

I. Supplementary 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 calculated spacer sequence), 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 MgCl₂: 2' @95°C, 15' @65°C, 5' @42°C, and then placed on ice). Hybridized dsDNA cassette were gel purified and digested with NheI before being used as an insert in the cloning step.

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 + σ⁵⁴ 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 (Fig. S1A) was constructed by digesting the template plasmids at the NheI sites, and either reannealing 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 (Fig. 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 (Fig. S1A). Selection was carried out via double Kan/Amp resistance (20 $\mu\text{g}/\text{ml}$ and 100 $\mu\text{g}/\text{ml}$). 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.

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 μ l of the mixed culture was dispensed on slides with 1.5% agar slabs in PBS. Cells were incubated for 30 minutes 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 (Fig. S1C).

Image Analysis

Single cell data was 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 (Fig. S1C – bottom histogram.) Note, the background brightness is omitted from Fig. S1C-bottom, as it is clearly separable from the fluorescence signature in the data of the single strain sample Fig. 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 $p_o(L)/p_o(70)$ (Fig. S1D) shows that both measurements yield data that is identical within experimental error, which supports the conclusion about the robustness of the data in the text.

II. Supplementary Data

The Importance of NRI#3,4,5

In the natural system, 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 detectable 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.

III. Enhancer Activation and Repression Models

A. Model for NRI~P- σ^{54} looping activation

In order to model the transcriptional kinetics, we make the following assumptions (see Fig. S2C schematic):

- 1). There is always a “poised” polymerase waiting at the glnAp2 promoter awaiting an activation signal.
- 2). For the hexamerization process, a cooperative process, the appropriate expression for equilibrium binding is given by

$$\frac{\left(\frac{[N]}{K_H}\right)^n}{1 + \left(\frac{[N]}{K_H}\right)^n}, \quad (S1)$$

where $[N]$ is the concentration of phosphorylated NRI~P dimers, K_H is the hexamer 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 ~ 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:

$$\left(\frac{[N]}{K_H}\right)^n \gg 1. \quad (S2)$$

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 $(\text{NRI~P})^6$ -DNA- σ^{54} complex gets to equilibrate. This in turn means that the DNA-bound $(\text{NRI~P})^6$ 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 $(\text{NRI~P})^6$ 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:

$$p_{loop} = \frac{\frac{J(L)}{K_{NR}} \left(\frac{[N]}{K_H}\right)^n}{1 + \left(\frac{[N]}{K_H}\right)^n + \frac{J(L)}{K_{NR}} \left(\frac{[N]}{K_H}\right)^n}, \quad (S3)$$

where $J(L)$ is the J-factor for DNA looping described in Box 1 in the main text, 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 Box 1 of the text.

Given assumptions 2 and 3, equation S3 reduces to:

$$p_{loop} \cong \frac{\frac{J(L)}{K_{nr}}}{1 + \frac{J(L)}{K_{nr}}} \quad (S4)$$

From this point on we will make the notational replacement $\chi(L) = J(L) / K_{nr}$,

exploiting the looping capacity also defined in Box 1.

B. Modeling gInAp1 activity

Throughout this paper we quantify the level of gene expression from our synthetic enhancer by measuring the fluorescence of mCherry expressed by the gInAP2 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: gInAP1 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 gInAP1 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 gInAP2 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 gInAP1 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. In a simplified model

where we don't consider translation explicitly, the change in concentration of NRI~P, $[N]$ over time is given by

$$\frac{d[N]}{dt} = \alpha_{70} \frac{1}{1 + \left(\frac{[N]}{K_{nr}}\right)^n} + \alpha_{54} \frac{\chi(L) \left(\frac{[N]}{K_{nr}}\right)^n}{1 + (1 + \chi(L)) \left(\frac{[N]}{K_{nr}}\right)^n} - \beta[N], \quad (S5)$$

where 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

$$\frac{d[mCherry]}{dt} = \alpha_{70} \frac{1}{1 + \left(\frac{[N]}{K_{nr}}\right)^n} + \alpha_{54} \frac{\chi(L) \left(\frac{[N]}{K_{nr}}\right)^n}{1 + (1 + \chi(L)) \left(\frac{[N]}{K_{nr}}\right)^n} - \beta [mCherry], \quad (S6)$$

where 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

promoter explicitly. In order to simplify our expressions and the exploration of different parameter choices we replace the looping capacity by

$$\chi(L) = \gamma \chi_{max}, \quad (S7)$$

where 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 S3B. We will then modulate the looping capacity by tuning the value of γ .

Before we proceed we invoke an experimental observation: the level of fluorescence in the absence of NRI is comparable to 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

$$[mCherry]([N] = 0) = [mCherry](L = 70), \quad (S8)$$

which, using equation (S6) in steady state, results in

$$\alpha_{70} = \alpha_{70} \frac{1}{1 + \left(\frac{[N](L=70)}{K_{nr}} \right)^n} + \alpha_{54} \frac{\chi_{max} \left(\frac{[N](L=70)}{K_{nr}} \right)^n}{1 + (1 + \chi_{max}) \left(\frac{[N](L=70)}{K_{nr}} \right)^n}. \quad (S9)$$

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

$$\left(\frac{[N](L=70)}{K_{nr}} \right)^n \gg 1, \quad (S10)$$

resulting in the condition

$$\chi_{max} = \frac{\alpha_{70}}{\alpha_{54} - \alpha_{70}}. \quad (S11)$$

Before we proceed we wish to obtain estimates for the dissociation constant, K_{nr} , 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 (S5)

$$\text{Normalized transcription rate} = \frac{\chi(L) \left(\frac{[N]}{K_{nr}} \right)^n}{1 + (1 + \chi(L)) \left(\frac{[N]}{K_{nr}} \right)^n} \frac{1 + \chi(L)}{\chi(L)}. \quad (S12)$$

In Figures S3B and S3C we show their data combined with the normalized transcription rate from equation (S12) 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. Interestingly, the value of K_{nr} has been estimated for an *in vitro* NRI~P transcription reaction and yielded a value of 100 nM (Schulz et al., 2000), which implies a slightly weaker binding constant than that suggested by our model and the *in vivo* data.

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 (S7). First, we determine the value of χ_{max} from α_{54} and α_{70} using equation (S11). Second, we solve for the concentration of NRI~P in steady state using equation (S5). 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 Figures S3D and S3E we present the probability of finding NRI~P bound to the NRI#1,2 sites as a function of α_{54} and the ratio $\alpha_{70} / \alpha_{54}$. According to (Magasanik, 1996) the level of NRI inside the cell under maximal activating conditions corresponding to nitrogen starvation is roughly 70 per cell. Assuming an NRI decay rate given by dilution due to cell division of $\beta = 0.0116 \text{ min}^{-1}$ (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,

$$\alpha_{54} > 70 \text{ molecules/cell} \times 0.0116 \text{ min}^{-1} \approx 0.8 \text{ nM/min}, \quad (\text{S13})$$

where we have used the simple rule of thumb that one molecule inside the cell corresponds to a concentration of 1nM. From Figures S3D and S3E 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. However, in order to determine this unequivocally, the *in vivo* rates will have to be measured explicitly.

C. Enhancer Repression Model

The main features displayed by the regulatory modules, 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 Fig. S5). 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.

i. 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

$$T = T_{tot} \frac{1}{1 + 2\frac{[A]}{K_{at}} + \left(\frac{[A]}{K_{at}}\right)^2}, \quad (\text{S14})$$

for free TetR,

$$AT = T_{tot} \frac{2 \frac{[A]}{K_{at}}}{1 + 2 \frac{[A]}{K_{at}} + \left(\frac{[A]}{K_{at}} \right)^2}, \quad (S15)$$

for TetR bound to a single aTc molecule, and

$$ATA = T_{tot} \frac{\left(\frac{[A]}{K_{at}} \right)^2}{1 + 2 \frac{[A]}{K_{at}} + \left(\frac{[A]}{K_{at}} \right)^2}, \quad (S16)$$

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 semi-quantitative 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.

ii. 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 $N_{NS} = 5 \times 10^6$ base pairs. Following (Bintu et al., 2005b) the partition function for TetR binding to the DNA is given by

$$Z_{1Tet} = 1 + \frac{T}{N_{NS}} e^{-\beta \Delta \varepsilon_{TD}} + \frac{AT}{N_{NS}} e^{-\beta \Delta \varepsilon_{ATD}} + \frac{ATA}{N_{NS}} e^{-\beta \Delta \varepsilon_{ATAD}}, \quad (\text{S17})$$

where 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 $\Delta \varepsilon_{TD} = \varepsilon_{TD}^S - \varepsilon_{TD}^{NS}$, where ε_{TD}^S is the specific binding energy of TetR to operator DNA and ε_{TD}^{NS} 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 aTc₂-TetR. To connect directly to the biochemical parameters presented in Table S4 we switch to a description of the partition function in eqn. (S17) in terms of dissociation constants. In order to do this we define (Bintu et al., 2005b):

$$e^{-\beta \Delta \epsilon_{TD}} = e^{-\beta (\epsilon_{TD}^S - \epsilon_{TD}^{NS})} = \frac{K_{TD}^{NS}}{K_{TD}^S} = \frac{1}{K_{TD}}, \quad (S18)$$

which results in the partition function

$$Z_{1Tet} = 1 + \frac{T}{N_{NS}} \frac{1}{K_{TD}} + \frac{AT}{N_{NS}} \frac{1}{K_{ATD}} + \frac{ATA}{N_{NS}} \frac{1}{K_{ATAD}}, \quad (S19)$$

The resulting effective dissociation constants are shown in Table S4 for the O1 operator. Note, 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

$$\frac{ATA}{N_{NS}} \frac{1}{K_{ATAO1}} = \frac{300}{5 \times 10^6} \frac{1}{3 \times 10^{-3}} = 0.02, \quad (S20)$$

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 (S19) we can calculate the probability of the single binding site being empty,

$$p_{1Tet,0} = \frac{1}{Z_{1Tet}}, \quad (S21)$$

or occupied by either T or AT

$$p_{1Tet,1} = \frac{T}{N_{NS}} \frac{1}{K_{TD}} + \frac{AT}{N_{NS}} \frac{1}{K_{ATD}}. \quad (S22)$$

Now, we consider the effect of having a TetR bound near NRI~P on DNA looping. As shown in Box 1, the partition function for activation by NRI~P in the absence of TetR is given by

$$Z_{NRI\sim P} = 1 + \frac{J(L)}{K_{NR}} = 1 + \chi(L), \quad (S23)$$

where 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

$$Z_{NRI\sim P,1Tet} = 1 + \chi_{long}(L) + (1 + \chi_{short,1}(L)) \left(\frac{T}{N_{NS} K_{TD}} + \frac{AT}{N_{NS} K_{ATD}} \right). \quad (S24)$$

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

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

(S25)

We can combine this expression with equations (S14) and (S15) 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 unrepresed levels. The latter corresponds to saturating concentrations of aTc, which we assume is equivalent to taking the limit of no TetR

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

(S26)

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

iii. 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 $\chi_{int}(L)$, corresponding to having either one of the binding sites occupied by 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:

$$Z_{NRI \sim P, 2Tet} = 1 + \chi_{long}(L) + (1 + \chi_{int}(L)) \left(\frac{2T}{N_{NS} K_{TD}} + \frac{2AT}{N_{NS} K_{ATD}} \right) + \\ (1 + \chi_{short,2}(L)) \left(\left(\frac{T}{N_{NS} K_{TD}} \right)^2 + \left(\frac{AT}{N_{NS} K_{ATD}} \right)^2 + \frac{2(T)(AT)}{\left(N_{NS} \right)^2 K_{TD} K_{ATD}} \right) \omega_s, \quad (S27)$$

which leads to the following expression for the looping probability:

$$p_{r,2}(L) = \left[\begin{array}{l} \chi_{long}(L) + \chi_{int}(L) \left(\frac{2T}{N_{NS}K_{TD}} \chi_{int}(L) + \frac{2AT}{N_{NS}K_{ATD}} \chi_{int}(L) \right) + \\ \chi_{short,2}(L) \left(\left(\frac{T}{N_{NS}K_{TD}} \right)^2 + \left(\frac{AT}{N_{NS}K_{ATD}} \right)^2 + \frac{2(T)(AT)}{\left(N_{NS} \right)^2 K_{TD} K_{ATD}} \right) \omega_s \end{array} \right] / Z_{NRI \sim P, 2Tet}. \quad (S28)$$

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:

$$\frac{p_{r,2}(L)}{p_o} = \left[\begin{array}{l} 1 + \frac{\chi_{int}(L)}{\chi_{long}(L)} \left(\frac{2T}{N_{NS}K_{TD}} + \frac{2AT}{N_{NS}K_{ATD}} \right) + \\ \frac{\chi_{short,2}(L)}{\chi_{long}(L)} \left(\left(\frac{T}{N_{NS}K_{TD}} \right)^2 + \left(\frac{2AT}{N_{NS}K_{ATD}} \right)^2 + \frac{2(T)(AT)}{\left(N_{NS} \right)^2 K_{TD} K_{ATD}} \right) \omega_s \end{array} \right] / \frac{Z_{NRI \sim P, 2Tet}}{1 + \chi_{long}(L)}. \quad (S29)$$

iv. 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. The reason for introducing such a parameter is that there is no parameter set that is consistent with all of the data without including such an additional parameter. 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:

$$\begin{aligned}
Z_{NtrC,3Tet} = & 1 + \chi_0 + (1 + \chi_{int1}) \frac{1}{N_{NS}} \left(3 \frac{T}{K_{TD}} + 3 \frac{AT}{K_{ATD}} \right) + \\
& (1 + X_{int2}) \omega_s \frac{2}{(N_{NS})^2} \left(\left(\frac{T}{K_{TD}} \right)^2 + 2 \frac{T}{K_{TD}} \frac{AT}{K_{ATD}} + \left(\frac{AT}{K_{ATD}} \right)^2 \right) + \\
& (1 + \chi_{int2}) \omega_l \frac{1}{(N_{NS})^2} \left(\left(\frac{T}{K_{TD}} \right)^2 + 2 \frac{T}{K_{TD}} \frac{AT}{K_{ATD}} + \left(\frac{AT}{K_{ATD}} \right)^2 \right) + \\
& (1 + \chi_{short,3}) \omega_s^2 \omega_l \frac{1}{(N_{NS})^3} \left(\left(\frac{T}{K_{TD}} \right)^3 + 3 \frac{AT}{K_{ATD}} \left(\frac{T}{K_{TD}} \right)^2 + 3 \left(\frac{AT}{K_{ATD}} \right)^2 \frac{T}{K_{TD}} + \left(\frac{AT}{K_{ATD}} \right)^3 \right).
\end{aligned} \tag{S30}$$

Here we have omitted the explicit dependence of the different $\chi(L)$ on length.

Using (S30) we can compute the ratio $p_{r,3}(L)/p_o$ as for the 1 and 2-tet cases respectively.

v. Estimating the looping-capacity values

One of the major shortcomings of the model presented above is the proliferation of “free” looping capacity parameters with increasing number of binding sites on the cassette. A minimalistic approach to address this issue permitting us to reduce the model to a tractable 3-4 fitting parameters (i.e. binding constants and anti-cooperativity values) is based upon the observation that in the limit of no aTc and assuming saturating concentrations of TetR the expressions for repression (e.g. eqn. S29) reduce to:

$$r_n(L) \equiv \lim_{[A] \rightarrow 0} \frac{p_{r,n}(L)}{p_o(L)} = \frac{\chi_{short,n}(L)}{1 + \chi_{short,1}(L)} \frac{1 + \chi_{long}(L)}{\chi_{long}(L)}, \tag{S31}$$

where n is a subscript that signifies the number of TetR binding sites on the cassette. This allows us to obtain the different $\chi_{short,n}(L)$ as a function of $\chi_{long}(L)$ provided that we have either a measurement or theoretical expression for $r_n(L)$. While a theoretical expression is outside the scope of this paper, the value for $r_1(L)$, for instance, can be extrapolated from figure 2B for any length (as indicated by dashed line on the figure), or obtained directly from measurements such as the ones shown in Fig. 3B and 3D for L=115 and L=150 bp respectively.

vi. Fitting the Bimodal Repression Data

Given the long-length limit derived in the Box 2, and the strong repression observation stemming from increased rigidity of the nucleo-protein complex for short lengths, 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:

$$\chi_{short1}(L) = \gamma_{s1} \chi_{long}(L - L_{s1}) + \varphi_1, \quad (\text{S32})$$

where L_{s1} is a length scale that corresponds approximately to the TetR footprint plus the unprotected DNA between the TetR and the driver complex ($\sim 45\text{-}50$ 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:

$$\gamma_{s1} \equiv \frac{\gamma_{short}}{\gamma_{long}} \quad (\text{S33})$$

in order to converge on the entropic limit result for repression shown in eqn. (5) of Box 2.

The qualitative rationalization for the length shift approximation is based on the assumption that in the short-range rigid looping regime the cassette sequesters the portion of the DNA it binds to and severely hinders it from bending (Fig. 2A light blue shades). As a result of this length sequestration, the looping segment is now effectively shortened thus shifting the looping capacity function $\chi(L)$ to the right leaving only the upstream portion of the DNA ($L-L_{s1}$) free to bend and form a loop. In Fig 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. (S32) into eqn. (S31), we generate the following expression which can be used for fitting the repression data for the 1-Tet case presented in Fig. 2A:

$$r_1(L) = \frac{\frac{\gamma_{s1}\chi_{long}(L - L_{s1}) + \varphi_1}{\chi_{long}(L)}}{1 + \frac{\gamma_{s1}\chi_{long}(L - L_{s1}) + \varphi_1}{1 + \chi_{long}(L)}}, \quad (\text{S34})$$

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:

$$r_2(L) = \frac{\frac{\gamma_{s2}\chi_{long}(L - L_{s2}) + \varphi_{s2}}{\chi_{long}(L)}}{\frac{1 + \gamma_{s2}\chi_{long}(L - L_{s2}) + \varphi_{s2}}{1 + \chi_{long}(L)}}, \quad (S35)$$

and

$$r_3(L) = \frac{\frac{\gamma_{s3}\chi_{long}(L - L_{s3}) + \varphi_{s3}}{\chi_{long}(L)}}{\frac{1 + \gamma_{s3}\chi_{long}(L - L_{s3}) + \varphi_{s3}}{1 + \chi_{long}(L)}}, \quad (S36)$$

where 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 Fig. S4C, we plot the results of eqn. (S34-S36) using the approximated repressed looping capacity functions plotted in Fig. S4B, and a clear bimodal behavior is obtained for all approximated functions.

We use eqn. (S34-S36) to fit the data in Fig. 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 ($L_{NRI} \sim 60$ bp – (Hervas et al., 2009)) to the contour length of the total bound enhancer (driver+TetR cassette), expressed as follows:

$$\gamma_{sn} \sim \frac{L_{NRI}}{L_{NRI} + L_{sn}}. \quad (S37)$$

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): $\gamma_{s1}=0.60\pm0.07$, $\gamma_{s2}=0.37\pm0.06$, $\gamma_{s3}=0.23\pm0.05$, 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:

$$\chi_{sn}(L) = \frac{L_{NRI}}{L_{NRI} + L_{sn}} \chi(L - L_{sn}). \quad (S38)$$

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

vii. Four levels of complexity to the multi-TetR binding site model and sample fits

Given our assumption that TetR can bind in two forms to its binding site (i.e. with zero or one aTc molecules bound), thus effectively defining it as two different DNA binding proteins, the expression for the anti-cooperativity parameters can take into account the different couplings that may be possible between the different species. Of course, accounting for all of these eventualities

leads to a large proliferation in the allowed parameters. It is possible to express these parameters as follows:

$$\overrightarrow{\omega}_s = (\omega_{tt}^s, \omega_{at}^s, \omega_{aa}^s) \quad (S39)$$

and

$$\overrightarrow{\omega}_l = (\omega_{tt}^l, \omega_{at}^l, \omega_{aa}^l) \quad (S40)$$

which we term the vector model to differentiate from the version developed above with a single value for ω_s and ω_l , now termed the scalar model. In this case $\omega_{tt}^s, \omega_{tt}^l, \omega_{at}^s, \omega_{at}^l, \omega_{aa}^s, \omega_{aa}^l$ correspond to the Tet-Tet, Tet-(atc-Tet), and (atc-Tet)-(atc-Tet) short and long range interactions respectively. Using the vector model for anti-cooperativity the partition function for the 1-Tet case does not change, but the 2-Tet function now becomes:

$$Z_{NRI \sim P, 2Tet} = 1 + \chi_{long}(L) + (1 + \chi_{int}(L)) \left(\frac{2T}{N_{NS} K_{TD}} + \frac{2AT}{N_{NS} K_{ATD}} \right) + \\ (1 + \chi_{short,2}(L)) \left(\omega_{tt}^s \left(\frac{T}{N_{NS} K_{TD}} \right)^2 + \omega_{aa}^s \left(\frac{AT}{N_{NS} K_{ATD}} \right)^2 + \omega_{at}^s \frac{2(T)(AT)}{(N_{NS})^2 K_{TD} K_{ATD}} \right). \quad (S41)$$

Likewise, the partition function of the 3-Tet case is generalized as follows:

$$\begin{aligned}
Z_{NtrC,3Tet} = & 1 + \chi_0 + (1 + \chi_{int1}) \frac{1}{N_{NS}} \left(3 \frac{T}{K_{TD}} + 3 \frac{AT}{K_{ATD}} \right) + \\
& (1 + \chi_{int2}) \frac{2}{(N_{NS})^2} \left(\omega_{tt}^s \left(\frac{T}{K_{TD}} \right)^2 + 2 \omega_{at}^s \frac{T}{K_{TD}} \frac{AT}{K_{ATD}} + \omega_{aa}^s \left(\frac{AT}{K_{ATD}} \right)^2 \right) + \\
& (1 + \chi_{int2}) \frac{1}{(N_{NS})^2} \left(\omega_{tt}^l \left(\frac{T}{K_{TD}} \right)^2 + 2 \omega_{at}^l \frac{T}{K_{TD}} \frac{AT}{K_{ATD}} + \omega_{aa}^l \left(\frac{AT}{K_{ATD}} \right)^2 \right) + \\
& (1 + \chi_{short,3}) \frac{1}{(N_{NS})^3} \left(\omega_{tt}^l (\omega_{tt}^s)^2 \left(\frac{T}{K_{TD}} \right)^3 + \left(2 \omega_{at}^l \omega_{at}^s \omega_{tt}^s + \omega_{tt}^l (\omega_{at}^s)^2 \right) \frac{AT}{K_{ATD}} \left(\frac{T}{K_{TD}} \right)^2 \right. \\
& \left. + \left(2 \omega_{at}^l \omega_{at}^s \omega_{aa}^s + \omega_{aa}^l (\omega_{at}^s)^2 \right) \left(\frac{AT}{K_{ATD}} \right)^2 \frac{T}{K_{TD}} + \omega_{aa}^l (\omega_{aa}^s)^2 \left(\frac{AT}{K_{ATD}} \right)^3 \right).
\end{aligned} \tag{S42}$$

Note that we assume for simplicity that the looping capacity parameters only reflect the number of bound proteins, and therefore are assumed to be the same for all the combinations of T and AT binding for particular occupancies.

In order to quantitatively analyze our fold-induction data (Fig.3 and 4) using the thermodynamic model, we may also need to further increase the complexity of the model by accounting for the fact that we use both TetO1 and TetO2 binding sites in all of our 2-Tet and 3-Tet cassettes. Thus, instead of having K_{TD} and K_{ATD} for binding constants, we now have: K_{TD01} , K_{TD02} , K_{ATD01} , and K_{ATD02} (i.e. two additional binding constants). Therefore, in order to analyze our data it is possible to consider four variations on the model in order of increasing complexity:

1. Scalar ω – single binding site
2. Scalar ω – different binding site (i.e. TetO1 and TetO2)
3. Vector ω – single binding site.

4. Vector ω – different binding site (i.e. Tet01 and Tet02)

Using these different levels of complexity (and their associated numbers of free parameters) we can generate many types of curves that recapitulate the qualitative as well as quantitative behavior of our data, as shown in Fig. 6 for Model 1. However, quantitatively analyzing the data using model fits requires us to simultaneously account for the steps and cooperative transitions in the three types of data sets, while using the same free parameter values (i.e. binding constants and anti-cooperativity parameters) in order to extract microscopic information on the binding of TetR to DNA *in vivo* and the associated anti-cooperative protein-protein interactions. While this analysis is outside of the scope of the present work, for completeness we show in Fig. S5 and Table S5 fits and associated parameter values for the 1-Tet, 2-Tet, and 3-Tet data sets shown in the text. These fits are computed using Model 4 with looping capacity values extracted from the bar graphs of fig. 3B and 3D, and come the closest to generating good quantitative fits with the same free-parameter values simultaneously as Table S5 shows.

Supplementary Figure 1: Plasmid Schematic, Effect of NRII2302 on strain Growth, and Single Cell Measurements Methodology – associated with 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 saturating 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 μ m. 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.

Supplementary Figure 2: Model for enhancer-based transcription – associated with Box 1

A). The panel shows several candidate looping capacity functions, which when plugged into eqn. 4 of box 1 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. 5 – Box 1 (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 Fig. 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 (Muller 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 $\frac{3}{4}$ power law decline, which denotes approximately the long length scale power law decline exhibited by the red curve. 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.

Supplementary Figure 3: Model for glnAp1 transcription – associated with Box 2.

Parameters and limits of the simple thermodynamic model. (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 glnAP1 promoter. Binding to the NRI#1,2 sites renders the σ^{70} promoter inactive. In the second case the levels of NRI~P are too low to repress the σ^{70} promoter. (B,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, E) The steady state probability of occupancy of NRI~P on the NR#1,2 sites is calculated from equation (S5) assuming $\beta = 0.0116 \text{ min}^{-1}$ for (D) $\gamma = 0$, and (E) $\gamma = 1$.

Supplementary Figure 4: Fitting the enhancer repression data – associated with 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 $(\text{NRI~P})^6$ driver and a “poised” σ^{54} -RNAP complex are always bound to the DNA. B) Approximated repressed looping capacity functions (see eqn. S32 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 $L_s=45, 87$, and 129 for the red, green, and purple curves respectively. We utilized the unshifted looping capacity function used to fit the data in Fig. 1B (see Box 1 for parameters) 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 (S34-S36) from the sample functions shown in Fig. S3B. The color of each repression curve is identical to the color of the corresponding repressed looping capacity function shown in Fig. S3B used for the computation.

Supplementary Figure 5: Model Fits to the Repression Ratio Data – associated with Figure 3-6

Fits to the repression ratio data using level 4 of the model (i.e vector model with two binding sites). In panels (A), (B), (C), we show fits to the 1-Tet data used in Fig 4A, 2-Tet data shown in Fig. 3A, and 3-Tet data shown in Fig. 3C respectively. In all cases the model qualitatively reproduces the major features observed in the data. We used a simple “digital” version of the vector model with $\vec{\omega}_s = (1,0,0)$ and $\vec{\omega}_l = (1,0.1,0.1)$ which implies that all states where aTc-Tet is bound next to some other form of TetR are excluded. This interpretation implies that the formation of steps in our system may be due to excluded or forbidden states whose absence allows the single occupancy configuration to become more prominent.

IV. Supplementary References

- Atkinson, M.R., Pattaramanon, N., and Ninfa, A.J. (2002). Governor of the glnAp2 promoter of Escherichia coli. *Molec Micro* **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.
- 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.
- Hervas, 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.
- 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., 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.
- 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.
- 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.
- 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 Cellular and Molecular Biology*, N.F.E. et. al., ed. (Washington DC, ASM Press), p. 1344.

Muller, J., Oehler, S., and Muller-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.

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 U S A* 80, 5554-5558.

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.

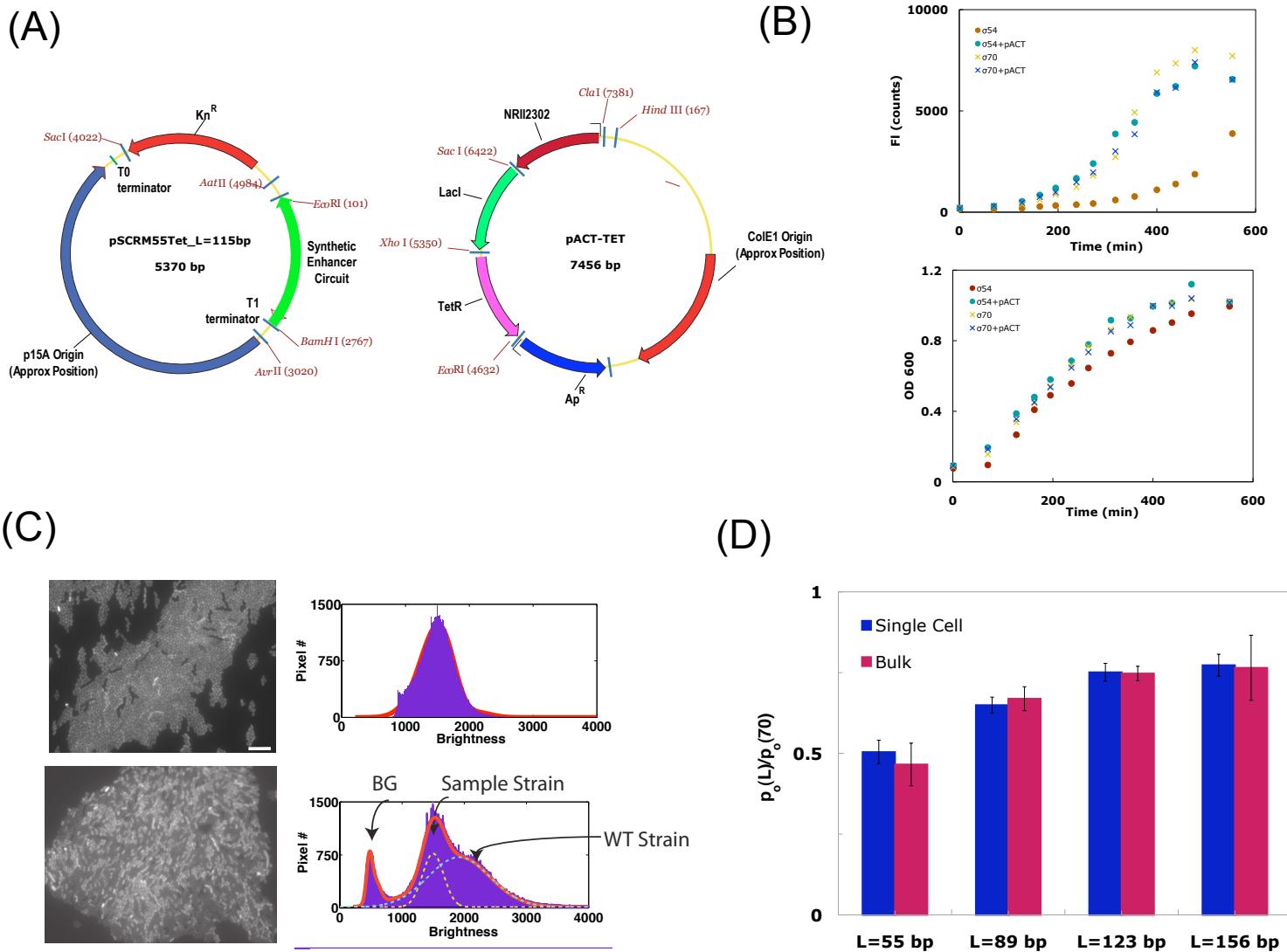


Figure S1 associated with Figure 1

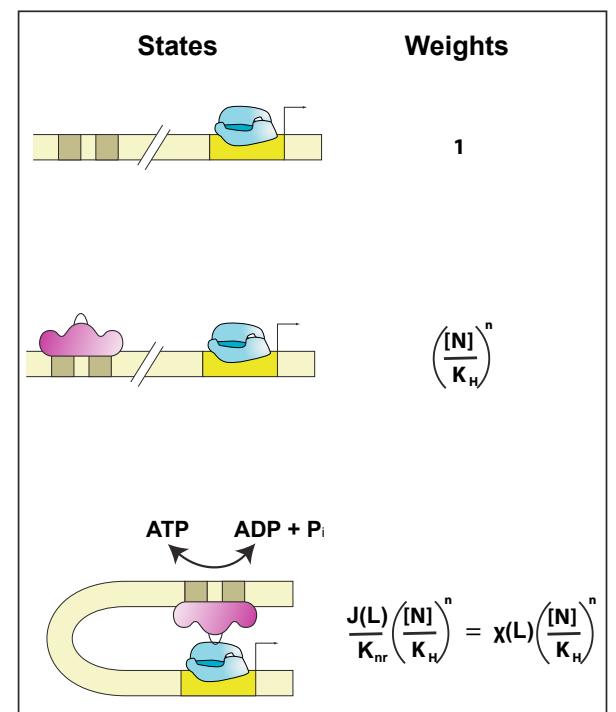
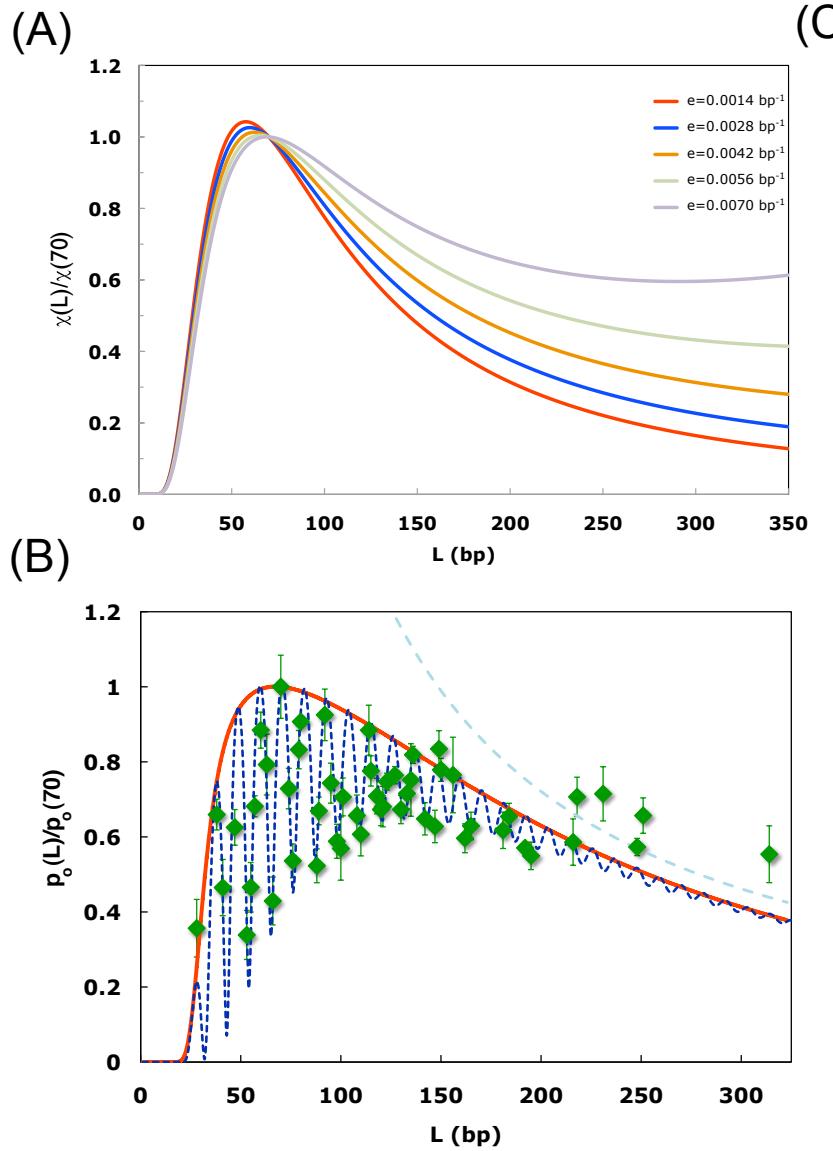


Figure S2 associated with Box 1

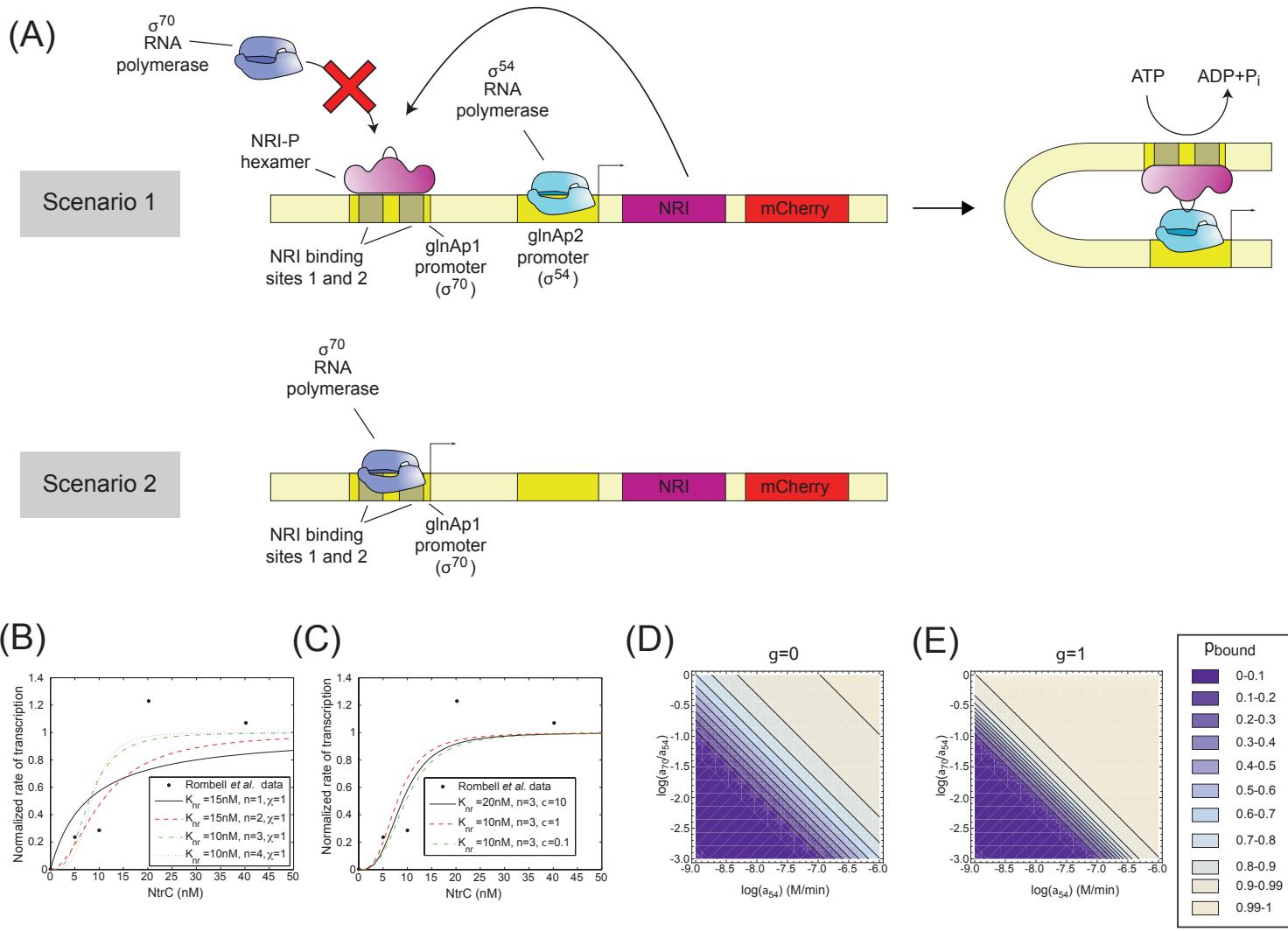


Figure S3 associated with Box 2.

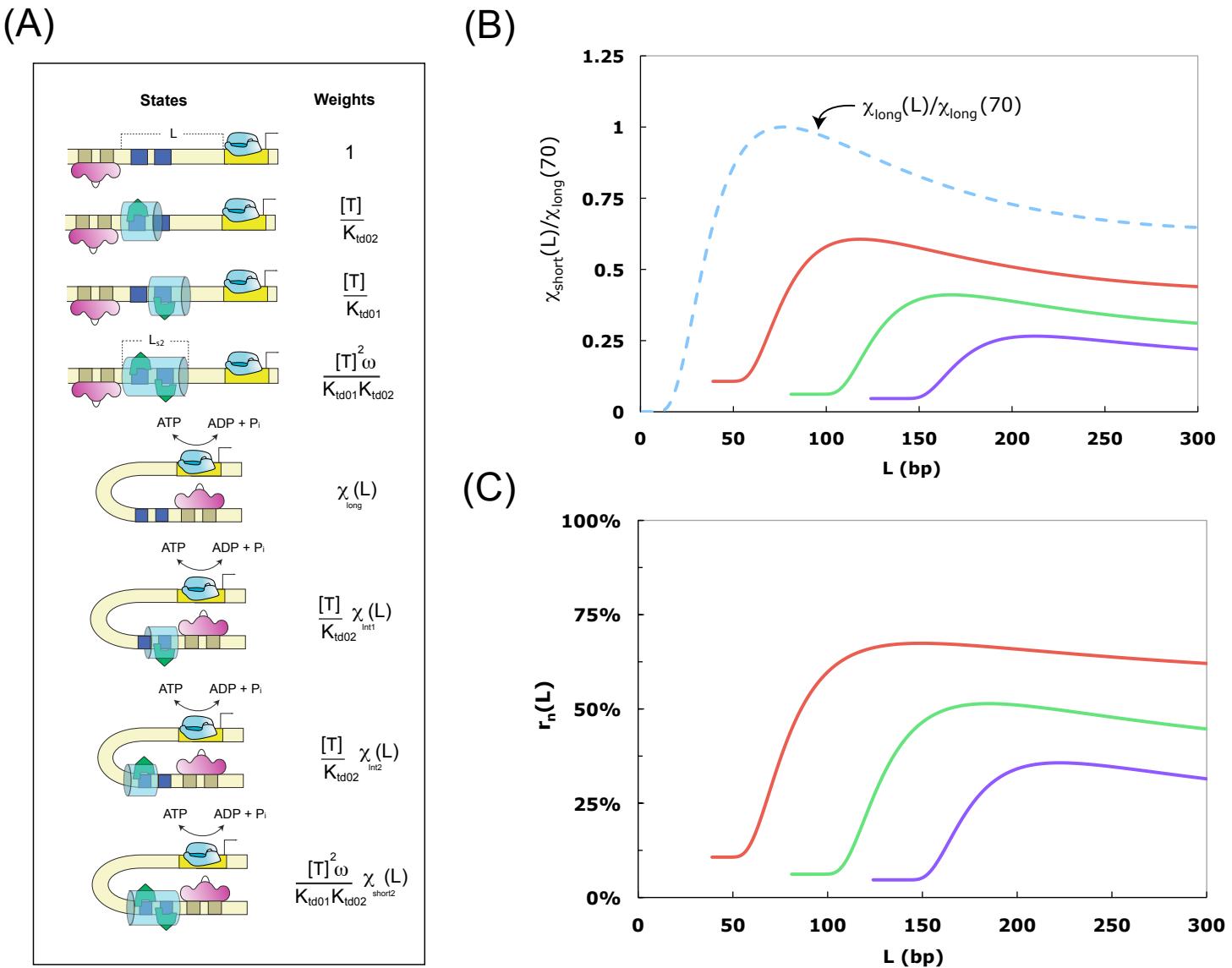


Figure S4 associated with Figure 2

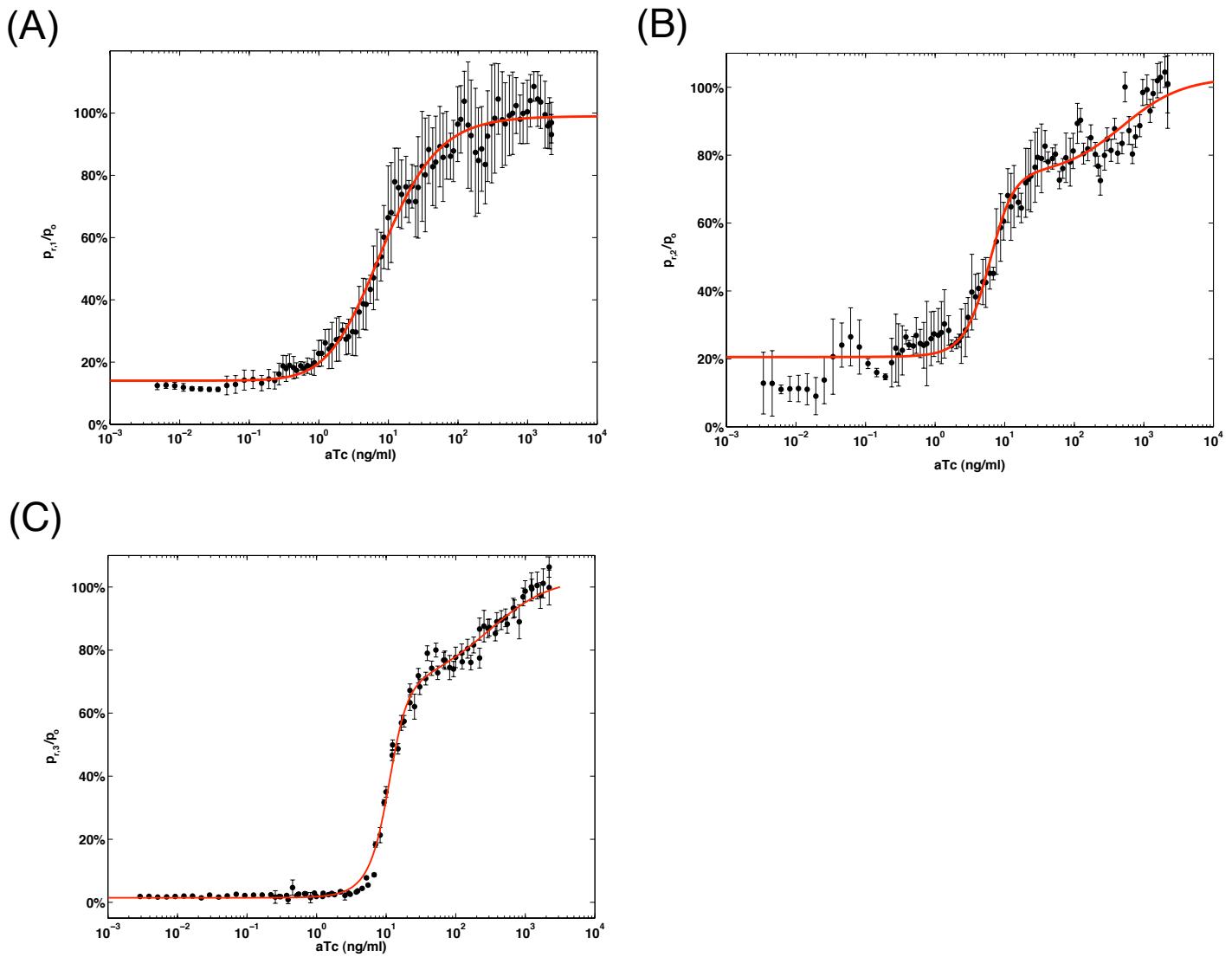


Figure S5 associated with Fig. 3-6

Cassette	Length	Sequence
Spacer19	19	CTAGCTCCGGCCAAGCTCG
Spacer92	92	CTAGAAATCTCCGGCGCTGCGCGGGATCGCCTGCACGGAGCTGGTCCGCCAGACTCGAGCTTGCGCCACCGTGAGGTCTGGCCACACGG GTCGGCGCTTCGGCAGCGTA CTAGGAGG
Spacer64	64	CTAGCTCCGGCGCTGCGCGGATCGCCTGCACGGAGCTGGTCCGCCAGACTCGAGCTTG
1-Tet	25	CTAGCTCCCTATCAGT GATA AGAGA G
2-Tet	60	CTAGCTCCCTATCAGT GATA AGAGATCCGGCCAGACTCGAG ACTCTATCATTGATA AGAG TG
3-Tet	95	CTAGCTCCCTATCAGT GATA AGAGATCCGGCCAGACTCGAG ACTCTATCATTGATA AGAG TGCGAGCCTCGACATCCC TATCAGT GATA AGAG A
3-Tet-S	125	CTAGAAATCCGGTCCGCTCCCTATCAGT GATA AGAGATCCGGCCAGACTCGAG ACTCTATCATTGATA AGAG TGCGAGCCTCGACATCCC TATCAGT GATA AGAG A
6-Tet	190	CTCCCTATCAGT GATA AGAGATCCGGCCAGACTCGAG ACTCTATCATTGATA AGAG TGCGAGCCTCGACATCCC TATCAGT GATA AGAG A GCTAG
2-Tet-10	60	CTAGCTCCCTATCAGT GATA AGAGACTCCGGCGCTGCGCGGATCGCCTGCACGGAGCG ACTCTATCATTGATA AGAG TG
2-Tet-01	60	CTAGCTCCCGCGCTGCGCGGATCGCCTGCACGGAGCG ACTCTATCATTGATA AGAG TG
3-Tet-110	95	CTAGCTCCCTATCAGT GATA AGAGATCCGGCCAGACTCGAG ACTCTATCATTGATA AGAG TGCTCCGCGCTGCGCGGATCGCTGACGGAGCG
3-Tet-101	95	CTAGCTCCCTATCAGT GATA AGAGACTCCGGCGCTGCGCGGATCGCCTGCACGGAGCG ACTCTATCATTGATA AGAG TGCGAGCCTCGACATCCC TATCAGT GATA AGAG A
3-Tet-011	95	CTAGCTCCCGCGCTGCGCGGATCGCCTGCACGGAGCG ACTCTATCATTGATA AGAG TGCGAGCCTCGACATCCC TATCAGT GATA AGAG A
3-Tet-100	95	CTAGCTCCCTATCAGT GATA AGAGATTGGTCCGCCAGACTCGAGCTTGCGCCACCGT CTCCGCGCTGCGCGGATCGCTGACGGAGCG
3-Tet-010	95	CTAGCTGGTCCGCCAGACTCGAGCTTGCGCCACCGT ACTCTATCATTGATA AGAG TGCTCCGCGCTGCGCGGATCGCTGACGGAGCG
3-Tet-001	95	CTAGCTCCCGCGCTGCGCGGATCGCCTGCACGGAGCGAGGTCTGGCCACACGGTCGGCGTTGCAGCA CTCCCTATCAGT GATA AGAG A

Table S1 - associated with Figure 1 – Synthetic Enhancer Cassettes.

Sequence

Table S2 – associated with Box 1 – Sequences used for Fig. 1C

L (bp)	Sequence (1 - TetR binding sites - Tet02)
150	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTTCCGACGCTAGCTCCCTATCAGTAGATAGATAGATGGTCCGGCAGACTCGAGCTTGGGCCACGGCTCGCCGCGATCGCCGCGCTCGCCGCGCTAGCTCCCTCCCATGATAACGCCCTTAAAGGGCAATTAAAAGTTGGCACAGATTTCGCTT
115	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTTCCGACGCTAGCTCCCTATCAGTAGATAGATAGAAGCTAGCAGCGCGCGAGATTCCGCTGCACGGAGCCTAGCGTCCCCATGATAACGCCCTTAAAGGGCAATTAAAAGTTGGCACAGATTTCGCTT
114	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTTCCGACGCTAGCTCCCTATCAGTAGATAGAAGCTAGCAGCGCGCGAGATTCCGCTGCACGGAGCCTAGCGTCCCCATGATAACGCCCTTAAAGGGCAATTAAAAGTTGGCACAGATTTCGCTT
95	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTAGCAGCGCGAGATTCCGCTGCACGGAGCCTAGCGTCCCCATGATAACGCCCTTAAAGGGCAATTAAAAGTTGGCACAGATTTCGCTT
66	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTAGCAGCGCGAGATTCCGCTGCACGGAGCCTAGCGTCCCCATGATAACGCCCTTAAAGGGCAATTAAAAGTTGGCACAGATTTCGCTT
63	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTAGCAGCGCGAGATTCCGCTGCACGGAGCCTAGCGTCCCCATGATAACGCCCTTAAAGGGCAATTAAAAGTTGGCACAGATTTCGCTT
53	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTAGCAGCGCGAGATTCCGCTGCACGGAGCCTAGCGTCCCCATGATAACGCCCTTAAAGGGCAATTAAAAGTTGGCACAGATTTCGCTT

L (bp)	Sequence (2 - TetR binding sites - Tet02/Tet01)
181	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTCTATCATTTGATAGAGTCTAGCGACGAGCTTGGGCCACGGATTTCGCTT
150	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTCTATCATTTGATAGAGTCTAGCGACGAGCTTGGGCCACGGATTTCGCTT
149	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTCTATCATTTGATAGAGTCTAGCGACGAGCTTGGGCCACGGATTTCGCTT
130	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTCTATCATTTGATAGAGTCTAGCGACGAGCTTGGGCCACGGATTTCGCTT
115	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTCTATCATTTGATAGAGTCTAGCGACGAGCTTGGGCCACGGATTTCGCTT
101	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTCTATCATTTGATAGAGTCTAGCGACGAGCTTGGGCCACGGATTTCGCTT
98	GCTCACAAATTGACCAACATGGCTTAAATTGTCATTGAAGCACAATTATTGGTCAAACATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTCTATCATTTGATAGAGTCTAGCGACGAGCTTGGGCCACGGATTTCGCTT

L (bp)	Sequence (3 - TetR binding sites - Tet02/Tet01/Tet02)
252	GCTCACARTGGACCAACATGGCTTAAATTGTCATTGAAGCACAATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGACGAGCTGGCGCGAGCTTGCGCTTGGGCCACGGATTTCGCTT
216	GCTCACARTGGACCAACATGGCTTAAATTGTCATTGAAGCACAATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGACGAGCTGGCGCGAGCTTGCGCTTGGGCCACGGATTTCGCTT
184	GCTCACARTGGACCAACATGGCTTAAATTGTCATTGAAGCACAATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGACGAGCTCCGGCCAGACTCGAGCTTGCGCTTGGGCCACGGATTTCGCTT
165	GCTCACARTGGACCAACATGGCTTAAATTGTCATTGAAGCACAATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGACGAGCTCCGGCCAGACTCGAGCTTGCGCTTGGGCCACGGATTTCGCTT
150	GCTCACARTGGACCAACATGGCTTAAATTGTCATTGAAGCACAATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGACGAGCTCCGGCCAGACTCGAGCTTGCGCTTGGGCCACGGATTTCGCTT
136	GCTCACARTGGACCAACATGGCTTAAATTGTCATTGAAGCACAATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGACGAGCTCCGGCCAGACTCGAGCTTGCGCTTGGGCCACGGATTTCGCTT
133	GCTCACARTGGACCAACATGGCTTAAATTGTCATTGAAGCACAATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGACGAGCTCCGGCCAGACTCGAGCTTGCGCTTGGGCCACGGATTTCGCTT
123	GCTCACARTGGACCAACATGGCTTAAATTGTCATTGAAGCACAATTGTCCTCGACGCTAGCTCCCTATCAGTAGATAGAAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGAGCTCCGGCCAGACTCGAGCTACTCATTTGATAGAGTCTAGCGACGAGCTCCGGCCAGACTCGAGCTTGCGCTTGGGCCACGGATTTCGCTT

Table S3 – Associated with Figure 2 - Sequences used for Fig. 2B

Parameter	Value	Reference
K_{TD}^{NS}	1 mM bp	(Kleinschmidt et al., 1988)
K_{TO1}^S	10 pM	(Lederer et al., 1995)
K_{ATO1}^S	1 nM – 10 nM	(Lederer et al., 1995)
K_{ATAO1}^S	3 μM – 1mM	(Lederer et al., 1995)

Table S4 – Associated with Fig. 5 - Biochemical parameters from the Literature

Parameter	Value 1-Tet	Value 2-Tet	Value 3-Tet
T_0	351 molecules	351 molecules	351 molecules
K_{at}	100 pM	360 pM	360 pM
K_{TD01}	100 pM	100 pM	100 pM
K_{TD02}	15 pM	15 pM	15 pM
K_{ATD01}	0.7 nM	0.7 nM	0.7 nM
K_{ATD02}	7.5 nM	7.5 nM	7.5 nM
$\overline{\omega}_s = (\omega_{tt}^s, \omega_{at}^s, \omega_{aa}^s)$	NA	(1,0,0)	NA
$\overrightarrow{\omega}_l = (\omega_{tt}^l, \omega_{at}^l, \omega_{aa}^l)$	NA	(1,0,0)	(1,0.1,0.1)

Table S5 – Associated with Figure 6 – Parameters used for Fits with model level 4