetR-binding sites on the cassette by yielding discretely separated values for each cassette type. This is highlighted by the colored lines, which denote each of colored data sets representing the repression functions $r_1(L)$, $r_2(L)$, and $r_3(L)$ (see "Theory: Model for Looping Initiated Transcription" in the Supplemental Information for the functional form of these terms) on the graph.

To understand the origins of regulation at a distance in our synthetic enhancer system, the thermodynamic model framework tells us how to go beyond the two-state description introduced in Figure 1. In particular, we have to account for all of the different states of occupancy in which TetR can be bound to the DNA looping region. To that end, we add an additional set of states to our thermodynamic modeling framework, which provides a convenient scheme for characterizing the different states of the promoter and their relative probabilities. As shown in Figure 2C for the 1-Tet case, the model now has four states that come in two broad categories: unlooped and inactive and looped and active, each with and without TetR bound. Unfortunately, our knowledge of the geometric details of the loops in the repressed case (i.e., when the cassette is bound by TetR proteins) is too meager to adopt a "first principles" approach, which would allow us to relate the looping capacity in the presence of TetR to the looping capacity in its absence. As a result, the states and weights are still written in terms of the looping capacity, but now the looping capacities themselves are undetermined parameters.

However, for the long looping length limit ($L >> L_t$), simple polymer models can be used to develop intuition for the resulting repression (Phillips et al., 2009). Using these theoretical results and the model presented in "Theory: Model for Looping Initiated Transcription" in the Supplemental Information, we can derive an expression for long-distance repression that is a ratio of the repressed to the unrepressed looping capacity functions (Equations S23 and S24), which converges to a fixed value and gives a sense of the theoretical underpinnings for r(L). Consequently, at the very long loop length limit, both the model and experiment indicate that these repression values (denoted by the dashed lines) seem to converge on a particular constant for each cassette configuration, rather than approach the nonrepressed value of 100%.

Using the long-looping length limit and the repression values observed for the strong repression regime, we can approximate the functional form of the repressed looping capacity functions (fits in Figure 2B) for each cassette using the same functional form exploited earlier. Using these functions, the data can be compactly represented by a simple function that is consistent with both the transition lengths (L_t) and the saturation values that appear to be correlated with the number of TetR-binding sites and the distance between the beginning of the NRI#1 site and the last TetR-binding sites (see Figure S4).

Multiple TetR-Binding Sites Generate Step Functions from a Variable Input

The long-range repression capability of our synthetic enhancer system discussed above has further regulatory potential. This observation suggests a design strategy for constructing synthetic enhancers. By tuning the concentrations of an input signal, which alters the binding probability of the regulatory proteins, the level of gene expression can, in turn, be systematically tuned between different discrete values. In the case of TetR, this can be done simply by titrating variable amounts of a soluble ligand anhydrous-Tetracycline (aTc), which prevents the binding of TetR to its binding site by inducing a conformational change (Orth et al., 2000).

We studied the regulatory output of four different types of binding site cassettes—1-, 2-, 3-, and 6-Tet—in response to the variable input signal. In order to compare the output functions for the different cassettes, we plot the data (Figure 3) by constructing a ratio of the fluorescence level measured in the presence of a given ligand concentration divided by the maximal average fluorescence level (i.e., when the cassette is most likely unoccupied by TetR at saturating concentrations of aTc; labeled 100% on the plots).

In Figure 3A, the regulatory function for the 2-tet cassette is presented. We observe a response that exhibits three discrete values of expression: a repressed state, a sharp transition at ≅10 ng/ml aTc to an intermediate partially repressed level, and a final transition at ≅200 ng/ml aTc to an unrepressed expression level.

In order to understand the intermediate expression level of the regulatory output function, we constructed two additional synthetic enhancers. These enhancers were constructed with identical looping lengths to the 2-Tet enhancer and contain only a single binding site for TetR at either the distal or proximal binding site location of the 2-Tet construct. Examination of Figure 3B shows that the weak repression level ($r_1(L)$) measured for the single binding site cassettes is in reasonable accord with the intermediate level of the repression ratio in Figure 3A and with the weak repression regime for the 1-Tet cassette repression data (Figure 2B). Therefore, it is likely that the intermediate level observed for the 2-Tet enhancer reflects the partial TetR occupancy configuration (Figure S4A) for the two-binding site architecture.

The regulatory output function for the 3-Tet cassette shown in Figure 3C also exhibits a series of discrete expression levels. In particular, this case is characterized by four values: a fully repressed state and a sharp transition at 10 ng/ml to a set of three nearby expression levels that are located at values of roughly 70%–80%, 90%, and 100%, respectively. Alternatively, one may choose a more conservative interpretation of the data shown in Figure 3C as having a single intermediate level at ≅70%–80% and a shallow increase to 100% repression ratio thereafter.

The 3-Tet output function can be understood qualitatively using similar logic to that introduced in thinking about the 2-Tet cassette regulatory function. For this case (Figure 3D), there is one configuration for full occupancy, one for an unoccupied state, and three configurations each for single and double occupancies. To show that the steps shown in Figure 3C reflect these partial occupancy states, we measured the repression values for six additional cassettes that account for all possible occupancy configurations (Figure 3D). We found that only the triply occupied configuration is strongly repressed, whereas the other configurations are weakly repressed with values of 40%–45% and 60%–80% of full expression for double and single occupancy, respectively, thereby supporting the idea that the discrete jumps in the repression ratio levels are associated with either the single or double occupancy configurations. Interestingly, the repression ratio value of the first (and perhaps only) intermediate coincides approximately with the average repression level ($r_1(L)$; purple shade; Figure 3D) of the three single occupancy configurations. This indicates that the dominant state at these aTc concentrations is the single occupancy configuration.

The next step in the progression of increasingly complex enhancer architectures corresponds to a case with six TetR-binding sites. The regulatory output function (Figure 3E) does not exhibit an increase in the number of intermediates but instead is characterized by two intermediates with more evenly spaced repression ratio values and with sharper transitions that produce a more distinct step-like function than for the 2- and 3-Tet cassettes (see also Figure 4). Here, the first intermediate repression ratio state is located at 65% of the unoccupied cassette maximum and the second at 75%–80% of the maximum. These values are markedly different from the 80% and 90% values that were measured for the 3-tet cassette.

Combinatorial Control in a Synthetic Enhancer

Examining the data for the 1-, 2-, 3-, and 6-binding site cassettes more closely, we find additional regulatory features that likely would not have been guessed a priori. The dose-response for each TetR cassette type indicates that the transition

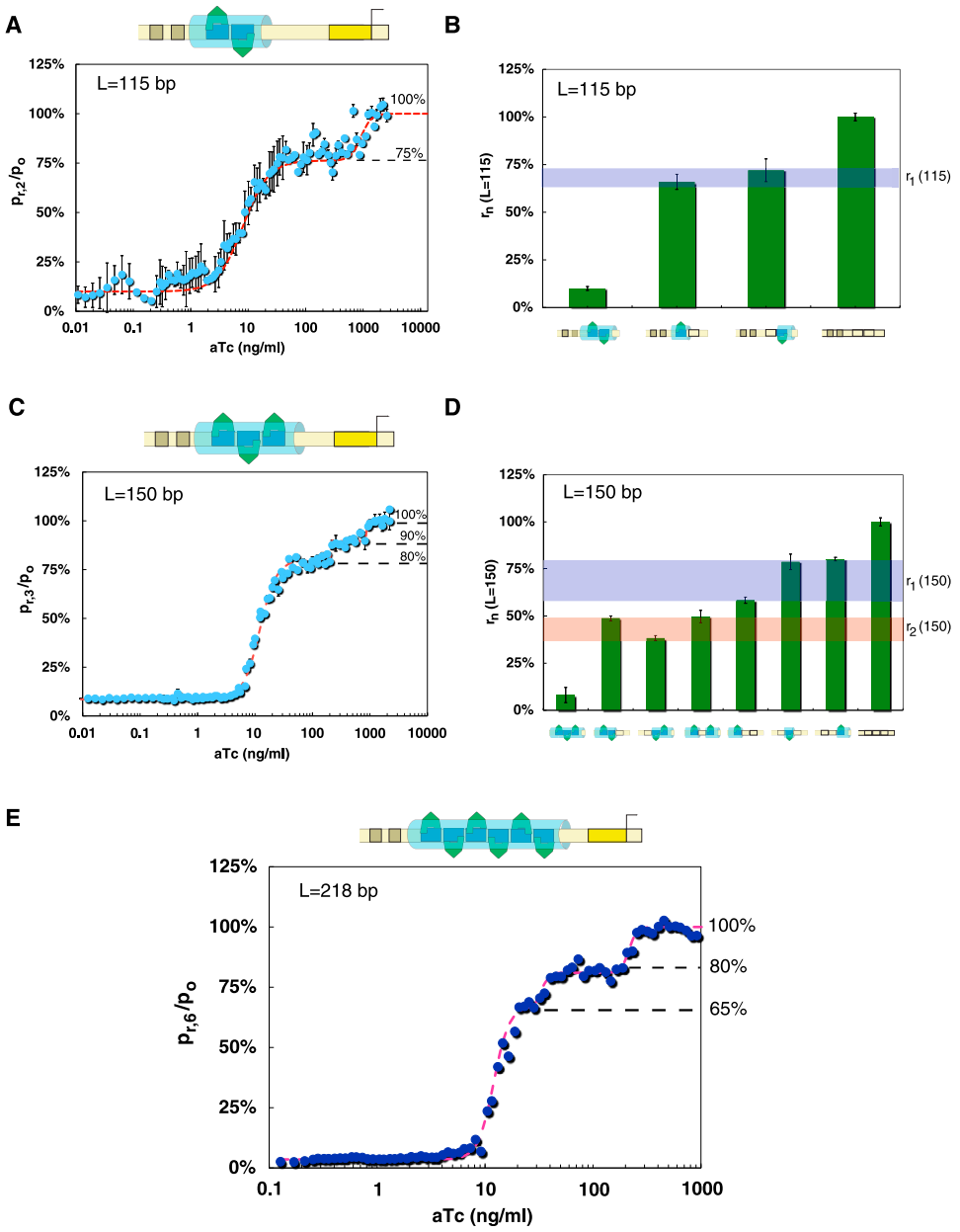

Figure 3. Synthetic Enhancers Convert Variable Ligand Input to Discrete Output Step Function

(A) High-resolution titration in 48-well plates of aTc with a 2-tet cassette at L = 115 bp. The data show three discrete states separated by transitions.

(B) Repression levels measured for synthetic enhancers characterized by a deletion of either one or both of the TetR-binding sites at L = 115 bp. The purple shading corresponds to the weak repression value r_1 (L = 115 bp).

(C) 3-Tet repression ratio at L = 150 bp exhibiting four discrete states, with the upper three closely clustered at average repression ratio values of ~80%, ~90%, and ~100%.

(D) Repression levels measured for synthetic enhancer cassettes (L = 150 bp) containing zero, one, or two TetR-binding sites arranged in configurations that mimic the three binding site enhancers' partial occupancy states due to aTc titrations. The purple and orange shading corresponds to the weak repression values r_1 (L = 150 bp) and r_2 (L = 150 bp).

(E) Data for the 6-Tet cassette showing only four states, characterized by increased separation and sharper transitions between the intermediate states.

The dashed red lines in (A), (C), and (E) correspond to empirical fits of two (A) or three (C and E) Hill functions stitched together in a piece-wise continuous fashion that highlight the transitions and levels observed in the data. Error bars correspond to the standard deviation from multiple measurements. See also Figure S5.

(Figures 4A–4D) between the low repressed state and the first intermediate are characterized by an increasingly steeper transition that can be empirically quantified by a Hill coefficient greater than one. Interestingly, the Hill coefficients that were extracted turn out to be roughly equal to the number of TetR-binding sites. This result seems to imply that the regulatory function reflects an

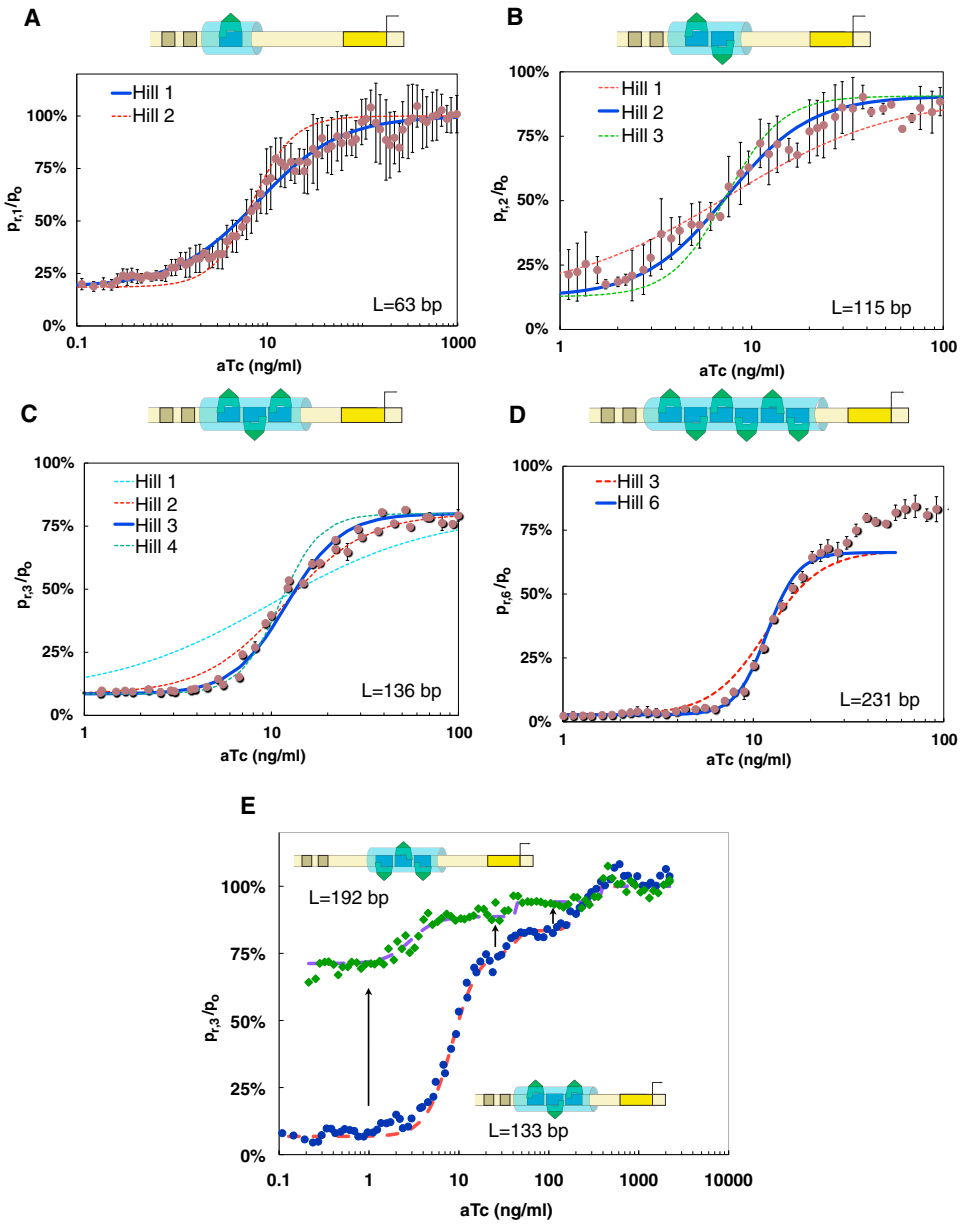

Figure 4. Coding and Computational Characteristics of Synthetic Enhancers

(A–D) Transition from the strongly repressed state to first intermediate level. The transition in all Tet cassettes (1, 2, 3, and 6) is best fitted by a Hill function of order (n), which roughly equals the number of binding sites. Dashed lines in each curve signify fits with Hill functions of n + 1 or n − 1, typically showing that only Hill functions of order n fit the data well.

(E) By shifting the cassette toward the σ^{54} promoter and away from the driver NRI#1 and #2 sites, a similarly shaped regulatory function (top) is observed. Error bars correspond to the standard deviation from multiple measurements. See also Figure S5.

effective interaction in the factors that bind to the cassette, which can be interpreted as a form of molecular counting.

To further examine the mechanistic underpinnings of our measurements, we examined the output function of additional synthetic enhancers with the binding site cassettes moved upstream a larger distance from the end of the NRI#2 site. This serves to further explore the effects of looping modification on the regulatory output and also as a control for whether or not our placement of the binding site cassette 28 bp upstream of the NRI#1,2 sites interferes in some nontrivial fashion with the binding of NRI∼P. Figure 4E shows that, for a synthetic enhancer with the three TetR-binding site cassette placed 45 bp downstream of the end of the NRI#2 site, the output function keeps its elementary characteristics (i.e., a strongly repressed state, a transition to one or two weakly repressed or unrepressed states, and transition steepness characterized by

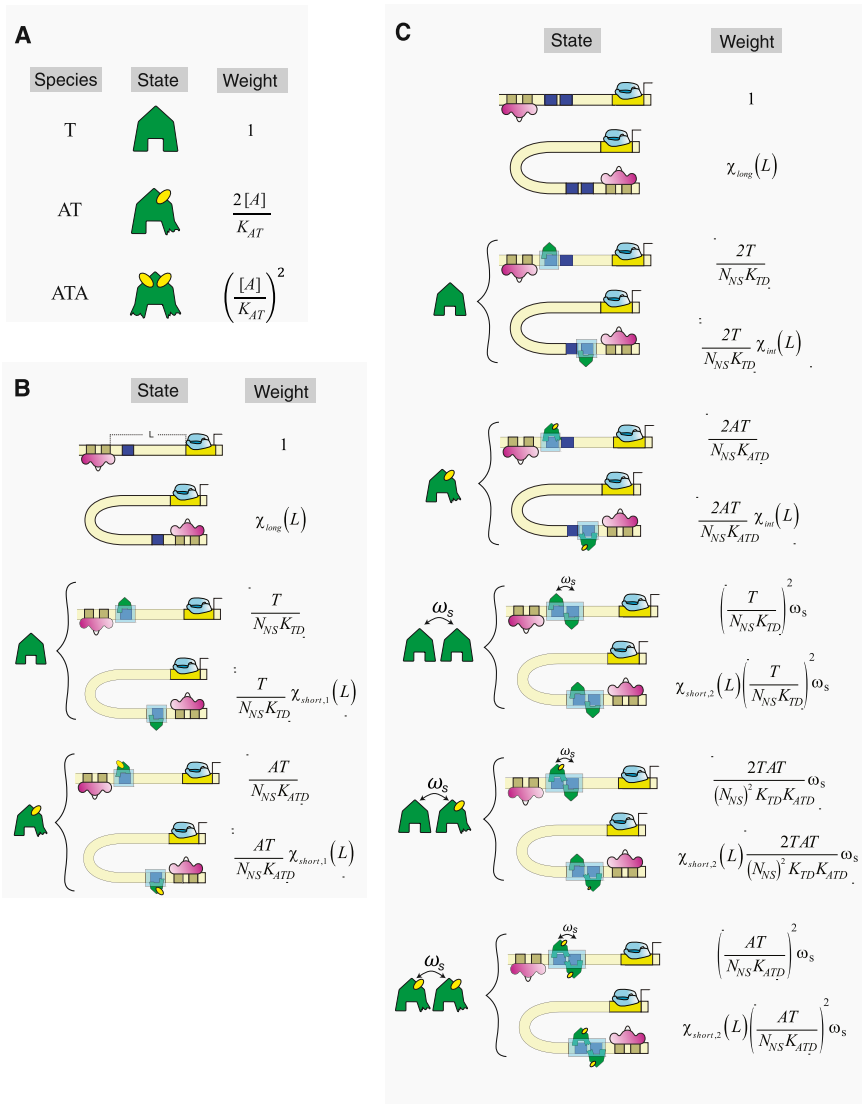

Figure 5. Generalized Model Schematic for Repression Ratio Data

(A–C) The models and their corresponding states and statistical weights are shown for (A) the interaction between aTc and TetR in solution, (B) the states and weights used for computing the repression ratio model function for the cases of a single TetR, and (C) two TetR. In (B) and (C), we now include states with the single aTc-bound TetR form. This protein has a binding affinity to the specific binding sites of TetR, which is two to three orders of magnitude lower than the free form of TetR. Furthermore, the two TetR model in (C) has a new parameter ω_s, which describes the interaction between adjacent TetR molecules. This interaction is crucial for the formation of steps in our model.

See also Figure S5 for model fits.

a Hill coefficient of three) regardless of where the cassette is positioned within the spacer region. Thus, the results shown in Figure 4E and the different response functions for the 1-, 2-, and 6-Tet cassettes (see Figure 4A, Figure 3A, and Figure 3E, respectively) suggest that each cassette type apparently encodes a particular output function, whose characteristic dose-response output depends on the geometry and binding site arrangement of the various TetR-binding cassettes and a possible interaction between TetR proteins bound on the cassette.

Modeling the Enhancer Output Functions

Given the modeling framework discussed in "Theory: Model for Looping Initiated Transcription" in the Supplemental Information, which were used to model the looping and the bimodal repression data, is it possible to generalize this scheme to reproduce the output functions shown in Figure 3 and Figure 4? In order to address this question, we need to develop a proper mechanism by which to extend the thermodynamic model to account for the aTc titrations. In doing so, we incorporate the following assumptions: the observation (Lederer et al., 1995) that up to two aTc ligand molecules can bind a single TetR dimer and that TetR can bind its DNA-binding site in two forms: unoccupied and occupied by a single aTc ligand but with different K_ds (see Table S4). These assumptions are based on crystal structure analysis (Orth et al., 2000) and in vitro binding experiments (Lederer et al., 1995, 1996). In the former, the ligand is shown to increase the distance between the DNA-binding motifs on the dimer, thus reducing the binding affinity to DNA of a protein bound by a single ligand and abolishing it altogether when both ligands are bound. In the latter, binding curve analysis suggests that more than one bound ligand is required to abolish TetR binding to the DNA.

These assumptions allow us to formulate states and weights prescriptions (see schematic in Figure 5A), which generate mathematical expressions (see "Theory: Model for Enhancer Repression via Induction" and Equations S32–S34 in the Supplemental Information) for the number of TetR molecules in various states of aTc occupancy—*T*, *AT*, and *ATA* corresponding to the number of free TetR proteins, TetR occupied by a single molecule of aTc, and doubly occupied TetR, respectively. Given this relationship between TetR and aTc, we were then able to install those results into our states and weights schemes for the various enhancer occupancies, which in turn allowed us to formulate a model for the repression ratio data (Figures 5B and 5C for generalized model schematics), which not only accounts for the looping size effect due to TetR binding, but also illustrates how this binding is altered in the presence of different concentrations of aTc.

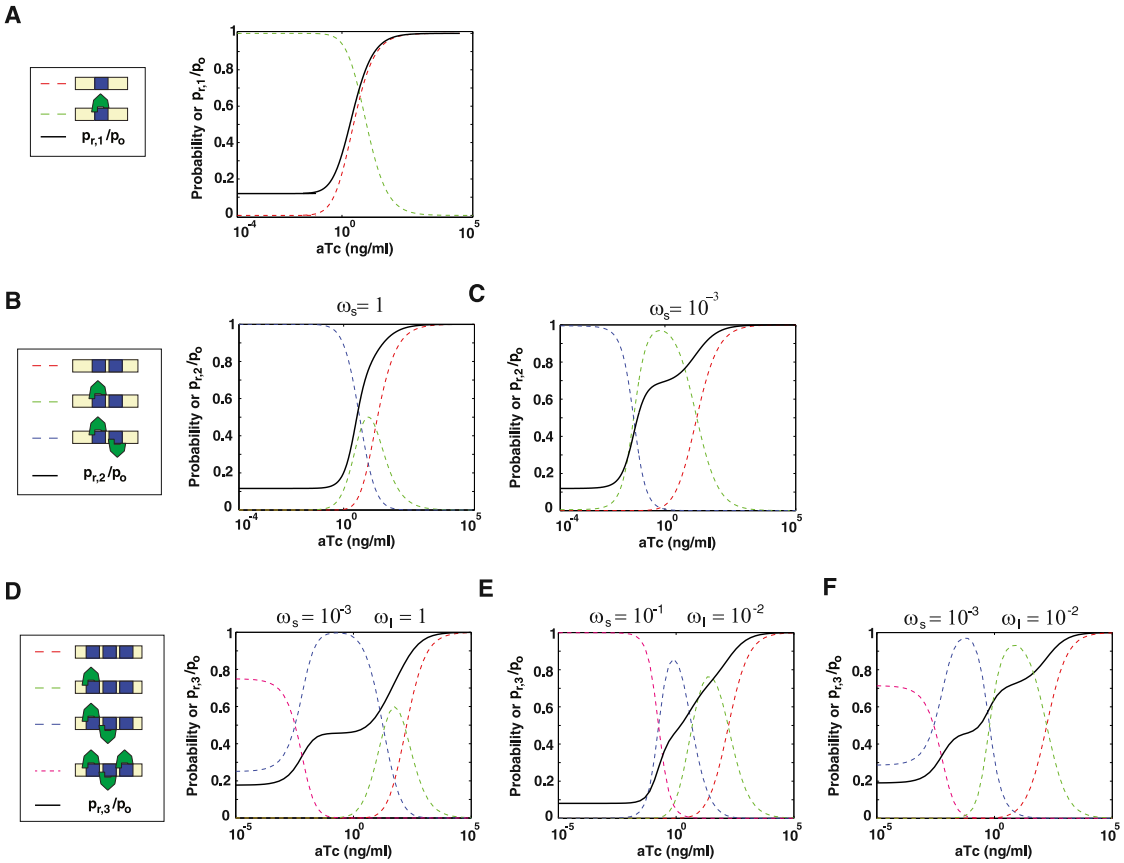

Figure 6. Theoretical Repression Ratio Curves and Associated Probabilities

In all panels the red, green, blue, and violet dashed lines correspond to the no occupancy, single, double, and triple occupancy state probability distributions respectively plotted as a function of aTc concentration. The thick black line corresponds to the theoretical repression ratio dose-response function computed at each aTc concentration from the individual probability distributions.

(A) Occupancy states and dose-response function for the single binding site case.

(B) Occupancy probability states of the two TetR-binding site model and associated dose-response function constructed using the parameters from (A) and $\omega_s = 1$.

(C) Same model as described in (B) but with $\omega_s = 10^{-3}$, implying that it is energetically unfavorable to have two TetR molecules bound next to each other.

(D–F) Occupancy probability states and associated dose-response functions for the three TetR-binding site model for cases in which the short- and long-range interaction parameters take the values (D) $\omega_s = 10^{-3}$ and $\omega_l = 1$, (E) $\omega_s = 0.1$ and $\omega_l = 10^{-2}$, and (F) $\omega_s = 10^{-3}$ and $\omega_l = 10^{-2}$.

Sample fits of the model to the data sets in Figure 3 and Figure 4 are shown in Figure S5.

First, we consider a model for the 1-Tet cassette. Figure 6A shows a typical repression ratio curve and associated occupancy state probabilities that can be obtained for a wide array of parameter combinations. The model for the 1-Tet case captures the essential features of the 1-Tet data (Figure S5A), as does the empirical fit given by a Hill function with Hill coefficient one, as shown in Figure 4A.

In order to extend the model to the 2-Tet case (see "Theory: Model for Enhancer Repression via Induction" in the Supplemental Information), we incorporate an additional parameter (ω_s) that accounts for any interaction that may be incurred between bound proteins on neighboring TetR sites. If this parameter is less than one, then the bound proteins exhibit anticooperative behavior, which leads to increased stability for the single occupancy configurations as compared with the double occupancy one. On the other hand, if ($\omega_s > 1$), then this parameters amounts to a cooperative interaction, which leads to a preference for the doubly occupied state as compared with other cassette occupancy states (data not shown).

In Figures 6B and 6C, we plot the individual probabilities (Equation S47) for the cassette suboccupancies as a function of ligand concentration for the 2-Tet case for two values of (ω_s): 1 and 0.001. The blue dashed lines in both panels correspond to the double occupancy probability, which approaches one for very low ligand concentrations and declines sharply thereafter. Likewise, the red lines correspond to the no occupancy configuration, and as expected, the probability of this state approaches one for very high ligand concentrations. The single occupancy probability (green lines) varies sharply between both panels. For values of ($\omega_s \cong 1$) (Figure 6B), it overlaps significantly with the other two probabilities, leading to a relatively small overall contribution from the single occupancy

configurations, which results in an output function that lacks an intermediate step (Figure 6B, black line). However, for values of (ω_s) that promote anticooperativity in the protein-protein interaction, the overlap of the probabilities is significantly reduced (Figure 6C), which in turn leads to an intermediate step in the output function. Thus, according to our model, the reduced stability of the double occupancy configuration is critical for the formation of the step function.

Extending the model further to the 3-Tet case (Figures 6D–6F and Figure S5C for fits) and varying the value for (ω_s) leads to the emergence of a step function for decreasing values of ω_s characterized by a single intermediate, as for the 2-Tet case. The plot in Figure 6D shows a clear signature for a step at a repression ratio level of $\cong$0.4–0.5, with a second additional sharp transition to the top level corresponding to the unoccupied cassette configuration. For slightly lower values of ω_s, the model produces an output function (Figure 6E) that looks similar to the data in Figure 3C. However, no matter what value of ω_s is chosen, the model is unable to produce two intermediate states. In order to generate a step function with two intermediates (Figure 6F), one has to introduce a second weaker anticooperativity term (ω_l) for the next to nearest neighbor interaction. As a result, we conclude that the existence of anticooperativity interaction parameters seems to be a crucial feature of any model that attempts to reproduce the particular discrete output functions obtained by the experiments, with the number of intermediates steps reflecting the extent of the protein-protein interactions (i.e., nearest neighbor, next-nearest neighbor, etc). However, a full microscopic understanding of the function of these synthetic enhancers requires a deeper knowledge of both the DNA mechanics and the ways in which the repressors interact both with each other and with their DNA substrate.

Conversion of the σ^{70} Activator TraR to a Repressor Using Synthetic Enhancers

We reasoned that there was nothing special about the character of TetR as a DNA-binding protein that led to the observed behavior of our synthetic enhancer. To the extent that this hypothesis is correct, we should be able to replace TetR with some other DNA-binding protein and obtain a qualitatively similar regulatory output. To that end, we constructed additional synthetic enhancer cassettes containing binding sites for the activator TraR. In particular, under normal circumstances, TraR, a LuxR homolog found in *Agrobacterium tumefaciens*, acts as a transcriptional activator of σ^{70} promoters. In *E. coli*, however, its transcriptional activation capability is abolished, though the specific DNA-binding activity remains (Qin et al., 2009 and references within). Thus, in our case, we can use this protein in the enhancer context to alter the looping region just as we did with TetR.

The results obtained previously for the TetR systems (Figure 2, Figure 3, Figure 4, Figure 5, and Figure 6) indicate that the behavior of the output functions that are generated by the class of models presented here depends strongly on three parameters: the values of the looping capacities for the different enhancer states of occupancy by the enhancer binding protein (Figure 7A), the number of binding sites (Figure 7B), and the protein-protein interaction parameter (Figure 7C). In particular, the protein-protein interaction parameter determines whether the regulatory output will exhibit a smoothly decreasing expression level function ($\omega_s \cong 1$) or be characterized by sharp transitions and an intermediate expression level step ($\omega_s << 1$). Because the presence of a step in the regulatory output function indicates that the states with several enhancer-binding proteins bound are relatively unstable, the model predicts that this effect is attainable experimentally if a large mutual exclusion effect is engineered into the synthetic enhancer design.

Due to the fact that the DNA binding probability for TraR increases as a function of ligand concentration (see Figure 7D and "Theory: Model for Enhancer Repression via Induction" in the Supplemental Information), the model predicts that it is possible to obtain a regulatory output function that is qualitatively a mirror image of the output function obtained for the synthetic enhancer architecture with three TetR-binding sites (for states and weights, see Figure S6). Consequently, we opted to design the TraR synthetic enhancer with 6 bp spacing between the binding sites to ensure that a mutual exclusion effect will be present as a result of presumed excluded volume effects between the bound TraR dimers. Figures 7E and 7F show the experimental results and model predictions. At low ligand concentrations of the small inducer molecule that is necessary for TraR to bind to DNA, N-(3-oxo-octanoyl)-L-homoserine (3OC8), the enhancer regulatory response is characterized by a small magnification ($\cong$7%) of expression levels as compared with the unoccupied enhancer for 3OC8 concentrations that are less than 10 nM. For larger concentrations, repression characterized by clearly detectible steps is observed with a minimal value of $\cong$60% of the unoccupied enhancer expression level. The data indicate that a well-separated intermediate in repression values occurs at $\cong$90% of unoccupied expression level and ranges from $\cong$30 to 500 nM in 3OC8 concentration, validating the model's qualitative predictions and our general approach for inducing regulatory response in synthetic enhancer design.

DISCUSSION

We explored transcriptional and regulatory characteristics of an enhancer-based transcriptional system by constructing increasingly complex enhancer elements from the ground up. Our approach was predicated on the assumption that a bacterial enhancer can be constructed as a modular object made of three connected components: driver-binding sites, a poised σ^{54} promoter, and small DNA cassettes containing several binding sites for DNA-binding proteins. In this work, we restricted ourselves to using the same module for the driver and poised promoter while varying the enhancer-binding protein binding site module. However, we suspect that any of the other modules can be altered to access an even richer space of regulatory effects.

We then proceeded to characterize our synthetic enhancers' regulatory output functions using experimental measurements and a set of thermodynamic models. Our results show that, unlike the conventional model for repression, wherein a repressor inhibits transcription by competing

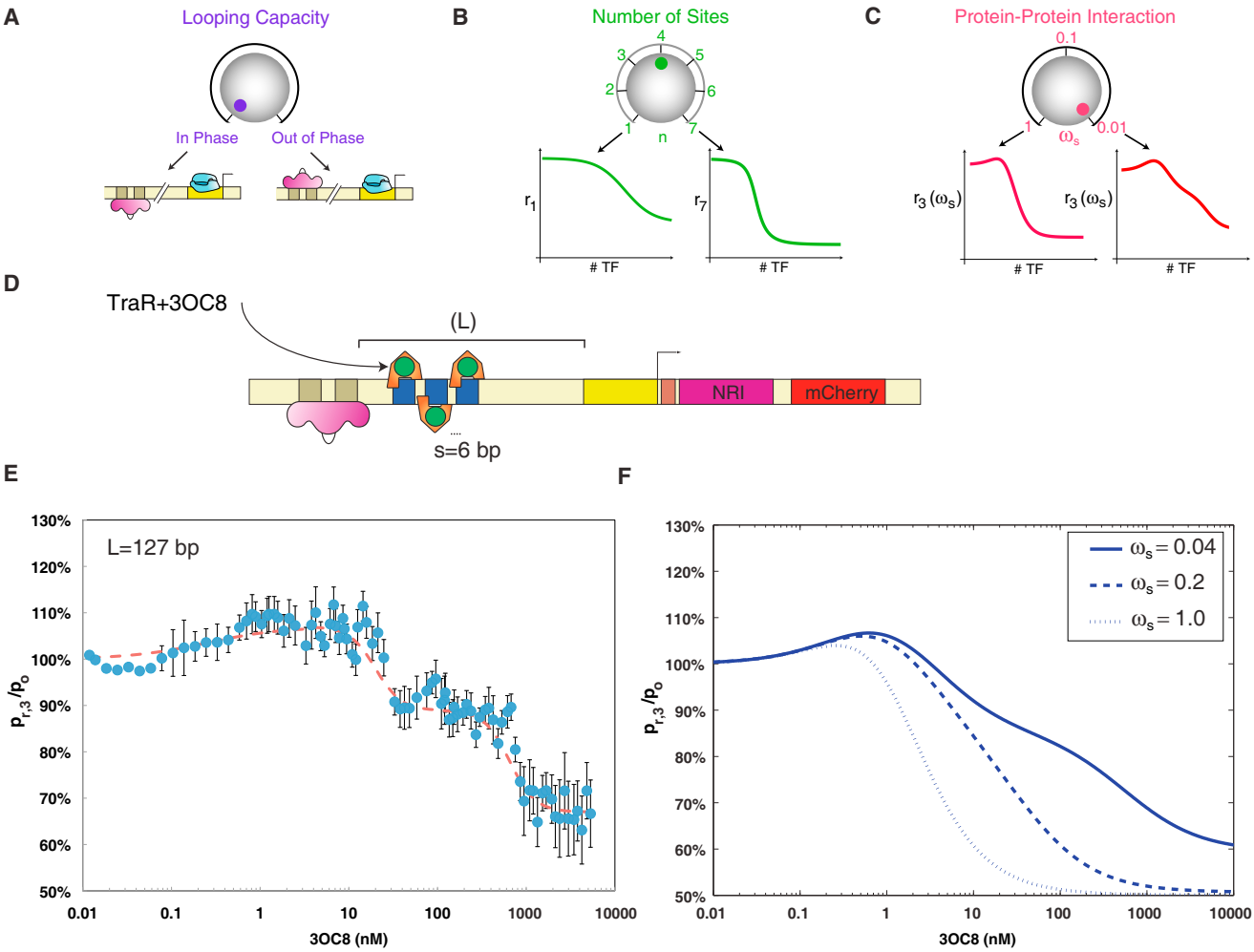

Figure 7. Prediction and Regulatory Behavior of an Enhancer Designed from the Ground Up

(A–C) Schematic representation of the different experimental knobs for controlling enhancer regulatory output.

(D) Circuit schematic for the TraR synthetic enhancers with three TraR-binding sites arranged with 6 bp between each site (as compared with 16 bp used for TetR). The cartoon for the TraR protein signifies that only the dimeric isoform of TraR bound to the cognate ligand 3OC8 can bind DNA.

(E) Regulatory dose-response function for the TraR synthetic enhancer over six decades of 3OC8 concentration. The dashed red line corresponds to an empirical fit of two Hill functions stitched together in a piece-wise continuous fashion.

(F) Model prediction (Equation S50) for regulatory output of the TraR synthetic enhancer showing examples with three values of the short-range interaction parameter. We used the following normalized looping capacity values (i.e., each value is divided by χ_o) for all three curves: $[\chi_L, \chi_{int1}, \chi_{int2}, \chi_{short,3}] = [1, 1.15, 0.85, 0.5]$. See Figure S6 for a detailed graphical representation of the statistical states and weights for this model. See also Figure S7 for schematic of naturally occurring bacterial enhancers.

for the RNAP-binding site or by interfering with RNAP initiation, the synthetic enhancers exhibit repression by a modification of the DNA's capacity to loop. This leads to a regulatory output that is characterized by two key modes: a strongly repressed state in which the enhancer is unlikely to loop and a weakly repressed state in which looping is more likely at short and long looping lengths, respectively. Within each mode, the resultant level of repression depends on the enhancer element properties (i.e., number of binding sites, transcription factor binding regions, binding site arrangement and spacing, etc.) and weakly on the length of the loop (Figure 2B). Therefore, these results provide a mechanistic model for regulatory action at a distance by showing that regulatory effects can be systematically generated when the transcription factors are bound at large distances (i.e., hundreds of bps) from the basal promoter.

One striking outcome induced by the various repression states observed for our synthetic enhancers is the emergence of step-like dose-response regulatory output functions. In the Results section, we showed that the steps that form in the response for the 2-Tet, 3-Tet, and 6-Tet cases can be explained by repression levels of preferred cassette occupancy states. The preferred states, in turn, are determined by various anticooperativity parameters, which are used to model a destabilizing interaction between two TetR proteins that are bound in the vicinity of one another.

Given this analysis, we then asked whether it is possible to utilize these underlying mechanisms that are responsible for enhancer regulatory output and design a new synthetic enhancer from the ground up with a predetermined output function using a completely different enhancer-binding protein. We showed that, if we replace the TetR protein by another DNA-binding protein (TraR) and conserve the binding geometry (i.e., proteins are bound in opposite orientation with spacing of 6 bp for TraR and 16 bp for TetR), the same step-like regulatory response is observed in accordance with the model's qualitative predictions. As a result, our data suggest that the specific identity of the enhancer-binding protein (TraR and TetR are generic choices of DNA-binding proteins) is not as crucial to the regulatory output as the arrangement and number of its binding sites. Consequently, the design of enhancer regulatory output is reduced to a consideration of the variable looping geometry induced by the presence of DNA-binding proteins within the loop.

The observed discrete levels of the regulatory output (Figure 2) and the transitions between steps (Figure 3,4,7) of this output illustrate that a form of molecular counting is taking place at the synthetic enhancer. Because regulation has traditionally been used to explain the phenomenon of gene switching from "on" to "off" and vice versa, how do we then classify cases like that described here, wherein there are apparently more than two discrete regulatory states that can be accessed within a singular regulatory motif?

The regulatory effects observed with our synthetic enhancers can be interpreted via our model as a cumulative outcome of three analog knobs individually tuned to particular values (Figures 7A–7C). These knobs are the looping capacity values, the number and arrangement of transcription factor binding sites, and the character of the protein-protein interaction. All three of these tuning variables are distinct yet affected by the particular state of the others. For instance, we showed that the ability to loop is affected by the presence or absence of DNA-binding proteins and by the number of binding sites. Furthermore, the number of bound proteins for a given concentration of inducer is, in turn, affected by the protein-protein interaction parameter, which reflects the number of active proteins present in the cell.

Even though our experiment and model allowed us to conveniently identify or isolate these control parameters, at present, the models serve primarily as a conceptual framework for understanding the behavior of the synthetic enhancers as a function of the various regulatory knobs that can be tuned. Unfortunately, for the time being, it is not possible to predict either the looping capacity or the protein-protein interaction parameters from first principles. In particular, for the cases presented here, we showed that the looping capacity can be repressive for the case of TetR or repressive and activating for TraR. Both of these observables are apparently related to the particular localized protein-DNA interactions, yet we are unable to formulate a first principles theoretical model for these quantities. These uncertainties are an inheritance of our current limited understanding of in vivo DNA mechanics, protein-DNA interactions, and protein-protein interactions for neighboring transcription factors. At the same time, we view the kind of interplay between experiment and theory played out here as precisely the type of approach that will allow us to begin to develop quantitative intuition for all of these phenomena.

Given these limitations, what are the practical lessons learned from our synthetic enhancer's capability to count molecules or "measure" cellular concentration of proteins? Recently, molecular counting was demonstrated using gene regulatory networks via both systems (Long et al., 2009) and synthetic biology (Friedland et al., 2009) approaches. When comparing these two examples, we find that they describe two different forms of counting. In Friedland et al. (2009), the authors demonstrate a chemical pulse counter, which yields a singular output once a particular pulse number is reached. On the other hand, the quorum-sensing counter shown by Long et al. (2009) generates an output expression level, which is a discrete function of the number of inputs integrated (in their case, two). The behavior of the circuits that we have constructed are analogous to integrative counters but exhibit a capability to integrate more than two inputs in a compact DNA sequence architecture. As a result, it is tempting to speculate that gene regulatory circuits, which utilize enhancers as input integrators, can therefore enable an enriched regulatory potential.

Finally, the motivation for building synthetic enhancers from the ground up is to not only generate some complex regulatory phenomenon, which in this case tests our understanding of protein-DNA interactions and poised transcription, but to also try to isolate underlying mechanisms that are responsible for natural regulatory phenomenon. Similar constructionist approaches have been used often in recent years to study gene regulatory networks, and in the many examples published to date (e.g., Basu et al., 2005; Elowitz and Leibler, 2000; Gardner et al., 2000), gene circuits synthesized de novo often yielded important insights into the underlying mechanism of protein networks in biology. Hence, the question remains of whether any of the above results and their interpretations provide new insight into regulatory phenomena observed in natural bacterial enhancers.

As an example of natural bacterial enhancers, the wild-type NRI∼P system in *E. coli* contains three additional NRI sites (#3–#5) (see Figure S7 and "Theory: Model for Enhancer Repression via Induction" in the Supplemental Information) that flank the #1 and #2 sites and σ^{54} promoter, in what we defined as the looping region (see Figure 1). Deletion of these sites (Atkinson et al., 2002) has been shown to increase expression in discrete amounts driven by the hexamer bound at the #1 and #2 sites. These additional sites have been dubbed "governor sites" as a tribute to the fact that they limit or inhibit the overall expression level. Thus, we can effectively consider this natural system as analogous to the synthetic enhancer considered here with a "cassette" of three additional NRI-binding sites.

To explore this analogy further, we examined the binding site architecture of three additional bacterial enhancers (http://regulondb.ccg.unam.mx/) (Figure S7). In a manner similar to the synthetic enhancers, these natural enhancers form entities that are capable of integrating multiple inputs upstream of a poised σ^{54} promoter. The binding site architectures imply that the regulatory output exhibited by these enhancers may be characterized by a similar modeling approach to the one used here. Because the ingredients used to construct our synthetic

enhancer are all common architectural elements in real transcriptional networks, we argue that the capacity to assemble these elements as done here can provide a predictive model for deciphering the regulatory output of additional bacterial enhancers in the natural context as well. Given these similarities, it is tempting to speculate that the modification of the looping capacity mechanism explored in our work might actually be a strategy adopted for the regulation of natural enhancers in bacteria.

EXPERIMENTAL PROCEDURES

Synthetic Enhancer Cassette Design

Synthetic enhancer cassettes (Table S1) were designed as follows. First we computationally designed 100 bp sequences that had a minimal probability to bind DNA-binding proteins. This was done by constructing an algorithm that randomly generated a set of 1 million 34 bp sequences. The sequences were compared to the roughly 1900 known specific DNA-binding sites for *E. coli* transcription factors obtained from RegulonDB (http://regulondb.ccg.unam.mx/). Each calculated sequence was scored by first computing the percent homology with a particular binding site, weighting that number by an exponential weight that heavily favors low homologies, and finally totaling the values obtained for each of the 1900 binding sites (sequences that matched a known binding site were eliminated). After obtaining the sequences with the lowest scores, a second run was carried out on the complementary sequence of the lowest-scoring 1% of the original sequences. The scores of the two runs were combined, and sequences with the lowest combined scores were listed in order. The sequences were predominantly GC rich (~75%) with very low A and T content (~25%). We ordered the spacer92 (Table S1) sequence using two complementary primers (IDT).

Cassettes containing TetR-binding sites were designed as follows (all containing a tandem of NheI sites). The 1-Tet cassette included the high-affinity (10 pM) TetO2 site (Hillen and Berens, 1994) (Table S1). The 2-Tet cassette included the TetO1 site (30–50 pM) site (Hillen and Berens, 1994), a 16 bp spacer (obtained from the calculated spacer sequence; see above), and a TetO2 site. The 3-Tet cassette contained two TetO2 sites, two spacer sequences of 16 bps (determined using the above algorithm), and one TetO1 site. The 3-Tet-S cassette has additional spacer sequences placed in front of the first TetO1 site and after the last site. The 6-Tet cassette is effectively a double cassette made of a tandem of 3-Tet cassettes.

All cassettes were ordered as complimentary oligos from IDT. Oligos were hybridized as follows (in saline solution containing 10 mM $MgCl_2$) and then placed on ice: 2′ @95°C, 15′ @65°C, 5′ @42°C. Hybridized dsDNA cassettes were gel purified and digested with NheI before being used as an insert in the cloning step.

Looping Length Dependence Assay

20 ml of fresh LB with appropriate antibiotics was inoculated in 125 ml flasks with overnight starters of synthetic enhancer strains characterized by different looping lengths (i.e., 3.300LG cells + synthetic enhancer plasmid + p3Y15 plasmid; Atkinson et al., 2003). Cultures were vigorously shaken at 37°C (Innova), and fluorescence measurements were taken at 30 min intervals for roughly 5 hr to cover the midlog growth range. For each measurement, 200 μl of culture was dispensed in each of four wells of a 96-well plate (Corning Costar–Fisher Scientific). The 96-well plates were read by a plate reader (Tecan–Infinite 200) at 580/610 excitation/emission with gain 100 and appropriate controls for autofluorescence and glnAp1 leakage. The fluorescence results for the four wells were averaged and normalized by a reading of the culture's OD600. S/N was > 10 for all synthetic enhancer strains tested with respect to leakage and > 20 with respect to auto fluorescence (obtained from a null strain).

Repression Measurement Assay

Repression level measurements were carried out as follows: first, synthetic enhancer plasmids were transformed with either pACT-Tet (Figure S1A) or pACT-Tra plasmids in 3.300LG (Atkinson et al., 2003) cells (in which the *traR* gene replaces the *tetR* gene). Next, synthetic enhancer strains were grown in fresh LB with appropriate antibiotics (Kan/Amp) to midlog range, as measured by a spectrophotometer (Pharmacia Biotech) OD600 of ≅0.6 and were resuspended in low-growth/low-autofluorescence BA buffer (for 1 l – 0.5 g Tryptone [Bacto], 0.3 ml Glycerol, 5.8 g NaCl, 50 ml 1M $MgSo_4$, 1 ml – 10 × PBS buffer – pH 7.4, 950 ml DDW). 1 mM IPTG was added to induce the circuit at this point to deactivate the LacI protein that represses the glnAp2 promoter. 2 ml of resuspended culture with IPTG were dispensed in each well of a 48-well plate. The plates were then incubated in a 37°C shaker until cultures reached growth steady state. Measurements of fluorescence levels were taken by dispensing 200 μl of culture in each well into a 96-well plate and were carried out on a plate reader as mentioned above. All repression measurements were done in triplicates with cultures grown from individual synthetic enhancer strain colonies.

To get the percentage of inhibition, autofluorescence levels were subtracted from expression levels measured for strains with and without endogenous TetR. Subsequently, the ratio of the adjusted fluorescence level for the +TetR strains to the –TetR strains was taken.

Repression Ratio Measurement Assay

Synthetic enhancer strains containing the pACT-Tet or pACT-Tra plasmid were initially grown in LB, resuspended in the low growth buffer, and dispensed in the 48-well plates. In this case, appropriate concentrations of aTc or 3OC8 (sigma) were dispensed in each well, spanning four to six orders of magnitude. For each strain, we used two plates to allow for 94 different readings of fluorescence as a function of aTc concentration (two wells were used as –IPTG controls). We carried out each measurement in duplicates, i.e., four plates per measurement.

To compute the repression ratio levels as a function of aTc or 3OC8 concentrations, each fluorescence ratio value was calculated using a running average algorithm. This entails averaging three to five raw fluorescence readings for every fluorescence value shown, whereby the averaging is carried over adjacent inducer concentrations. This algorithm is used to smooth out short-range fluctuations and highlights the large-scale features that span wide concentration ranges.

Strain Construction

See Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and five tables and can be found with this article online at doi:10.1016/j.cell.2011.06.024.

ACKNOWLEDGMENTS

We would especially like to thank Prof. Frances H. Arnold for providing lab space and the forum to conduct thorough discussions as this project was evolving. We would also like to thank Eric H. Davidson for important early discussions and Alex J. Ninfa for plasmids and strains. R.A. was supported by a NIH Ruth L. Kirschstein fellowship, a Caltech CBCD grant, and the NIH through award NIH ARRA R01 GM085286-01S. R.P. and H.G.G. gratefully acknowledge awards NIH ARRA R01 GM085286-01S, R01 GM085286, and the NIH Director's PIONEER Award DP1 OD000217.

Received: August 24, 2010
Revised: January 25, 2011
Accepted: June 14, 2011
Published: July 7, 2011

REFERENCES

Atkinson, M.R., Pattaramanon, N., and Ninfa, A.J. (2002). Governor of the glnAp2 promoter of Escherichia coli. Mol. Microbiol. *46*, 1247–1257.

Atkinson, M.R., Savageau, M.A., Myers, J.T., and Ninfa, A.J. (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. Cell *113*, 597–607.

Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H., and Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. Nature *434*, 1130–1134.

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

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.

Buck, M., Gallegos, M.T., Studholme, D.J., Guo, Y., and Gralla, J.D. (2000). The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. J. Bacteriol. *182*, 4129–4136.

Davidson, E.H. (2001). Genomic Regulatory Systems: Development and Evolution (Burlington, MA: Academic Press).

Davidson, E.H. (2006). The Regulatory Genome (Burlington, MA: Elsevier).

Driever, W., Thoma, G., and Nüsslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. Nature *340*, 363–367.

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

Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G., and Collins, J.J. (2009). Synthetic gene networks that count. Science *324*, 1199–1202.

Garcia, H.G., and Phillips, R. (2011). Quantitative dissection of the simple repression input-output function. Proc. Natl. Acad. Sci. USA. 10.1073/pnas.1015616108.

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

Hervás, A.B., Canosa, I., Little, R., Dixon, R., and Santero, E. (2009). NtrC-dependent regulatory network for nitrogen assimilation in Pseudomonas putida. J. Bacteriol. *191*, 6123–6135.

Hillen, W., and Berens, C. (1994). Mechanisms underlying expression of Tn10 encoded tetracycline resistance. Annu. Rev. Microbiol. *48*, 345–369.

Hillen, W., Schollmeier, K., and Gatz, C. (1984). Control of expression of the Tn10-encoded tetracycline resistance operon. II. Interaction of RNA polymerase and TET repressor with the tet operon regulatory region. J. Mol. Biol. *172*, 185–201.

Huo, Y.X., Tian, Z.X., Rappas, M., Wen, J., Chen, Y.C., You, C.H., Zhang, X., Buck, M., Wang, Y.P., and Kolb, A. (2006). Protein-induced DNA bending clarifies the architectural organization of the sigma54-dependent glnAp2 promoter. Mol. Microbiol. *59*, 168–180.

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.

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.

Law, S.M., Bellomy, G.R., Schlax, P.J., and Record, M.T., Jr. (1993). In vivo thermodynamic analysis of repression with and without looping in lac constructs. Estimates of free and local lac repressor concentrations and of physical properties of a region of supercoiled plasmid DNA in vivo. J. Mol. Biol. *230*, 161–173.

Lederer, T., Takahashi, M., and Hillen, W. (1995). Thermodynamic analysis of tetracycline-mediated induction of Tet repressor by a quantitative methylation protection assay. Anal. Biochem. *232*, 190–196.

Lederer, T., Kintrup, M., Takahashi, M., Sum, P.E., Ellestad, G.A., and Hillen, W. (1996). Tetracycline analogs affecting binding to Tn10-Encoded Tet repressor trigger the same mechanism of induction. Biochemistry *35*, 7439–7446.

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.

Long, T., Tu, K.C., Wang, Y., Mehta, P., Ong, N.P., Bassler, B.L., and Wingreen, N.S. (2009). Quantifying the integration of quorum-sensing signals with single-cell resolution. PLoS Biol. *7*, e68.

Magasanik, B. (1993). The regulation of nitrogen utilization in enteric bacteria. J. Cell. Biochem. *51*, 34–40.

Mukherji, S., and van Oudenaarden, A. (2009). Synthetic biology: Understanding biological design from synthetic circuits. Nature Rev. Genetics *10*, 859–871.

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.

Ninfa, A.J., and Atkinson, M.R. (2000). PII signal transduction proteins. Trends Microbiol. *8*, 172–179.

Ninfa, A.J., Reitzer, L.J., and Magasanik, B. (1987). Initiation of transcription at the bacterial glnAp2 promoter by purified E. coli components is facilitated by enhancers. Cell *50*, 1039–1046.

Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. Nat. Struct. Biol. *7*, 215–219.

Phillips, R., Kondev, J., and Theriot, J. (2009). Physical Biology of the Cell (New York: Garland Science).

Qin, Y., Keenan, C., and Farrand, S.K. (2009). N- and C-terminal regions of the quorum-sensing activator TraR cooperate in interactions with the alpha and sigma-70 components of RNA polymerase. Mol. Microbiol. *74*, 330–346.

Ramos, J.L., Martínez-Bueno, M., Molina-Henares, A.J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., and Tobes, R. (2005). The TetR family of transcriptional repressors. Microbiol. Mol. Biol. Rev. *69*, 326–356.

Rappas, M., Bose, D., and Zhang, X. (2007). Bacterial enhancer-binding proteins: unlocking sigma54-dependent gene transcription. Curr. Opin. Struct. Biol. *17*, 110–116.

Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., and Elowitz, M.B. (2005). Gene regulation at the single-cell level. Science *307*, 1962–1965.

Schulz, A., Langowski, J., and Rippe, K. (2000). The effect of the DNA conformation on the rate of NtrC activated transcription of Escherichia coli RNA polymerase.sigma(54) holoenzyme. J. Mol. Biol. *300*, 709–725.

Su, W., Porter, S., Kustu, S., and Echols, H. (1990). DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial glnA promoter. Proc. Natl. Acad. Sci. USA *87*, 5504–5508.

EXTENDED EXPERIMENTAL PROCEDURES

Strain Construction

Synthetic enhancer strains were constructed off of a basic template plasmid. To construct the template the following steps were taken: first the LacI inducible GlnG + σ^{54} promoter cassette was cloned via PCR from the pglnAp2 plasmid obtained from A. Ninfa (Atkinson et al., 2002; Atkinson et al., 2003) into a basic pPROLar plasmid used in the Frances Arnold lab for library construction. Second, a mCherry reporter gene was cloned with an associated strong ribosome binding site (AGGAGA) downstream of the GlnG gene. Third, the NRI#3,#4 sites were mutated to inactivate these sites (following (Atkinson et al., 2002)), and instead a single NheI was inserted 22bp upstream of the NRI#1,2 binding sites. Fourth, the spacer92 sequence (Table S1) was inserted into the NheI site to make SCRM10. Finally, an additional NheI site was inserted in several locations along the sequence flanking the σ^{54} promoter and first NheI site to allow for maximal variability in the base template plasmids.

The synthetic enhancer library (Figure S1A) was constructed by digesting the template plasmids at the NheI sites, and either re-annealing without a new segment of DNA or with the various dsDNA cassettes as described above. Altogether, 70 synthetic enhancer plasmids were constructed and sequence verified (Table S1 and S2).

The pACT family of plasmids (Figure S1A) was constructed by modifying p3Y15 (a gift of A. Ninfa (Atkinson et al., 2003)). Into the parent plasmid, we inserted a *lacI* gene, and either a *tetR* or *traR* gene all under the control of the same glnL promoter as the one controlling NRII2302 mutant.

The synthetic enhancer strains were constructed by transforming into a 3.300LG strain (a gift of A. Ninfa (Atkinson et al., 2003) with deletions for *glnL* and *glnG* genes) sequence verified pACT and synthetic enhancer plasmids (Figure S1A). Selection was carried out via double Kan/Amp resistance (20 μg/ml and 100 μg/ml). Candidate synthetic enhancer strains were tested for fluorescence in the presence and absence of IPTG and TetR as required on the plate reader (Tecan – Infinite 200) to ensure that a proper strain was constructed. All sequences are available upon request.

Single-Cell Microscopy

Single-cell fluorescence measurements were carried out as follows. A particular synthetic enhancer strain was grown to mid log range in the presence of IPTG and mixed in 1:1 ratio with the maximally expressing strain (L = 70 bp) adjusted to the same OD. This was done to avoid any systematic error that may be incurred by microscope gain and offset differences for separate samples. 2 μl of the mixed culture was dispensed on slides with 1.5% agar slabs in PBS. Cells were incubated for 30 min at room temperature, and then imaged with an Olympus IX81 microscope, with 60x 1.2NA water Immersion objective, and Hamamatsu ORCA ER digital cooled CCD camera. Several images of dispersed monolayer "cell-sheets" were taken to allow for a large collection of statistics, which was subsequently analyzed (Figure S1C).

Image Analysis

Single-cell data were analyzed by counting and binning pixel gray-scale values (using ImagePro and Matlab). Figure S1C, top, shows a single cell population image used as control to show that cell populations are well fit by a Gaussian distribution. Due to this control, we were able to define the range of pixel brightness that corresponded to single-cell fluorescence levels, which were then fit with a two-peak function. This yielded an average fluorescence value for each strain in the 1:1 mix (Figure S1C – bottom histogram.) Note, the background brightness is omitted from Figure S1C-bottom, as it is clearly separable from the fluorescence signature in the data of the single strain sample Figure S1C, top.

To get the value for $p_o(L)/p_o(70)$, a ratio of the means for the two populations in each frame was obtained, and error bars were calculated based on the variance of this value obtained from the different frames. Comparison of single cell and bulk data for $p_o(L)/p_o(70)$ (Figure S1D) shows that both measurements yield data that is identical within experimental error.

Theory: Model for Looping-Initiated Transcription

In the first part of the theory section, we construct a general model for enhancer-based or looping-initiated transcription. Next, we adjust the model to the specific molecular characteristics of our synthetic enhancer model system, showing that the general model is applicable for the experimental conditions used in our study. Finally, we show that the general looping-activation framework can be extended using the same equilibrium thermodynamic principles to account for the binding of additional transcription factors to the enhancer. This, in turn, is shown to generate long-range regulatory effects. The theory developed in this section is used to model and fit the data depicted in Figure 1 and Figure 2.

General Model for Looping-Initiated Transcription

To develop intuition for how the looping capacity of the regulatory network could be tuned, we resorted to simple models that have been put forth to describe DNA mechanics in the regulatory setting. Our approach does not assume a particular polymer model of DNA. Though there are subtleties in the description of the circuit that are likely beyond the reach of these simple models, they nevertheless provide a basis for thinking about how our synthetic enhancers would behave as the concentration of TetR that binds in the DNA looped region of our circuit was titrated. These ideas are sketched below.

The J-factor characterizes the propensity of a linear piece of DNA to form circles. In a cyclization experiment, the formation of DNA circles starting from linear DNA, the J-factor can be defined as the local concentration of one DNA end in the vicinity of the other end

(Flory et al., 1976; Jacobson and Stockmayer, 1950; Marky and Olson, 1982). Though defined with reference to the properties of naked DNA, an effective looping J-factor can be considered for in vivo situations like those addressed here as well. Previously, the J-factor has been used successfully in an analogous way to describe DNA looping in in vitro and in vivo settings (Bintu et al., 2005a; Han et al., 2009; Zhang et al., 2006a).

In order to model looping-initiated transcription, we start with a model (Schulz et al., 2000) originally introduced to analyze in vitro transcriptional assays carried out with NRI∼P, in which it is assumed that the process of DNA looping is an equilibrium process, which then allows us to exploit the J-factor and additionally defines a dissociation constant K_{nr} as a measure of the strength of the protein-protein interaction (NRI∼P-σ^{54}) in the looped conformation. In the particular case of our system we define the looping J-factor as the local concentration of the minimal enhancer (i.e., DNA-bound NRI∼P) in the vicinity of the ''poised'' RNAP-σ^{54} complex.

Using the definitions provided above we capture the kinetics of looping-initiated transcription in the following rate equation

$$\frac{d[mRNA]}{dt} = \alpha P_{loop}(L) - \beta[mRNA], \tag{S1}$$

wherein α is the maximum rate of mRNA production and β is an mRNA degradation rate constant, and $P_{loop}(L)$, the probability of transcription start, is given by:

$$P_{loop} \cong \frac{\frac{J(L)}{K_{nr}}}{1+\frac{J(L)}{K_{nr}}}, \tag{S2}$$

wherein *J(L)* is the J-factor for a length L. To develop intuition for the significance of the $J(L)/K_{nr}$ ratio and simplify the discussion below, we define the ''looping capacity'' as,

$$\chi(L) \equiv \frac{J(L)}{K_{nr}}. \tag{S3}$$

As a result of these definitions, we can now write the steady state transcription and fluorescence reporter levels as:

$$[P] \propto [mRNA] \cong \frac{\alpha}{\beta}\left(\frac{\chi(L)}{1+\chi(L)}\right) \tag{S4}$$

wherein *[P]* corresponds to reporter concentration (or fluorescence level readout), and *[mRNA]* corresponds to the mRNA concentration.

There has been a great deal of effort invested in working out the characteristics of looping J-factors from first principles (Lilja et al., 2004; Merlitz et al., 1998; Rippe et al., 1995; Semsey et al., 2004; Swigon et al., 2006; Towles et al., 2009; Zhang et al., 2006a; Zhang et al., 2006b). However, for most purposes it suffices to represent the J-factor via a simple functional form motivated by the worm-like chain model of DNA bending (Bintu et al., 2005a). In this way geometrical details of the protein-DNA interaction as well as properties of the DNA such as supercoiling can be absorbed into effective parameters. It is crucial to note that the particular details of the worm-like chain are *not* realized in vivo. The periodicity and peaked behavior corresponding to the interplay between the energy cost to bend the DNA at short distances and the entropy cost associated with the ends being in close proximity for large *L*. Since the looping capacity function differs from the J-factor by a multiplicative binding constant, we can fit the experimental data shown in Figure 1C using eqn. (S4) and the expression for the looping J-factor used in (Bintu et al., 2005a), modulated by a term that accounts for the torsional flexibility (Becker et al., 2005; Bloomfield et al., 1999)

$$\chi(L) = \exp\left(-\frac{d}{L} + eL - b\log(L) + f + \log\left(\sum_n \exp\left(-\frac{4\pi^2 C_{twist}}{2LP^2}(L - L_{opt} + nP)^2\right)\right)\right), \tag{S5}$$

wherein *n* corresponds to the different species that can loop differing by a 2π rotation and *P* corresponds to the helical repeat. The values for the free parameters are b = 2.52, d = 144.1 (bp), e = 5.6e-3 (bp^{-1}), f = 12.6, p = 11 bp, C_{twist} = 30 bp, L_{opt} = 70 bp. Fits are shown in Figure 1 and Figure S2.

Synthetic Enhancer Circuit Operational Principles

For more details on the NRI/NRII nitrogen regulation loci in *E. coli* see (Magasanik, 1993; Ninfa and Atkinson, 2000; Ninfa and Jiang, 2005) for reviews. In short, the *glnG* (*ntrC*) gene is activated first by a σ^{70} promoter glnAp1, which overlaps the NRI#1,2 sites. This promoter keeps a low basal level of the protein product NRI available to the cell (for a discussion of the role of glnAp1 in our

thermodynamic models refer to the next section). NRI can only bind the DNA if it is phosphorylated by NRII (the gene product of *ntrB* or *glnL*). Since NRII can function both as a phosphatase and kinase dependent on the PII signal transduction protein activation state, the expression levels of endogenous NRI are therefore tightly coupled to a complex signaling pathway. As a result, we chose to use the phosphatase deficient mutant NRII2302 in a 3.300 *E. coli* strain with deletions of the endogenous *glnL* and *glnG* genes (3.300LG (Atkinson et al., 2003)). This mutant effectively decouples the expression levels of NRI from the PII signal transduction pathway, and has been used before in a synthetic clock experiment carried out with a similar circuit to the one used here (3.300LG (Atkinson et al., 2003).

In order to activate the σ^{54} promoter (see Figure 1A for schematic), the cell must accumulate a sufficient amount of phosphorylated NRI (NRI∼P) proteins in order to assemble a hexamer on the DNA, which serves as the driver for the reaction. The hexameric protein-DNA complex is assumed to be highly stable with a large binding constant (Atkinson et al., 2003; Chen and Reitzer, 1995; Magasanik, 1993). Initiation occurs when the driver complex makes contact with the poised polymerase via DNA looping, bringing in contact an amino acid loop (GAFTGA) with a binding site cleft in σ^{54}. Subsequently, the energy obtained from hydrolysis of ATP in the hexameric complex is used to release the complex and induce the proper conformational changes on the poised polymerase, which in turn forms an open complex that allows transcription to progress (De Carlo et al., 2006; Rappas et al., 2007). Thus, the entire scheme may be viewed kinetically as an elaborate and highly processive molecular motor.

Model for the Synthetic Enhancer Looping Activation

In order to model the transcriptional kinetics for the NRI∼P-σ^{54} system, we make the following assumptions (see Figure S2C schematic):

(1) There is always a bound "poised" polymerase at the glnAp2 promoter awaiting an activation signal.

(2) For the NRI∼P hexamerization, a cooperative process, the appropriate expression for equilibrium binding is given by

$$\text{probability of NRI sites occupied} = \frac{\left(\frac{[N]}{K_H}\right)^n}{1+\left(\frac{[N]}{K_H}\right)^n}, \tag{S6}$$

wherein [N] is the concentration of phosphorylated NRI∼P dimers, K_H is the NRI∼P dissociation constant that incorporates the cooperativity of the binding interaction, and n is some coefficient > 1 that signifies the multimerization of NRI∼P at the NRI#1,2 sites. One can expect *n* to be as high as 6, but it could also be lower (about 3) since NRI∼P is a dimer in solution. Hence, we expect $3 < n < 6$ (Rombel et al., 1998).

(3) We assume that the glnAp1 promoter is only active when the concentration of NRI∼P ([N]) is vanishingly small (for a justification of this assumption see the next section). When a small amount of NRI∼P accumulates, the hexameric complex assembles, which simultaneously strongly represses the glnAp1 promoter while activating glnAp2. The subsequent constant production of NRI∼P allows glnAp2 to remain "on" in steady state. Therefore, for all of our experiments we posit that:

$$\left(\frac{[N]}{K_H}\right)^n >> 1. \tag{S7}$$

(4) Finally, we will assume that the rates of NRI∼P binding, oligomerization, those of looping, and unlooping are much faster than the subsequent rates involved in transcription. This means that before ATP can be hydrolyzed and an open complex be formed at the promoter the $(NRI\sim P)^6$-DNA-σ^{54} complex gets to equilibrate. This in turn means that the DNA-bound $(NRI\sim P)^6$ complex gets to explore its conformational space such that the DNA polymer is in equilibrium. Thus, there are three possible states that the system can adopt (Figure S2C-schematic), the NRI#1,2 unoccupied, NRI#1,2 occupied by the $(NRI\sim P)^6$ hexameric complex, and the occupied state in looped configuration. In order to initiate transcription, the complex must be in the occupied looped state, and as a result we obtain the following expression for the probability to initiate transcription:

$$p_{loop} = \frac{\frac{J(L)}{K_{NR}}\left(\frac{[N]}{K_H}\right)^n}{1+\left(\frac{[N]}{K_H}\right)^n+\frac{J(L)}{K_{NR}}\left(\frac{[N]}{K_H}\right)^n}, \tag{S8}$$

wherein *J(L)* is the J-factor for DNA looping and K_{nr} corresponds to the dissociation constant between the NRI∼P-DNA complex and σ^{54}. Details of the thermodynamic models leading to this expression are reviewed in (Bintu et al., 2005a; Bintu et al., 2005b) and in the previous section of the text. Given assumptions 2 and 3, equation S8 reduces to:

$$p_{loop} \cong \frac{\frac{J(L)}{K_{nr}}}{1+\frac{J(L)}{K_{nr}}} \tag{S9}$$

Modeling glnAp1 Activity

Throughout this paper we quantify the level of gene expression from our synthetic enhancer by measuring the fluorescence of mCherry expressed by the glnAp2 promoter. This is the promoter with the poised σ^{54} RNA polymerase waiting for NRI∼P to bind to the NRI#1,2 sites and loop in order to initiate transcription. In addition to mCherry the construct expresses the actual protein NRI, resulting in feedback within the circuit. However, there is an extra promoter in the system: glnAp1 as shown in Figure S3A. This σ^{70} promoter overlaps the NRI#1,2 sites such that it is only transcriptionally active if NRI∼P is not bound. The case of high NRI∼P concentration, which our enhancer is hypothesized to operate in is denoted as "Scenario 1" in Figure S3A. However, one concern might be that if not enough NRI∼P is present in the cells to oligomerize on the NRI#1,2 sites glnAp1 becomes active as shown as "Scenario 2" in Figure S3A and this would contaminate our measurements by producing an mCherry signal which would *not* reflect the action of the synthetic enhancer. Expression off of this promoter ensures a minimal level of NRI inside the cell (Magasanik, 1996).

The models developed in this paper aim at describing the role of looping in activation of the glnAp2 promoter. However, as noted above, it is not clear a priori if the levels of NRI could be so low at any point that the glnAp1 promoter would become active, even though there is strong experimental evidence to support this assumption (Lilja et al., 2004; Magasanik, 1993; Reitzer and Magasanik, 1983).

As a sanity check, we address this question by formulating a model that addresses transcription by both promoters. The model will be based on standard kinetic rate equation schemes commonly used to model promoter activity in bacteria, but in this case will be complemented by data obtained from the literature as well as measurements carried out by us (detailed below), which will be used as input parameters. Thus, the predictions of the model will be tested with parameters whose values are based on real in vivo measurements. Together the model and experimental data obtained from the literature will comprise the consistency check that we seek.

We formulate the model with a simplified scheme where we don't consider translation explicitly. As a result, the change in concentration of NRI∼P, [*N*] over time is given by

$$\frac{d[N]}{dt} = \alpha_{70}\frac{1}{1+\left(\frac{[N]}{K_H}\right)^n} + \alpha_{54}\frac{\chi(L)\left(\frac{[N]}{K_H}\right)^n}{1+(1+\chi(L))\left(\frac{[N]}{K_H}\right)^n} - \beta[N], \tag{S10}$$

wherein α_{70} and α_{54} are the maximum production rates of the σ^{70} glnAp1 promoter and σ^{54} glnAp2 promoter, respectively. Notice that glnAp1 is repressed by the binding of NRI∼P to the NRI#1,2 sites. The decay rate of NRI∼P is given by β and n corresponds to the cooperativity in NRI∼P assembling and binding to the NRI#1,2 sites, and as before is assumed to be: $3 < n < 6$. We do not measure the NRI∼P concentration directly. Instead we measure the amount of fluorescence due to the mCherry reporter, [*mCherry*]. Its change over time is given by:

$$\frac{d[mCherry]}{dt} = \alpha_{70}\frac{1}{1+\left(\frac{[N]}{K_H}\right)^n} + \alpha_{54}\frac{\chi(L)\left(\frac{[N]}{K_H}\right)^n}{1+(1+\chi(L))\left(\frac{[N]}{K_H}\right)^n} - \beta[mCherry], \tag{S11}$$

wherein for simplicity we have assumed that the production and degradation rates for NRI∼P and mCherry are the same. Our goal is to determine if the concentration of NRI∼P is ever low enough such that the glnAp1 promoter would become active. If, for example, the looping capacity is ever low enough such that the resulting concentration of NRI∼P is below the threshold for binding to NRI#1,2 and repression of glnAp1 then our model will have to account for this additional promoter explicitly. In order to simplify our expressions and the exploration of different parameter choices we replace the looping capacity by

$$\chi(L) = \gamma\chi_{max}, \tag{S12}$$

wherein γ is a number between 0 and 1 and χ_{max} corresponds to the maximum value of the looping capacity. Presumably, this corresponds to $L \approx 70$ as shown in Figure 1C and S2B. We will then modulate the looping capacity by tuning the value of γ.

Before we proceed, we invoke an experimental observation made by us: the level of fluorescence in the absence of NRI (measured on a synthetic enhancer construct where the NRI gene had a frame shift mutation 10 bp away from the transcriptional ATG start site) is

comparable to the maximum level of fluorescence when NRI is present at $L \approx 70$: 4580 ± 90 and 4800 ± 100 arbitrary fluorescent units, respectively. Mathematically, we represent this condition as

$$[mCherry]([N]=0) = [mCherry](L=70), \tag{S13}$$

which, using equation S11 in steady state, results in

$$\alpha_{70} = \alpha_{70}\frac{1}{1+\left(\frac{[N](L=70)}{K_H}\right)^n} + \alpha_{54}\frac{\chi_{max}\left(\frac{[N](L=70)}{K_H}\right)^n}{1+(1+\chi_{max})\left(\frac{[N](L=70)}{K_H}\right)^n}. \tag{S14}$$

We now assume that for $L = 70$ the NRI#1,2 sites are always occupied by NRI, namely that

$$\left(\frac{[N](L=70)}{K_H}\right)^n \gg 1, \tag{S15}$$

resulting in the condition

$$\chi_{max} = \frac{\alpha_{70}}{\alpha_{54}-\alpha_{70}}. \tag{S16}$$

Next, we wish to obtain estimates for the NRI∼P dissociation constant, K_H, and the Hill-coefficient, n. We do this by examining the data by (Rombel et al., 1998). Here, the in vitro transcription rate off of glnAp2 was measured as a function of the concentration of NRI∼P in solution. Since glnAp1 is not active in this case, the rate of transcription in steady state normalized by the maximum rate of transcription can be calculated from equation

$$\text{Normalized transcription rate} = \frac{\chi(L)\left(\frac{[N]}{K_H}\right)^n}{1+(1+\chi(L))\left(\frac{[N]}{K_H}\right)^n}\frac{1+\chi(L)}{\chi(L)}. \tag{S17}$$

In Figures S3B and S3C we show their data combined with the normalized transcription rate from equation S17 for several choices of parameters. Regardless of the choice of value for the looping capacity, χ, it is clear that the cooperativity in binding of NRI∼P to DNA, n, needs to be higher than three and its dissociation constant is on the order of 10nM.

With this information in hand we now calculate the expected level of NRI∼P inside the cell given several choices of α_{54}, α_{70}, and γ as defined in equation S12. First, we determine the value of χ_{max} from α_{54} and α_{70} using equation S16. Second, we solve for the concentration of NRI∼P in steady state using equation S10. Finally, given the calculated concentration of NRI∼P we determine the probability of finding it bound to the NRI#1,2 sites thus repressing glnAp1. In Figure S3D-E, we present the probability of finding NRI∼P bound to the NRI#1,2 sites as a function of α_{54} and the ratio α_{70}/α_{54}. According to (Magasanik, 1996) the level of NRI inside the cell under maximally activating conditions (corresponding to nitrogen starvation) is roughly 70 per cell. Assuming an NRI decay rate given by dilution due to cell division of $\beta = 0.0116$ min^{-1} (corresponding to a division time of one hour) this puts a lower bound on the value of the transcription rate off of the σ^{54} promoter, namely,

$$\alpha_{54} > 70 \text{ molecules/cell} \times 0.0116 \text{ min}^{-1} \approx 0.8 \text{ nM/min}, \tag{S18}$$

wherein we have used the simple rule of thumb that one molecule inside the cell corresponds to a concentration of 1nM. From Figure S3D-E we can see that for values of α_{54} higher than this bound, there is a significant range of parameters that results in an almost maximal presence of NRI∼P at the NRI#1,2 sites. This gives us confidence in the thermodynamic models developed throughout the text that just account for transcription from glnAp2. In order to determine this conclusively, the in vivo rates will have to be measured explicitly. At present we only possess relationship (S16) between the transcriptional rates, which is a result of direct experimental measurements. Thus, experimental estimates for the value of χ_{max} and β are still needed (the in vivo decay rate of NRI∼P) to complement this result and the values for n and K_H which we already possess.

General Model for Enhancer-Based Repression

In the conventional model for simple repression, one assumes that a repressor inhibits expression by either competing for the RNAP binding site, or by interfering with the ability of polymerase to initiate transcription (Alberts et al., 2002). Using thermodynamics models of transcription, it is then possible to derive a simple expression for "repressed" protein expression levels: (Phillips et al., 2009):

$$\frac{P_r}{P_o} = \frac{1}{1 + \frac{[R]}{K_d}} \tag{S19}$$

wherein P_r corresponds to the concentration of a repressed gene product, P_o is the unrepressed concentration, K_d the repressor dissociation constant, and *[R]* the repressor concentration. The behavior obtained for this simple model is a decay, which depends approximately inversely on the repressor concentration.

Given this analysis, what kind of expression do we expect to obtain for repression in looping-initiated transcriptional architectures? In Figure 2C, we envision a thermodynamic model, where a transcription factor binding site for TetR is placed at a certain distance downstream from the NRI#1,2 binding sites. The distance is chosen such that the repressor does not interfere with the cooperative assembly of the $(NRI{\sim}P)^6$ driver complex. Unlike the simple case schematized in Figure 1C, here there are two additional states that we need to enumerate: a looped state with TetR bound and an unlooped state with TetR bound. Following the same logic in constructing the thermodynamic model, the looped state with TetR bound can also generate transcription, but with a different looping capacity function we term $\chi_{short}(L)$ (see associated Figure S3) as opposed to $\chi_{long}(L)$, which now refers to the looping capacity in the absence of TetR on the DNA. As a result, we can derive the following expression in steady state for the repressed protein expression levels:

$$p_{r,1} \propto \frac{\chi_{long}(L) + \frac{[T]}{K_{TD}}\chi_{short}(L)}{1 + \chi_{long}(L) + \frac{[T]}{K_{TD}}(1 + \chi_{short}(L))} \tag{S20}$$

wherein *[T]* corresponds to the concentration of TetR inside the cell, and K_{TD} is the binding constant for the Tet 02 site (~10 pM (Hillen and Berens, 1994)). In order to obtain an expression for repression as quantified in Figure 2B, we divide eqn. (S20) by the level of expression in the absence of repressor (eqn. (S4)) to get:

$$\frac{p_{r,1}}{p_o} = \frac{1 + \frac{[T]}{K_{TD}}\frac{\chi_{short}(L)}{\chi_{long}(L)}}{1 + \frac{[T]}{K_{TD}}\left(\frac{1 + \chi_{short}(L)}{1 + \chi_{long}(L)}\right)}. \tag{S21}$$

At saturating concentrations of TetR this expression reduces to

$$r_1(L) \equiv \lim_{[T]\to\infty} \frac{p_{r,1}}{p_0} = \frac{\frac{\chi_{short}(L)}{\chi_{long}(L)}}{\frac{1 + \chi_{short}(L)}{1 + \chi_{long}(L)}}. \tag{S22}$$

This corresponds to the magnitudes shown in Figure 2B and Figure 3B,D. Using such data, we can relate $\chi_{long}(L)$ and $\chi_{short}(L)$.

Though understanding the actual quantitative details of the functional form of $\chi_{short}(L)$ is beyond the scope of this paper, it is instructive to consider a toy model of the long length or entropic limit for the polymer chain. In this limit the length of the loop is much bigger than the typical size of the intervening molecular players. Thus, the looping J-factor can be approximated by the entropy of bringing the two DNA sites together. We can quantify this by the J-factor for DNA closure, which in the simple model of a Gaussian chain scales as $L^{-3/2}$. Therefore, in this limit the two looping capacities reduce to

$$\begin{aligned} \lim_{L\to\infty} \chi_{long}(L) &= \gamma_{long} L^{-\frac{3}{2}} \\ \lim_{L\to\infty} \chi_{short}(L) &= \gamma_{short} L^{-\frac{3}{2}} \end{aligned} \tag{S23}$$

wherein γ_{short} and γ_{long} are constants that are dependent on the particular physical properties of the loop. These looping capacities will approach $\chi = 1$ for an appropriate large length, which implies that in this limit eqn. (S22) approaches the constant value:

$$\lim_{L\to\infty} r_1(L) = \frac{\gamma_{short}}{\gamma_{long}}. \tag{S24}$$

The point of this argument is to illustrate how each of the different constructs could have a different value of the repression (r(L)) in the large L limit. The long length regime for each of the constructs is highlighted in Figure 2B by the dashed lines. However, a more rigorous treatment of this limit would require knowledge of the absolute values of the looping capacities.

Fitting the Bimodal Repression Data of Figure 2

Given the long-length limit derived in the previous section, and the data shown in Figure 2B, we assume that we can approximate the 1-Tet case looping capacity function by relating it to the non-repressed looping capacity function by a simple shift and rescaling term:

$$\chi_{short1}(L) = \gamma_{s1}\chi_{long}(L - L_{s1}) + \varphi_1, \tag{S25}$$

wherein L_{s1} is a length scale that corresponds approximately to the TetR footprint plus the unprotected DNA between the TetR and the driver complex (~45-50 bp), φ_1 corresponds to some small constant intended to quantify the approximately constant strong repression observed in the short looping-length regime (L < 70 bp), and γ_{s1} is the geometric scaling factor, which is defined as:

$$\gamma_{s1} \equiv \frac{\gamma_{short}}{\gamma_{long}} \tag{S26}$$

The qualitative rationalization for the length shift approximation is based on the assumption that in the short-range rigid looping regime the TF binding sites sequester the portion of the DNA it binds to and severely hinders it from bending (Figure 2A light blue shades). As a result of this length sequestration, the looping segment is now effectively shortened thus shifting the looping capacity function χ(L) to the right leaving only the upstream portion of the DNA (L-L_{s1}) free to bend and form a loop. In Figure S4B, we plot three examples of this approximated looping-capacity function using different lengths for the cassette length shift, short-range looping capacity values for the rigid regime, and geometric scale factor. Note, that we use the form that does not include the periodic modulation, as the repression data does not exhibit a modulation signature.

By plugging eqn. (S25) into eqn. (S22), we generate the following expression which can be used for fitting the repression data for the 1-Tet case presented in Figure 2A:

$$r_1(L) = \frac{\dfrac{\gamma_{s1}\chi_{long}(L - L_{s1}) + \varphi_1}{\chi_{long}(L)}}{\dfrac{1 + \gamma_{s1}\chi_{long}(L - L_{s1}) + \varphi_1}{1 + \chi_{long}(L)}}, \tag{S27}$$

Furthermore, we can derive identical expressions for the 2-Tet and 3-Tet binding cassettes, assuming that TetR is present in saturating concentrations as follows:

$$r_2(L) = \frac{\dfrac{\gamma_{s2}\chi_{long}(L - L_{s2}) + \varphi_{s2}}{\chi_{long}(L)}}{\dfrac{1 + \gamma_{s2}\chi_{long}(L - L_{s2}) + \varphi_{s2}}{1 + \chi_{long}(L)}}, \tag{S28}$$

and

$$r_3(L) = \frac{\dfrac{\gamma_{s3}\chi_{long}(L - L_{s3}) + \varphi_{s3}}{\chi_{long}(L)}}{\dfrac{1 + \gamma_{s3}\chi_{long}(L - L_{s3}) + \varphi_{s3}}{1 + \chi_{long}(L)}}, \tag{S29}$$

wherein L_{s2}, L_{s3}, γ_{s2}, γ_{s3}, φ_{s2}, φ_{s3} are the cassette length scale, geometric scale factors, and short range rigidity-regime capacity values for the 2 and 3-Tet cassettes respectively. In Figure S4C, we plot the results of eqn. (S27-S29) using the approximated repressed looping capacity functions plotted in Figure S4B, and a clear bimodal behavior is obtained for all approximated functions.

We use eqn. (S27-S29) to fit the data in Figure 2B with the cassette length shift value and geometric scale factors as free parameters. In the case of the 1-Tet and 2-Tet cassettes the fits to the data (green and red lines) yield L_{S1} and L_{S2} values, which match well with the designed cassette lengths of 45 and 87 bp, respectively. The fit to the 3-Tet cassette (purple line) yields a shift length of ~129 bp which is somewhat larger than the designed cassette size of 122 bp.

In addition, the fits to the geometric scale factors yield values that are surprisingly close to a ratio of the length of driver footprint (L_{NRI}~60 bp – (Hervás et al., 2009)) to the contour length of the total bound enhancer (driver+TetR cassette), expressed as follows:

$$\gamma_{sn} \sim \frac{L_{NRI}}{L_{NRI} + L_{sn}}. \tag{S30}$$

However, it is not clear what the significance of this ratio might be. The results obtained from fitting the data with γ_{sn} (where subscript *n* corresponds to the number of TetR binding sites): $\gamma_{s1} = 0.60 \pm 0.07$, $\gamma_{s2} = 0.37 \pm 0.06$, $\gamma_{s3} = 0.23 \pm 0.05$, compare favorably with the values that can be computed based directly on sequence data of $\gamma_{s1} = 0.51$, $\gamma_{s2} = 0.37$, and $\gamma_{s3} = 0.29$. Thus, the fit to the data seems to suggest that the repressed looping capacity function can be derived phenomenologically as a simple modification of the non-repressed looping capacity as follows:

$$\chi_{sn}(L) = \frac{L_{NRI}}{L_{NRI} + L_{sn}} \chi(L - L_{sn}). \tag{S31}$$

It is important to note, this result is based purely on empirical considerations, and a theoretical derivation from first principles is outside of the scope of the model presented in this paper.

Theory: Model for Enhancer Repression via Induction

Our second class of models is designed to expand on the first class of models to include regulatory effects that are generated by small molecule induction. We will show that the main experimental features displayed by the synthetic enhancers (in the presence of an inducer - Figure 3, Figure 4, and Figure 7), namely the existence of different discrete steps and the transitions between them, can be recapitulated using a thermodynamic model (see the fits for the 1-Tet, 2-Tet, and 3-Tet data in Figure S5). As before, we assume that this is an equilibrium process given by the probability of the intervening DNA looping and NRI∼P contacting RNA polymerase and that the presence of different TetR molecules modifies this probability.

We develop the repression via induction model, by first formulating the description for TetR induction. We then couple the induction model to the looping initiation model described in the previous section for the 1, 2 and 3 binding site cases. Finally, we generalize the TetR induction model for a general inducer, to model the TraR data.

TetR Induction

We begin by considering the induction of TetR by aTc. Each monomer of the TetR dimer can be bound to one aTc molecule. We assume that this binding is independent, that is, that there is no cooperativity in the binding of the second aTc molecule once the first one is already bound. This assumption is in part supported by experimental evidence (Lederer et al., 1996; Lederer et al., 1995).

With these assumptions we define the different species of TetR-aTc as shown in Figure 5A. Here, [*A*] is the concentration of aTc and K_{AT} is the dissociation constant of one aTc molecule and one binding site on TetR. If T_{tot} is the total number of TetR molecules inside the cell then the number of molecules of the different species is given by

$$T = T_{tot} \frac{1}{1 + 2\frac{[A]}{K_{at}} + \left(\frac{[A]}{K_{at}}\right)^2}, \tag{S32}$$

for free TetR,

$$AT = T_{tot} \frac{2\frac{[A]}{K_{at}}}{1 + 2\frac{[A]}{K_{at}} + \left(\frac{[A]}{K_{at}}\right)^2}, \tag{S33}$$

for TetR bound to a single aTc molecule, and

$$ATA = T_{tot} \frac{\left(\frac{[A]}{K_{at}}\right)^2}{1 + 2\frac{[A]}{K_{at}} + \left(\frac{[A]}{K_{at}}\right)^2}, \tag{S34}$$

for TetR bound to two aTc molecules. With these results in hand we can now consider the binding of the different species to DNA and its effect on NRI∼P-DNA looping.

To attempt a semiquantitative confrontation of the model and our data, certain additional facts such as the number of TetR molecules per cell are needed. TetR is expressed off of the glnL promoter on the pACT-Tet plasmid. Under nitrogen excess conditions this promoter expresses on the order of 10 molecules per cell (Reitzer and Magasanik, 1983). The pACT-Tet plasmid has a ColE1 origin of replication which results in about 60 plasmids per cell (Lutz and Bujard, 1997). Thus we expect on the order of 300 TetR dimers per cell.

Single TetR-Binding Site

When a single binding site for TetR is present downstream from the NRI#1,2 binding sites it affects the probability of looping between the NRI∼P and σ^{54} poised polymerase. In the absence of TetR the looping capacity is given by $\chi_{long}(L)$, whereas in the presence of TetR the looping capacity will be reduced to $\chi_{short}(L)$.

In order to describe the binding of TetR to DNA we assume that when not specifically bound, most TetR is bound non-specifically to the *E. coli* genome. This is consistent with in vitro measurements of the non-specific dissociation constant of around to 1mM/bp (Kleinschmidt et al., 1988), which would result in more than 80% of the proteins being bound to any of the $N_{NS} = 5x10^6$ base pairs. Following (Bintu et al., 2005b) the partition function for TetR binding to the DNA is given by

$$Z_{1Tet} = 1 + \frac{T}{N_{NS}} e^{-\beta\Delta\varepsilon_{TD}} + \frac{AT}{N_{NS}} e^{-\beta\Delta\varepsilon_{ATD}} + \frac{ATA}{N_{NS}} e^{-\beta\Delta\varepsilon_{ATAD}}, \tag{S35}$$

wherein T is the number of TetR molecules per cell not bound to aTc, AT corresponds to the number of molecules bound to a single aTc molecule and ATA is the number of TetR molecules bound to two aTc molecules. The binding energy is defined, for example, as $\Delta\varepsilon_{TD} = \varepsilon^{S}_{TD} - \varepsilon^{NS}_{TD}$, where ε^{S}_{TD} is the specific binding energy of TetR to operator DNA and ε^{NS}_{TD} is its binding energy to non-specific DNA. The difference in binding energy between specific and non-specific DNA are defined in an analogous fashion for aTc-TetR and aTc_2-TetR. To connect directly to the biochemical parameters presented in Table S4 we switch to a description of the partition function in eqn. (S35) in terms of dissociation constants. In order to do this we define (Bintu et al., 2005b):

$$e^{-\beta\Delta\varepsilon_{TD}} = e^{-\beta\left(\varepsilon^{S}_{TD} - \varepsilon^{NS}_{TD}\right)} = \frac{K^{NS}_{TD}}{K^{S}_{TD}} = \frac{1}{K_{TD}}, \tag{S36}$$

which results in the partition function

$$Z_{1Tet} = 1 + \frac{T}{N_{NS}}\frac{1}{K_{TD}} + \frac{AT}{N_{NS}}\frac{1}{K_{ATD}} + \frac{ATA}{N_{NS}}\frac{1}{K_{ATAD}}, \tag{S37}$$

The resulting effective dissociation constants are shown in Table S4 for the O1 operator. Note that we model the number of molecules per cell rather than concentrations, which requires the usage of dimensionless effective dissociation constants as defined above. One can convert this notation into concentrations by dividing the binding constants and molecule numbers by an estimated cell volume, typically assumed to be 10^{-15} L.

Notice how small the ATA term is. If we assume that we have on the order of 300 TetR molecules per cell even in the case where all of these molecules are bound to two aTc molecules we get

$$\frac{ATA}{N_{NS}}\frac{1}{K_{ATAO1}} = \frac{300}{5\times10^6}\frac{1}{3\times10^{-3}} = 0.02, \tag{S38}$$

whereas the T and AT terms would be on the order of 10^5 and 500, respectively. We therefore choose not to consider binding of the ATA species to DNA from now on. It will still be a relevant species in solution, but it will not have a direct effect on the inhibition of NRI∼P-σ^{54} looping.

Using the partition function in Equation (S37) we can calculate the probability of the single binding site being empty,

$$p_{1Tet,0} = \frac{1}{Z_{1Tet}}, \tag{S39}$$

or occupied by either T or AT

$$p_{1Tet,1} = \frac{1}{Z_{1Tet}}\left(\frac{T}{N_{NS}}\frac{1}{K_{TD}} + \frac{AT}{N_{NS}}\frac{1}{K_{ATD}}\right). \tag{S40}$$

Now, we consider the effect of having a TetR bound near NRI∼P on DNA looping. As shown above, the partition function for activation by NRI∼P in the absence of TetR is given by

$$Z_{NRI\sim P} = 1 + \frac{J(L)}{K_{NR}} = 1 + \chi(L), \tag{S41}$$

wherein the J-factor [*J*] corresponds to the local concentration of NRI∼P in the vicinity of the σ^{54} RNA polymerase and K_{NR} is the dissociation constant between these two complexes. These two quantities are collapsed into $\chi(L)$, which we earlier dubbed the looping capacity. In the absence of TetR we redefine the looping capacity as $\chi_{long}(L)$, whereas in the presence of TetR the corresponding looping capacity will be $\chi_{short,1}(L)$. With this in hand we can write the total partition function corresponding to the states and weight shown in Figure 5B

$$Z_{NRI\sim P,1Tet} = 1 + \chi_{long}(L) + \left(1 + \chi_{short,1}(L)\right)\left(\frac{T}{N_{NS}K_{TD}} + \frac{AT}{N_{NS}K_{ATD}}\right). \quad \text{(S42)}$$

The probability of NRI∼P contacting RNA polymerase is given by

$$p_{r,1}(L) = \frac{\chi_{long}(L) + \chi_{short,1}(L)\left(\frac{T}{N_{NS}K_{TD}} + \frac{AT}{N_{NS}K_{ATD}}\right)}{Z_{NRI\sim P,1Tet}}. \quad \text{(S43)}$$

We can combine this expression with equations (S32) and (S33) in order to obtain $p_{r,1}(L)$ as a function of the concentration of aTc. Finally, repression measurements are plotted throughout the text as a ratio of the repressed expression levels (i.e., with bound TetR) to unrepressed levels. The latter corresponds to saturating concentrations of aTc, which we assume is equivalent to taking the limit of no TetR

$$p_o(L) \equiv \lim_{T_{tot}\to 0} p_{r,1}(L) = \frac{\chi_{long}(L)}{1 + \chi_{long}(L)}. \quad \text{(S44)}$$

We will therefore compare the quantity $p_{r,1}/p_o$ to our experimental results.

Two TetR-Binding Sites

To extend the model to the two TetR binding site case, we begin by defining different looping capacities. As for the 1-Tet case, when TetR is absent the corresponding looping capacity is given by $\chi_{long}(L)$, and likewise, $\chi_{short,2}(L)$ is associated with having the cassette fully occupied by TetR proteins. In addition, we add a third looping capacity parameter $\chi_{\text{int}}(L)$, corresponding to having either one of the binding sites occupied TetR, while the other remains unoccupied. The subscripts make explicit reference to the length of the loop corresponding to the given state of TetR occupancy.

Furthermore, we include a short-range interaction term, ω_s, between the two DNA-bound TetR molecules. The parameter ω_s corresponds to a cooperativity measure between the two binding sites. If $\omega_s > 1$ then the proteins bind cooperatively and the doubly bound state is more stable. Alternatively, if $\omega_s < 1$ then the proteins bind anti-cooperatively and the doubly bound state is less stable. As discussed in the text, when this factor is smaller than 1, the model generates steps in gene expression similar to those seen experimentally.

Figure 5C gives a cartoon representation of all of the different microscopic states available to the system in this model and their corresponding statistical weights obtained by computing the product of the Boltzmann factor with their corresponding microscopic degeneracies. Summing over all of these statistical weights results in the partition function:

$$Z_{NRI\sim P,2Tet} = 1 + \chi_{long}(L) + (1 + \chi_{\text{int}}(L))\left(\frac{2T}{N_{NS}K_{TD}} + \frac{2AT}{N_{NS}K_{ATD}}\right) + \left(1 + \chi_{short,2}(L)\right)\left(\left(\frac{T}{N_{NS}K_{TD}}\right)^2 + \left(\frac{AT}{N_{NS}K_{ATD}}\right)^2 + \frac{2(T)(AT)}{(N_{NS})^2K_{TD}K_{ATD}}\right)\omega_s, \quad \text{(S45)}$$

which leads to the following expression for the looping probability:

$$p_{r,2}(L) = \frac{\left[\chi_{long}(L) + \chi_{\text{int}}(L)\left(\frac{2T}{N_{NS}K_{TD}}\chi_{\text{int}}(L) + \frac{2AT}{N_{NS}K_{ATD}}\chi_{\text{int}}(L)\right) + \chi_{short,2}(L)\left(\left(\frac{T}{N_{NS}K_{TD}}\right)^2 + \left(\frac{AT}{N_{NS}K_{ATD}}\right)^2 + \frac{2(T)(AT)}{(N_{NS})^2K_{TD}K_{ATD}}\right)\omega_s\right]}{Z_{NRI\sim P,2Tet}}. \quad \text{(S46)}$$

As in the case of the single binding site construct, we exploit the quantity $p_{r,2}/p_o$ to compare to the data. This can simply be expressed as follows:

$$\frac{p_{r,2}(L)}{p_o} = \frac{\left[1 + \frac{\chi_{\text{int}}(L)}{\chi_{long}(L)}\left(\frac{2T}{N_{NS}K_{TD}} + \frac{2AT}{N_{NS}K_{ATD}}\right) + \frac{\chi_{short,2}(L)}{\chi_{long}(L)}\left(\left(\frac{T}{N_{NS}K_{TD}}\right)^2 + \left(\frac{2AT}{N_{NS}K_{ATD}}\right)^2 + \frac{2(T)(AT)}{(N_{NS})^2K_{TD}K_{ATD}}\right)\omega_s\right]}{\frac{Z_{NRI\sim P,2Tet}}{1 + \chi_{long}(L)}}. \quad \text{(S47)}$$

Three TetR-Binding Sites

In order to extend the 2-Tet model to the 3-Tet case, we first need to consider the different combinations of binding configurations and their degeneracies. Furthermore, in addition to the short-range anti-cooperativity parameter ω_s, we include a second longer-range anti-cooperativity term ω_l that measures the interaction between the TetR molecules bound to the proximal and distal binding sites. Clearly, as the architecture of the synthetic enhancer becomes increasingly complex, merely enumerating all of the states becomes laborious and there is an attendant proliferation of parameters. The partition function that emerges in this case can then be expressed as follows:

$$Z_{NtrC,3Tet} = 1 + \chi_0 + (1+\chi_{int1})\frac{1}{N_{NS}}\left(3\frac{T}{K_{TD}} + 3\frac{AT}{K_{ATD}}\right) + (1+X_{int2})\,\omega_s\frac{2}{(N_{NS})^2}\left(\left(\frac{T}{K_{TD}}\right)^2 + 2\frac{T}{K_{TD}}\frac{AT}{K_{ATD}} + \left(\frac{AT}{K_{ATD}}\right)^2\right) + (1+\chi_{int2})\,\omega_I\frac{1}{(N_{NS})^2}$$
$$\times\left(\left(\frac{T}{K_{TD}}\right)^2 + 2\frac{T}{K_{TD}}\frac{AT}{K_{ATD}} + \left(\frac{AT}{K_{ATD}}\right)^2\right) + (1+\chi_{short,3})\,\omega_s^2\omega_I\frac{1}{(N_{NS})^3}\left(\left(\frac{T}{K_{TD}}\right)^3 + 3\frac{AT}{K_{ATD}}\left(\frac{T}{K_{TD}}\right)^2 + 3\left(\frac{AT}{K_{ATD}}\right)^2\frac{T}{K_{TD}} + \left(\frac{AT}{K_{ATD}}\right)^3\right). \quad \text{(S48)}$$

Here we have omitted the explicit dependence of the different $\chi(L)$ on length. Using (S48) we can compute the ratio $p_{r,3}(L)/p_o$ as for the 1 and 2-tet cases respectively. While the analysis is outside of the scope of the present work, for completeness we show in Figure S5 fits the 1-Tet, 2-Tet, and 3-Tet data sets shown in the text.

Model for Regulatory Output for the TraR Synthetic Enhancer

TraR is a quorum sensing transcriptional activator belonging to the LuxR family, and is found in the plant pathogen *Agrobacterium tumefaciens*. It can only bind specific DNA sequences (Luo and Farrand, 1999; White and Winans, 2007) provided a specific AHL ligand N-(3-oxo-octanoyl)-L-homoserine (3OC8) is available to the cell. The binding of 3OC8 is claimed to not only stabilize the protein from degradation (Zhu and Winans, 2001), but also allows it to bind its binding site on DNA in a specific fashion. There is substantial biochemical and crystallographic evidence (Qin et al., 2000; Vannini et al., 2002; Zhang et al., 2002) that shows that ligand-free TraR is monomeric, inactive, and has a short half-life, while the ligand-bound form of TraR is a stable, active dimeric protein with high affinity to DNA. Due to its incompatibility with the σ^{70} subunit of *E. coli*, TraR when bound by its cognate ligand, is simply a dimeric DNA binding protein that can bind specific sequences (i.e., Tra Box (Luo and Farrand, 1999)).

In order to generate model predictions for the TraR enhancer regulatory output, we note (Figure 7A) that unlike TetR, TraR binds to DNA only via the dimeric, ligand bound isoform. This, combined with experimental evidence (Zhang et al., 1993) that shows a simple Hill function relationship between the concentrations of active TraR and 3OC8, implies that we can represent the concentration of active TraR dimers in the cell using the following phenomenological relationship:

$$[OT_rOT_r] = \alpha\left(\frac{\frac{[O]}{K_o}}{1+\frac{[O]}{K_o}}\right) \quad \text{(S49)}$$

wherein $[OT_rOT_r]$ corresponds to the concentration of active TraR dimers, $[O]$ is the 3OC8 concentration inside the cell, K_o is an effective dissociation constant, and α is a constant of proportionality. However, it is important to note that the different degradation rates observed for TraR and inducer-bound TraR in *E. coli* (Zhu and Winans, 2001) might lead to slightly different functional forms relating the concentration of inducer to the concentration of active TraR dimers. As such, we think of eqn. S54 as a rough approximation of the induction function.

In order to obtain an expression for the regulatory output from the TraR synthetic enhancer, we first need to take into account the occupancy probabilities for the synthetic enhancer as function of 3OC8 concentration. Figure S6-left shows that there are eight such occupancy states corresponding to one unoccupied state, three states with one TraR molecule bound, three states with two TraR molecules bound, and one state with all three molecules bound. Next, we assume an additional eight individual looped states (Figure S6-right) characterized by different looping capacities corresponding to the transcriptionally active states. Finally, we sum up all the weights to generate the partition function, which in this case reduces to simply replacing the terms for T and AT in eqn. S48 by OT_rOT_r. This yields:

$$Z_{NtrC,3Tra} = 1 + \chi_L + (1+\chi_{int1})\frac{3}{N_{NS}}\left(\frac{OT_rOT_r}{K_{T_rD}}\right) + (1+\chi_{int2})\frac{1}{(N_{NS})^2}\left(\left(\frac{OT_rOT_r}{K_{T_rD}}\right)^2(2\omega_s+\omega_I)\right) + (1+\chi_{short,3})\frac{1}{(N_{NS})^3}\omega_s^2\omega_I\left(\frac{OT_rOT_r}{K_{T_rD}}\right)^3 \quad \text{(S50)}$$

wherein K_{T_rD} is the binding constant of active TraR to the Tra Box (3×10^{-8} M - White and Winans, 2007), assumed to be the same for all three binding sites. Finally, the probabilities can be obtained by analogous operations to the ones used in the previous sections.

The Importance of NRI #3,# 4, #5

In the natural system (Figure S7A for schematic), there are three additional putative NRI sites (#3, #4, and #5) that flank the main NRI#1,2 tandem sites and the σ^{54} binding sites. In (Atkinson et al., 2002), the authors mutated the #3 and #4 sites, and found that they act as inhibitors of expression, limiting the total output of the promoter. As a result, those sites were mutated out of our synthetic enhancer constructs. The NRI#5 site was not to our knowledge studied in a similar setting. However, another study (Lilja et al., 2004), which examined the dependence of σ^{54} expression levels on various sequences, mutated all 3 sites, and seemed to indicate that NRI#5 may play a crucial role if their results were to be reconciled with (Atkinson et al., 2002). In order to avoid uncertainty, we mutated the NRI#5 in several constructs, and checked for effects on the fluorescence levels. In the context of our experiment, where the NRI promoter is decoupled from the nitrogen levels in the cell, we found that NRI#5 does not play a crucial role in this transcriptional system. We used sequences of identical length with and without the NRI#5 site, and found no detectible signature on the output

expression levels Thus, we conclude that if the NRI#5 plays a role, it is only in the context of the more complex NRI-NRII system that is endogenous to the cells.

SUPPLEMENTAL REFERENCES

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell, Fourth Edition (New York: Garland Science).

Atkinson, M.R., Pattaramanon, N., and Ninfa, A.J. (2002). Governor of the glnAp2 promoter of Escherichia coli. Mol. Microbiol. *46*, 1247–1257.

Atkinson, M.R., Savageau, M.A., Myers, J.T., and Ninfa, A.J. (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. Cell *113*, 597–607.

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

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.

Bloomfield, V.A., Crothers, D.M., and Tinoco, I. (1999). Nucleic acids: structures, properties, and functions (Sausalito, Calif.: University Science Books).

Chen, P., and Reitzer, L.J. (1995). Active contribution of two domains to cooperative DNA binding of the enhancer-binding protein nitrogen regulator I (NtrC) of Escherichia coli: stimulation by phosphorylation and the binding of ATP. J. Bacteriol. *177*, 2490–2496.

De Carlo, S., Chen, B., Hoover, T.R., Kondrashkina, E., Nogales, E., and Nixon, B.T. (2006). The structural basis for regulated assembly and function of the transcriptional activator NtrC. Genes Dev. *20*, 1485–1495.

Flory, P.J., Suter, U.W., and Mutter, M. (1976). Macrocyclization equilibriums. 1. Theory. J. Am. Chem. Soc. *98*, 5733–5739.

Han, L., Garcia, H.G., Blumberg, S., Towles, K.B., Beausang, J.F., Nelson, P.C., and Phillips, R. (2009). Concentration and length dependence of DNA looping in transcriptional regulation. PLoS ONE *4*, e5621.

Hervás, A.B., Canosa, I., Little, R., Dixon, R., and Santero, E. (2009). NtrC-dependent regulatory network for nitrogen assimilation in Pseudomonas putida. J. Bacteriol. *191*, 6123–6135.

Hillen, W., and Berens, C. (1994). Mechanisms underlying expression of Tn10 encoded tetracycline resistance. Annu. Rev. Microbiol. *48*, 345–369.

Huo, Y.X., Tian, Z.X., Rappas, M., Wen, J., Chen, Y.C., You, C.H., Zhang, X., Buck, M., Wang, Y.P., and Kolb, A. (2006). Protein-induced DNA bending clarifies the architectural organization of the sigma54-dependent glnAp2 promoter. Mol. Microbiol. *59*, 168–180.

Jacobson, H., and Stockmayer, W.H. (1950). Intramolecular Reaction in Polycondensations. 1. The Theory of Linear Systems. J. Chem. Phys. *18*, 1600–1606.

Kiupakis, A.K., and Reitzer, L.J. (2002). ArgR-independent induction and ArgR-dependent superinduction of the astCADBE operon in Escherichia coli. J. Bacteriol. *184*, 2940–2950.

Kleinschmidt, C., Tovar, K., Hillen, W., and Porschke, D. (1988). Dynamics of repressor-operator recognition: the Tn10-encoded tetracycline resistance control. Biochemistry *27*, 1094–1104.

Lederer, T., Takahashi, M., and Hillen, W. (1995). Thermodynamic analysis of tetracycline-mediated induction of Tet repressor by a quantitative methylation protection assay. Anal. Biochem. *232*, 190–196.

Lederer, T., Kintrup, M., Takahashi, M., Sum, P.E., Ellestad, G.A., and Hillen, W. (1996). Tetracycline analogs affecting binding to Tn10-Encoded Tet repressor trigger the same mechanism of induction. Biochemistry *35*, 7439–7446.

Leonhartsberger, S., Huber, A., Lottspeich, F., and Böck, A. (2001). The hydH/G Genes from Escherichia coli code for a zinc and lead responsive two-component regulatory system. J. Mol. Biol. *307*, 93–105.

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.

Luo, Z.Q., and Farrand, S.K. (1999). Signal-dependent DNA binding and functional domains of the quorum-sensing activator TraR as identified by repressor activity. Proc. Natl. Acad. Sci. USA *96*, 9009–9014.

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.

Magasanik, B. (1993). The regulation of nitrogen utilization in enteric bacteria. J. Cell. Biochem. *51*, 34–40.

Magasanik, B. (1996). Regulation of nitrogen utilization. In *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, F.C. Neidhardt, ed. (Washington, DC: ASM Press), pp. 1344–1356.

Marky, N.L., and Olson, W.K. (1982). Loop formation in polynucleotide chains. 1. Theory of hairpin loop closure. Biopolymers *21*, 2329–2344.

Merlitz, H., Rippe, K., Klenin, K.V., and Langowski, J. (1998). Looping dynamics of linear DNA molecules and the effect of DNA curvature: a study by Brownian dynamics simulation. Biophys. J. *74*, 773–779.

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.

Ninfa, A.J., and Atkinson, M.R. (2000). PII signal transduction proteins. Trends Microbiol. *8*, 172–179.

Ninfa, A.J., and Jiang, P. (2005). PII signal transduction proteins: sensors of α-ketoglutarate that regulate nitrogen metabolism. Curr. Opin. Microbiol. *8*, 168–173.

Phillips, R., Kondev, J., and Theriot, J. (2009). Physical Biology of the Cell (New York: Garland Science).

Qin, Y., Luo, Z.Q., Smyth, A.J., Gao, P., Beck von Bodman, S., and Farrand, S.K. (2000). Quorum-sensing signal binding results in dimerization of TraR and its release from membranes into the cytoplasm. EMBO J. *19*, 5212–5221.

Reitzer, L.J., and Magasanik, B. (1983). Isolation of the nitrogen assimilation regulator NR(I), the product of the glnG gene of Escherichia coli. Proc. Natl. Acad. Sci. USA *80*, 5554–5558.

Rippe, K., von Hippel, P.H., and Langowski, J. (1995). Action at a distance: DNA-looping and initiation of transcription. Trends Biochem. Sci. *20*, 500–506.

Rombel, I., North, A., Hwang, I., Wyman, C., and Kustu, S. (1998). The bacterial enhancer-binding protein NtrC as a molecular machine. Cold Spring Harb. Symp. Quant. Biol. *63*, 157–166.

Schulz, A., Langowski, J., and Rippe, K. (2000). The effect of the DNA conformation on the rate of NtrC activated transcription of Escherichia coli RNA polymerase.sigma(54) holoenzyme. J. Mol. Biol. *300*, 709–725.

Semsey, S., Tolstorukov, M.Y., Virnik, K., Zhurkin, V.B., and Adhya, S. (2004). DNA trajectory in the Gal repressosome. Genes Dev. *18*, 1898–1907.

Swigon, D., Coleman, B.D., and Olson, W.K. (2006). Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. Proc. Natl. Acad. Sci. USA *103*, 9879–9884.

Towles, K.B., Beausang, J.F., Garcia, H.G., Phillips, R., and Nelson, P.C. (2009). First-principles calculation of DNA looping in tethered particle experiments. Phys. Biol. *6*, 025001.

Tucker, N.P., D'Autréaux, B., Studholme, D.J., Spiro, S., and Dixon, R. (2004). DNA binding activity of the Escherichia coli nitric oxide sensor NorR suggests a conserved target sequence in diverse proteobacteria. J. Bacteriol. *186*, 6656–6660.

Vannini, A., Volpari, C., Gargioli, C., Muraglia, E., Cortese, R., De Francesco, R., Neddermann, P., and Marco, S.D. (2002). The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. EMBO J. *21*, 4393–4401.

White, C.E., and Winans, S.C. (2007). The quorum-sensing transcription factor TraR decodes its DNA binding site by direct contacts with DNA bases and by detection of DNA flexibility. Mol. Microbiol. *64*, 245–256.

Zhang, L., Murphy, P.J., Kerr, A., and Tate, M.E. (1993). Agrobacterium conjugation and gene regulation by N-acyl-L-homoserine lactones. Nature *362*, 446–448.

Zhang, R.G., Pappas, T., Brace, J.L., Miller, P.C., Oulmassov, T., Molyneaux, J.M., Anderson, J.C., Bashkin, J.K., Winans, S.C., and Joachimiak, A. (2002). Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. Nature *417*, 971–974.

Zhang, Y., McEwen, A.E., Crothers, D.M., and Levene, S.D. (2006a). Analysis of in-vivo LacR-mediated gene repression based on the mechanics of DNA looping. PLoS ONE *1*, e136.

Zhang, Y., McEwen, A.E., Crothers, D.M., and Levene, S.D. (2006b). Statistical-mechanical theory of DNA looping. Biophys. J. *90*, 1903–1912.

Zhu, J., and Winans, S.C. (2001). The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. Proc. Natl. Acad. Sci. USA *98*, 1507–1512.

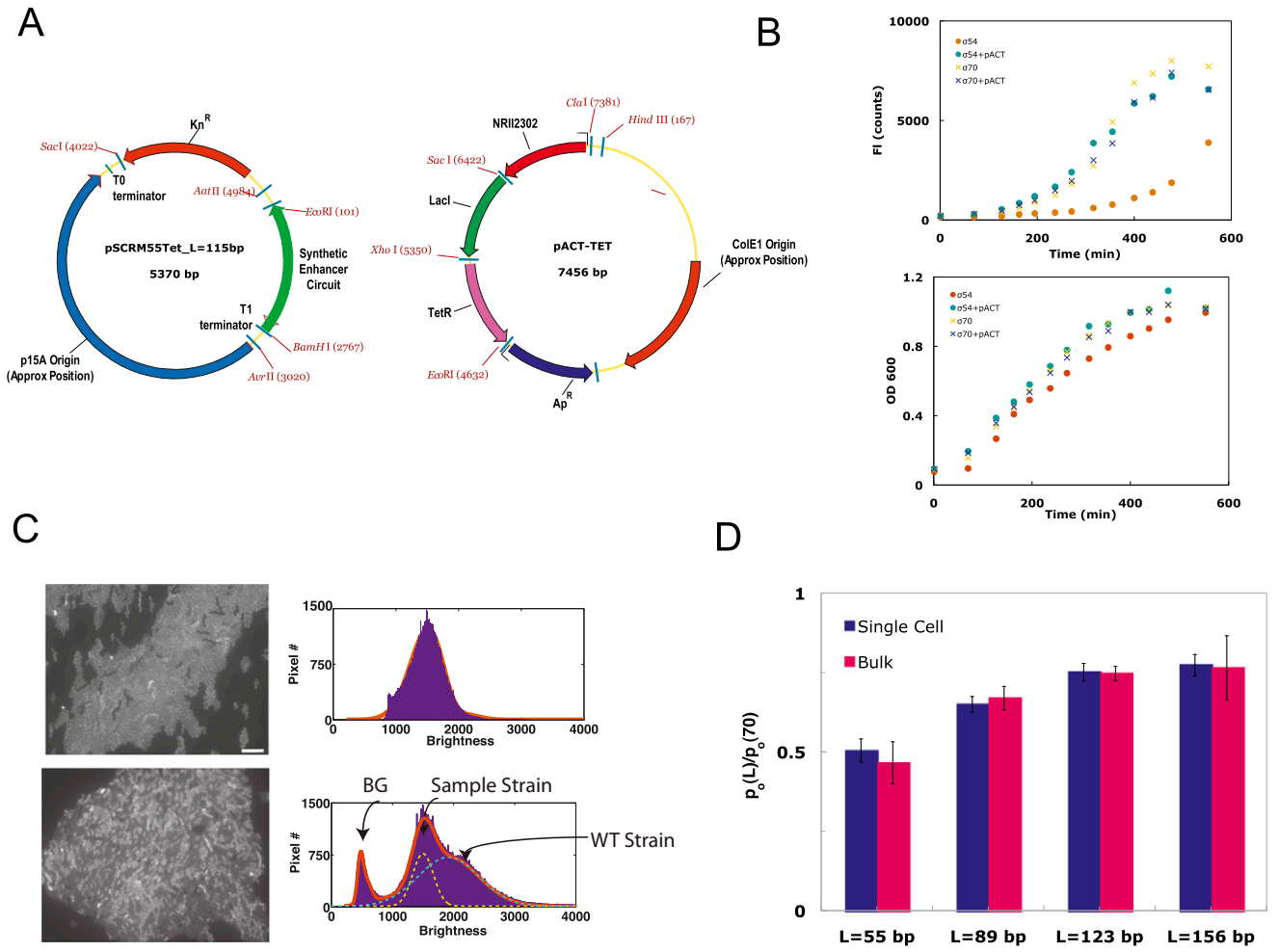

Figure S1. Plasmid Schematic, Effect of NRII2302 on Strain Growth, and Single-Cell Measurements Methodology, Related to Figure 1

(A) Plasmids schematic. The left drawing corresponds to a schematic of the synthetic enhancer plasmid containing kanamycin resistance and low copy number p15A Origin of replication. The right drawing corresponds to a schematic of the pACT-Tet plasmid, which contains the *NRII2302* mutant, *lacI*, and *tetR* genes. This plasmid has the high copy number ColE1 Origin of replication that provides high concentrations of the proteins encoded by those genes.

(B) Measurements of growth curves for synthetic enhancer strains in the presence and absence of pACT. In each graph, four strains are represented. The blue and yellow (x) correspond to strains containing the same vector background as the synthetic enhancer plasmid but with an *mCherry* gene expressed from an inducible σ^{70} pLac/Ara promoter rather than the synthetic enhancer circuit. The turquoise circles denote the synthetic enhancer strain in the presence of pACT, while the brown circles correspond to data where pACT is absent. pACT clearly does not affect the growth rate or fluorescence expression for the pLac/Ara control strain. Comparing the synthetic enhancer strains with and without pACT, we note that synthetic enhancer+pACT behaves like the control strains in growth rate and fluorescence maturation, while the synthetic enhancer strain not containing pACT is highly sensitive to growth conditions and behaves differently than the other three. The graphs indicate that pACT serves to insulate the synthetic enhancer circuit from the growth condition sensitivity that characterizes the NRI-NRII system.

(C) Top image and histogram - single cell imaging of one synthetic enhancer strain showing the uniformity in fluorescence. Scale bar corresponds to 10 μm. Bottom image and histogram - single cell measurements of expression level for synthetic enhancer strains of varying spacer lengths. Each image corresponds to a 1:1 mix by cell concentration of the brightest strain (L = 70 bp) with a sample strain. The bottom histogram displays the binning of all bright pixel values for the image to the left. The dashed lines correspond to the two cellular sub-populations obtained from a multi-peak fit (red line) to the histogram data. "BG" corresponds to the background brightness.

(D) Comparison of single cell and bulk measurements made with four different strains whose lengths are roughly separated by 35 bp. The data shows that within experimental error the results are virtually identical supporting the notion that both measurement methods are complementary.

Data and error bars for the single cell measurements are obtained from first and second moment estimates of the relevant distributions (examples are shown in Figure S1C).

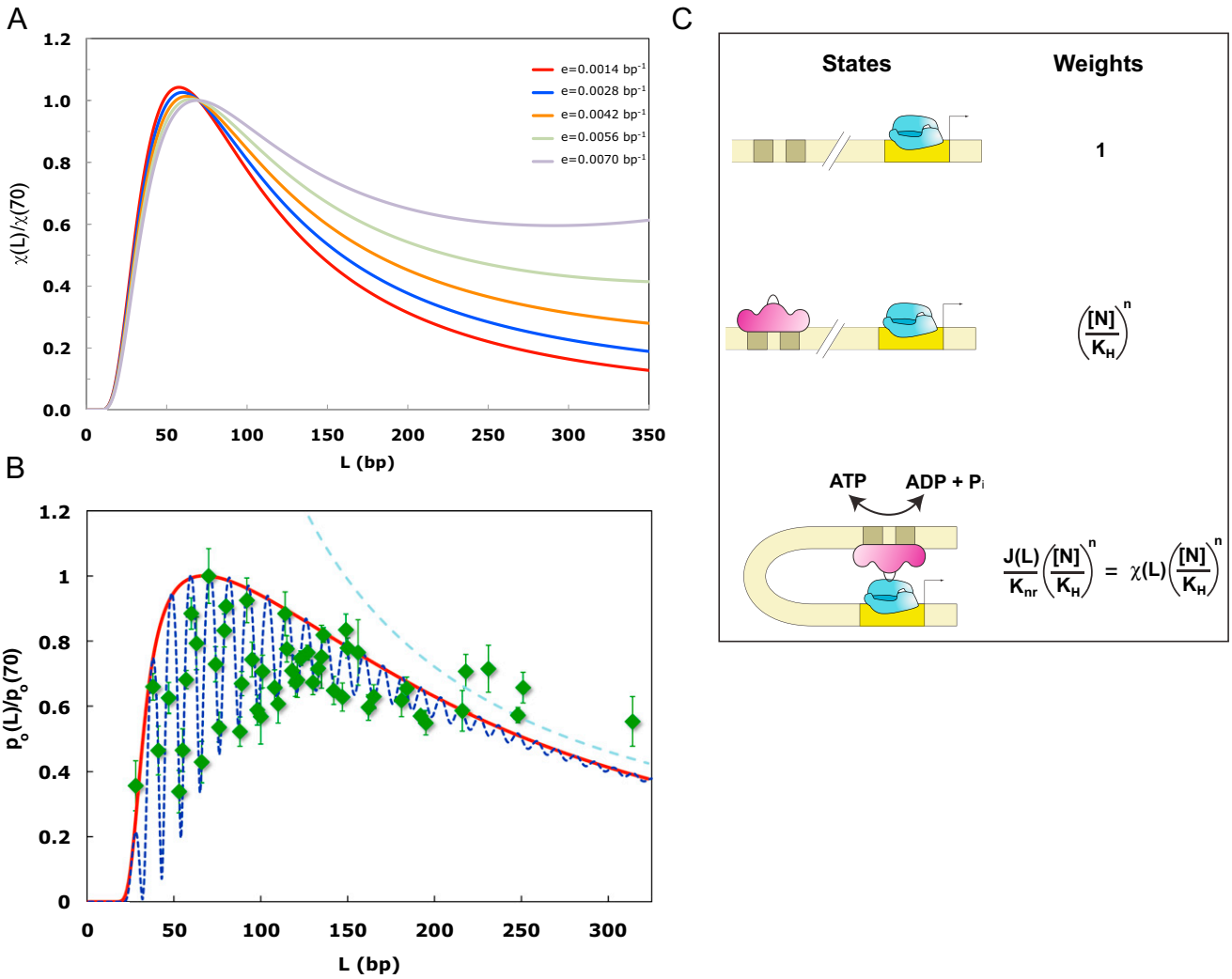

Figure S2. Model for Enhancer-Based Transcription, Related to Figure 1

(A) The panel shows several candidate looping capacity functions, which when plugged into eqn. S4 generate an adequate fit to the data. The functions differ in their long-range power law decline, which can be controlled by varying the parameter "e" in eqn. S5 (i.e., parameters "b" and "d" remain the same). This variation has important ramifications for the underlying biophysics of looping, and the binding constant associated with NRI∼P-σ^{54} interaction. The envelopes for the fits in Figure 1B and S2B are computed from the green and red curves respectively.

(B) While in the text we chose to show the "best" fit to our data generated by the model, here we fit the data with a looping capacity function that is identical (up to a multiplicative constant) to the one used to fit the LacI data obtained by (Müller et al., 1996). The fit here deviates from the data at longer lengths, which indicates that increased precision in constraining the looping capacity function may be attained by getting additional expression levels at longer loop lengths (>300 bp). The blue dashed line corresponds in this case to a 3/4 power law decline, which denotes approximately the long length scale power law decline exhibited by the red curve. Error bars correspond to the standard deviation from multiple measurements.

(C) Schematic for the thermodynamic model that includes the binding of NRI∼P. We assume throughout the paper that the large cooperativity associated with the formation of the hexameric complex is sufficient to neglect the state with NtrC unbound in all our modeling considerations.

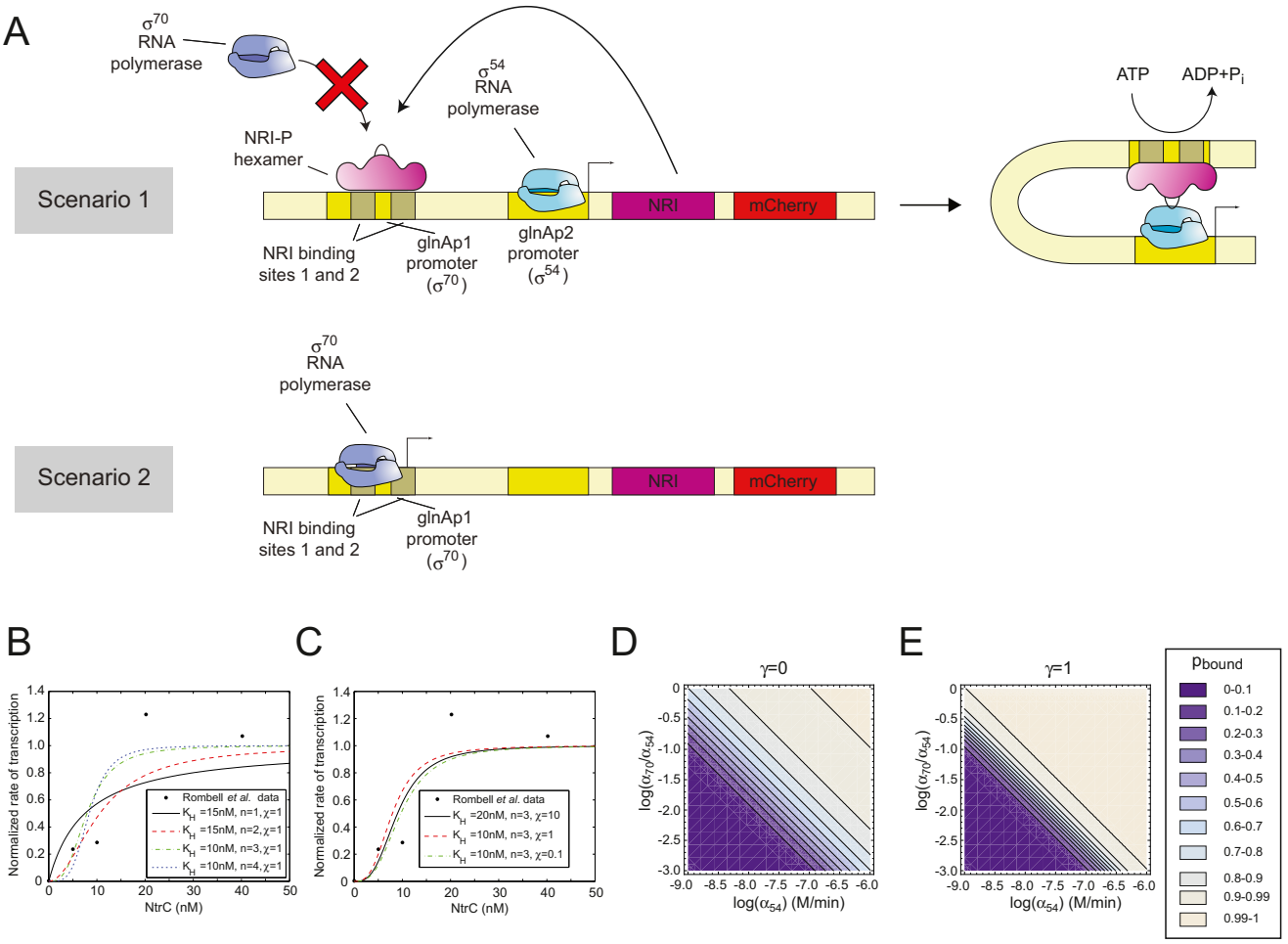

Figure S3. Model for glnAp1 Transcription, Related to Figure 1

Parameters and limits of simple thermodynamic model for the dual promoter glnAp1/glnAp2 system.

(A) Cartoon schematizing two possible scenarios. In the first case the level of NRI∼P in the cell is high enough to allow oligomerization at the NRI#1,2 sites and subsequent looping to contact the poised σ^{54} RNA polymerase at the glnAp2 promoter. Binding to the NRI#1,2 sites renders the σ^{70} glnAp1 promoter inactive. In the second case the levels of NRI∼P are too low to repress the σ^{70} glnAp1 promoter.

(B and C) The binding parameters of NRI∼P to the NR#1,2 sites are estimated from the in vitro transcription data by (Rombel et al., 1998).

(D and E) The steady state probability of occupancy of NRI∼P on the NR#1,2 sites is calculated from equation (S10) assuming $\gamma = 0$ for (D), and $\gamma = 1$ for (E). The plot in (D,E) were constrained by the value for n and K_H derived in (B,C) and by our measurement of absolute glnAp1 activity detailed by eqn. (S16).

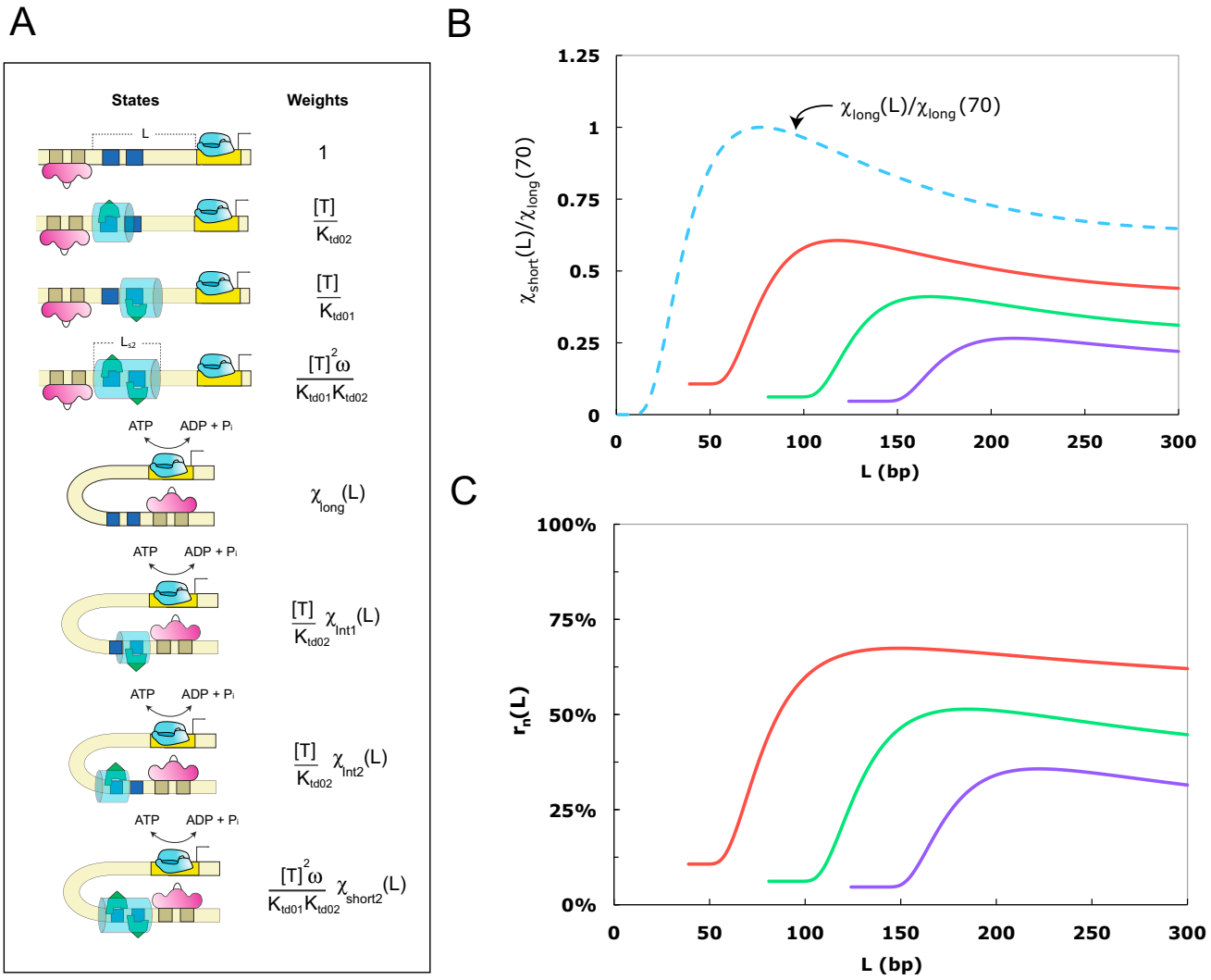

Figure S4. Fitting the Enhancer Repression Data, Related to Figure 2

(A) The schematic shows the four possible looped-active transcriptional states and four unlooped-inactive states with their associated weights for a construct containing two TetR binding site, assuming both a $(NRI{\sim}P)^6$ driver and a "poised" σ^{54}-RNAP complex are always bound to the DNA.

(B) Approximated repressed looping capacity functions (see eqn. S25 for definition) are shifted to the right and rescaled in three increments as compared with the unrepressed looping capacity functions (light blue-dashed line). We used L_t = 45, 87, and 129 for the red, green, and purple curves respectively. We utilized the unshifted looping capacity function used to fit the data in Figure 1B as a basis for computing the shifted functions. Note, that the looping capacity functions are normalized by the value of the unshifted function at L = 70 to be able to plot the curves on a 0 to 1 scale.

(C) The repression functions denoted by $r_n(L)$ are computed using equations (S27-S29) from the sample functions shown in Figure S4B. The color of each repression curve is identical to the color of the corresponding repressed looping capacity function shown in Figure S4B used for the computation.

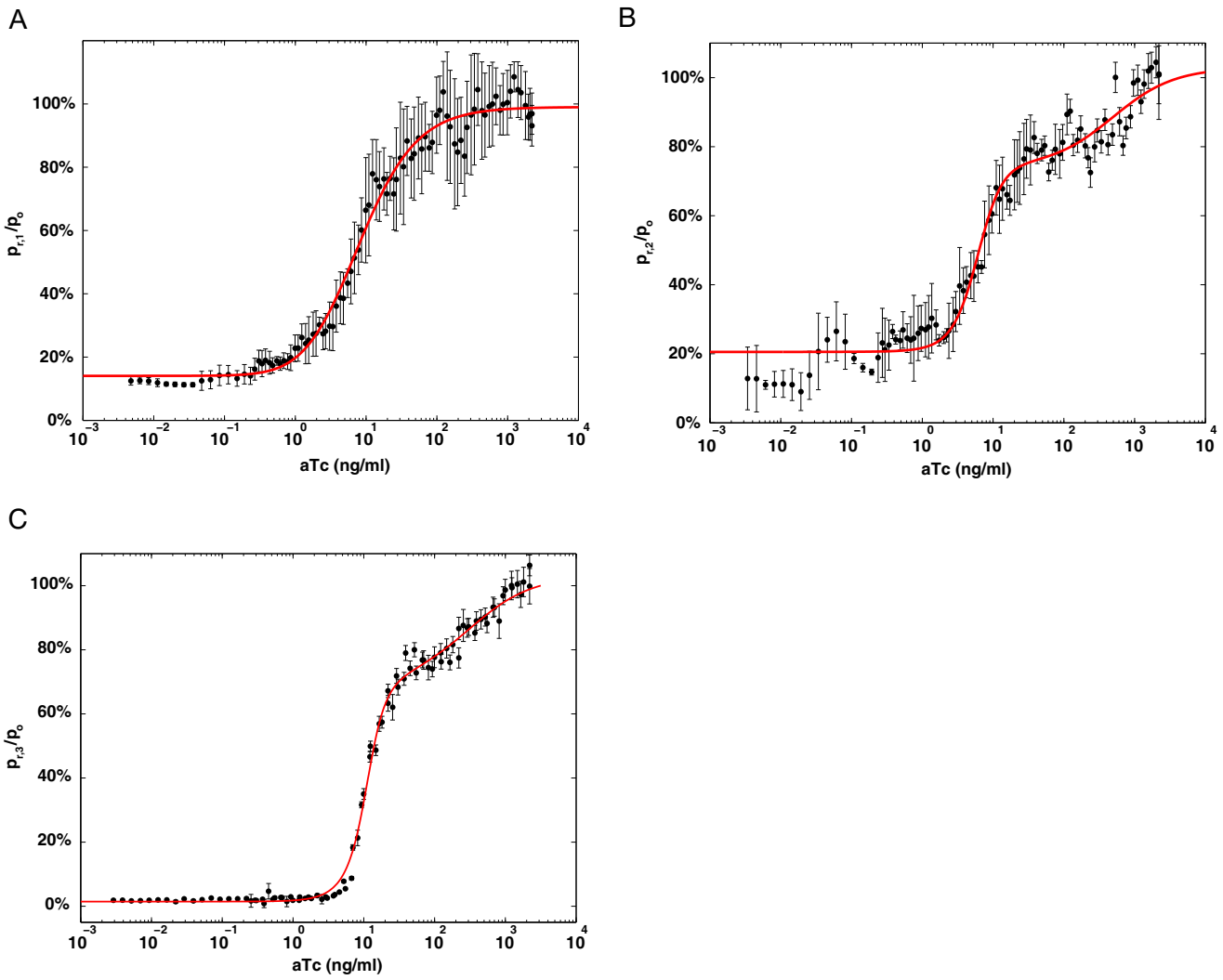

Figure S5. Model Fits to the Repression Ratio Data, Related to Figure 3, Figure 4, and Figure 6

Fits to the repression ratio data using the Model for Enhancer Repression via Induction. In panels (A), (B), (C), we show fits to the 1-Tet data used in Figure 4A, 2-Tet data shown in Figure 3A, and 3-Tet data shown in Figure 3C respectively. In all cases the model qualitatively reproduces the major features observed in the data. Error bars correspond to the standard deviation from multiple measurements for every aTc concentration.

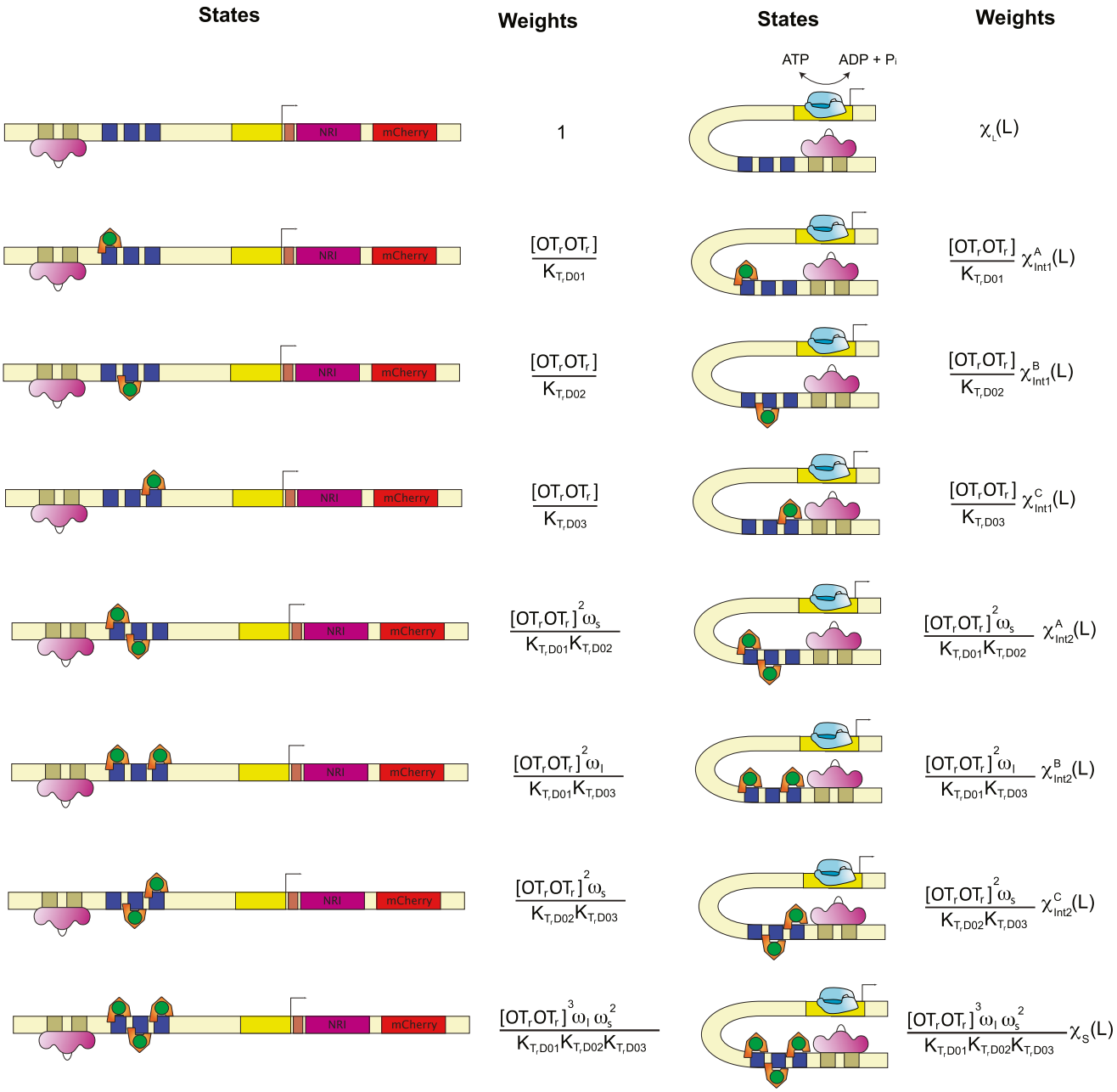

Figure S6. States and Weights Model for the 3-TraR Synthetic Enhancer, Related to Figure 7

The occupancy of the synthetic enhancer admits many different states of occupancy and DNA geometry. In this case, there are a total of 16 states with 8 unlooped and 8 looped states. For each state, there is a corresponding statistical weight obtained using the same logic as in the earlier TetR models.

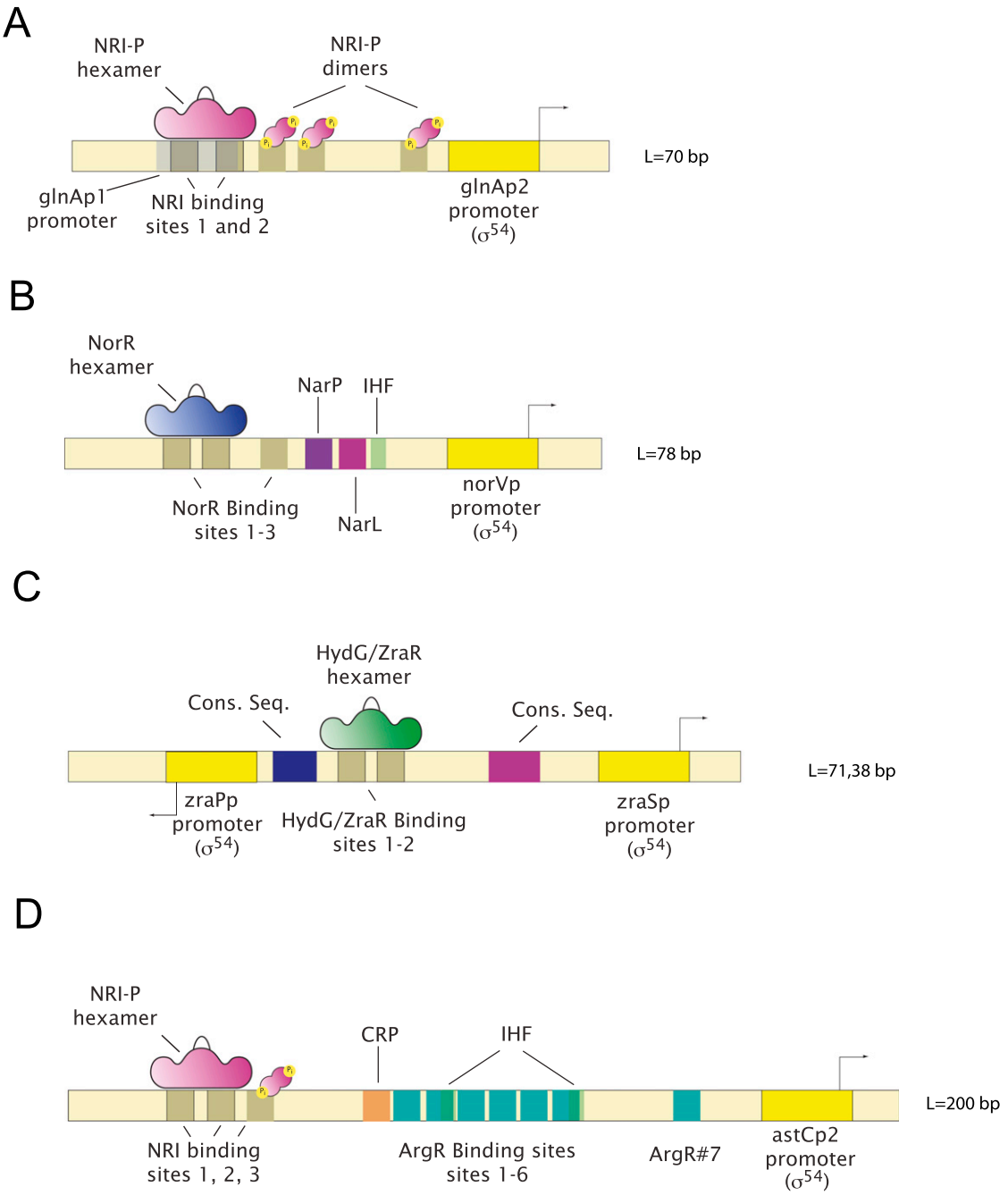

Figure S7. Examples of Natural Bacterial Enhancers, Related to Figure 7

Four examples of natural bacterial enhancers in several species of bacteria.

(A) The *E. Coli* glnAp2 enhancer, exhibiting two driver binding sites (NRI#1,2) and three additional NRI∼P binding sites shown in (Atkinson et al., 2002) to be repressing.

(B) The norVp enhancer showing two NorR driver binding sites and an additional three binding sites for the transcription factors NarP, NarL, and IHF (Tucker et al., 2004). IHF, in particular had been shown previously (Huo et al., 2006) to have the capability to both up and downregulate expression levels of bacterial enhancers based on its binding location relative to the σ^{54} promoter.

(C) A dual poised promoter system, where a single enhancer with ZraR/HydG driver binding sites drives both the zraPp and zraSp promoters. The conserved sequences are regions in the looping region that are conserved across four different species of bacteria (*E.coli, S. typhi, S. typhimurium, and K. oxytoca* – (Leonhartsberger et al., 2001)), and likely bind one or more unidentified transcription factors.

(D) The astCp2 enhancer in *E. coli*, exhibiting an ∼200 bp looping region, at least ten binding sites for three different kinds of TFs, and NRI∼P driver binding sites (Kiupakis and Reitzer, 2002). Notice that this architecture is similar to our synthetic enhancer with six TetR binding sites. It is therefore tempting to speculate that this enhancer also encodes some form of step-function for its regulatory output.