Building Enhancers From the Ground Up: A Synthetic Biology Approach

Roee Amit[1,3], Hernan G. Garcia[2], Rob Phillips[3], and Scott E. Fraser[1,3]

[1] Division of Biology, California Institute of Technology, Pasadena, CA 91125

[2] Department of Physics, California Institute of Technology, Pasadena, CA 91125

[3] Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125

Keywords: transcriptional regulation, synthetic biology, DNA looping, NtrC

SUMMARY

The study of transcription and its regulation have been cornerstones of biological research for decades. Recent efforts in both synthetic and systems biology have centered on quantitative treatments of the information processing capabilities of gene regulatory networks. One of the promises of the synthetic biology approach is that it concretely tests our understanding by forcing us to use what we think we understand to synthesize biological functions. In this paper, we exploit such an approach to study synthetic enhancers. To make it possible to systematically quantify the properties of the synthetic enhancers, we based our system on a bacterial circuit consisting of a minimally active promoter for the poised σ^{54} sigma factor, and an associated upstream minimal enhancer. This allowed us to construct a number of non-coding DNA cassettes that altered the enhancer region by changing its length and binding site content, and to probe their effect on reporter expression. Our findings suggest that enhancer-based transcriptional control depends critically and quantitatively on DNA looping, which leads to complex regulatory effects when the enhancer cassettes contain additional transcription factor binding sites. Repression and activation are achieved by altering the equilibrium thermodynamics of looping, which can lead to a bimodal output with a transition between states that is a function of the length of the regulatory cassette. In turn, looping-based regulation enables the enhancers to convert a variable ligand concentration input (e.g. of an inducer) into discrete or step-like output expression levels. As a result, the synthetic enhancers exhibit the

kind of computational and regulatory characteristics that are observed for the more-complex metazoan enhancers.

Introduction

The ability of living organisms to convert variable input signals into discrete output levels that in turn trigger cell differentiation and morphogenesis is one of the great mysteries of modern biology. One way this is achieved is through rich regulatory networks involving multiple “inputs”, almost all of which involve “action at a distance” through enhancers that can be widely separated from the promoter itself (Bolouri and Davidson, 2002; Davidson, 2006; Small et al., 1992; Stathopoulos and Levine, 2005). A large body of work has dissected a number of developmentally important genes, and has revealed a modular organization, which includes a basal promoter, to which the RNAP complex binds, and a set of enhancers or *cis* regulatory modules that are positioned nearby (located upstream or in introns) (Korzh, 2007). On its own, a basal promoter has the capability to execute little or no transcriptional regulation, but together with enhancers they can express their full regulatory potential (Davidson, 2001).

While the developmental gene regulatory subnetwork structure (Bolouri and Davidson, 2002; Stathopoulos and Levine, 2005) and the input/output relationship between different genes is becoming better defined, the underlying mechanisms of regulatory action at a distance responsible for integrating the various inputs in producing complex responses remain poorly understood. In the very few examples that have been quantitatively characterized thus far (Arnosti et al., 1996; Davidson, 2006; Gregor et al., 2007; Small et al., 1992; Yuh et al., 2001) regulatory regions have displayed output characteristics that can be interpreted to be a result of complex computational operations carried out on the

enhancer using the various input signals. For instance, the types of computations inferred from *cis* regulatory analysis carried out on the *endo16* (Yuh et al., 2001) and *cyIIIa* (Davidson, 2006) genes in the sea urchin *S.purpuratus* include arithmetic (i.e. addition, subtraction, etc.), logical (i.e. AND, OR, etc.), and analog operations (i.e discrete amplification, etc.).

To explore the kinds of mechanisms that can yield the complex regulatory and computational behavior associated with enhancers, we opted to construct synthetic enhancers *de novo*. This approach permits us to systematically construct enhancers in a modular fashion, starting with a minimal enhancer and progressively increasing the synthetic enhancer's complexity with the addition of discrete sets of defined binding sites for the well-known transcription factor TetR. The synthetic approach provides us with an experimental foundation that can be utilized to develop a cascade of nested 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 allow us to 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 increasingly complex cascade. At each experimental stage, where an increasingly rich 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

We selected the bacterial σ^{54} system to test our methodology because transcription through this sigma factor in bacteria has the hallmark characteristics of a minimal enhancer architecture: a minimal promoter with little intrinsic transcriptional capability (Atkinson et al., 2003), and an associated enhancer region that binds an activator required for “driving” the transcriptional process (Rappas et al., 2007). The enhancer is located upstream, widely separated from the promoter (100-1000 bp (Ninfa et al., 1987)), and thus transcription-initiation requires some sort of action-at-a-distance. In the absence of the ”driver”, the polymerase remains poised at the promoter, ready but unable to initiate transcription.

For the experiments detailed below, enhancer-initiated transcription was facilitated by the phosphorylated NRI protein (NRI~P) (Magasanik, 1993). The latter was transcribed from the *NtrC* or *GlnG* gene coded for by an auto-activating circuit (Atkinson et al., 2003), and when bound to the DNA cooperatively assembled a hexameric complex, which functioned as the transcriptional driver in our system. An mCherry reporter was used to measure the transcriptional activity of this promoter (for circuit detail see Experimental Procedures).

RESULTS

Expression levels are controlled by DNA looping

Since the driver that binds our synthetic enhancer is located upstream of the promoter, and there must be direct physical contact between it and the "poised" polymerase, DNA looping has been implicated (Schulz et al., 2000; Su et al., 1990) in the initiation of transcription (Fig. 1C). Therefore, we reasoned that systematically varying the length of the DNA sequence between the driver binding sites and σ^{54} promoter (Fig. 1A) 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 (Becker et al., 2005; Law et al., 1993; Lee and Schleif, 1989; Lewis and Adhya, 2002; Muller 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 S2) of variable length (28-315 bp - Fig. 1A and S1A). We carried out fluorescence measurements in bulk, while the strains were growing in mid-log 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 Fig. 1B seem to exhibit a strongly fluctuating behavior with a non-trivial dependence on looping length (*L*). However, a useful framework for considering this complex data is provided by the thermodynamic model schematized in Fig. 1C. The essence of the model schematized in that 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 experimental data can be described using 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 (RNAP). In order 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, inspired by the underlying DNA biophysics and transcriptional mechanics (see Box 1), 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. The value of the helical period is in good agreement with previous measurements (Becker et al., 2005; Law et al., 1993; Lee and Schleif, 1989; Muller et al., 1996). It is worth noting that while 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 (Fig. S2A-B), various candidates for the looping capacity function can generate plausible envelope functions for the data as shown by the red-line in Fig. 1B and S2B. In particular, we were able to fit our data adequately with the exact same looping capacity functional form used to fit (Bintu et al., 2005) the data obtained for LacI looping by (Muller et al., 1996), only in this case at *L=70* the probability to be in a looped state is roughly equal to the probability to be in the unlooped state. Discriminating between alternative looping capacity functions would require data from larger loop lengths than those observed here.

These measurements show a systematic relation between loop length and level of expression in the NtrC system *in vivo*. Not only is it interesting to observe that the features seen here are similar quantitatively to those found in other familiar bacterial examples such as the *lac* operon, but these results provide a hint 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 factor. As such, these results serve as the jumping off point for the construction of the synthetic enhancers, which are the central result of this work.

Enhancer repression is a bimodal function of spacer length

Given that the level of transcription depends critically upon DNA looping, we hypothesized that it is possible to generate regulatory action-at-a-distance by altering the looping capacity function via the binding of transcription factors within the DNA loop. 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. 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 NRI#2 binding site (Fig. 2A) to ensure that the first TetR (Hillen et al., 1984) binding site does not interfere with the binding

of the NRI~P complex (Hervas 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 Box 2. The extent of repression for each cassette was quantified by measuring the fluorescence of the reporter both in the presence of TetR proteins, and in their absence. In Fig. 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 Fig. 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 lengths L_t, which serve as a DNA length scale setting a sharp transition between the two repression regimes, varies for each cassette type and are labeled as L_{t1}, L_{t2} and L_{t3}, 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, TetR is able to repress expression dramatically, and at longer lengths ($L>L_{t1}$, L_{t2}, and L_{t3}) expression is less repressed. One simple interpretation of this length dependence is that the DNA-TetR complex behaves like a “rigid” nucleoprotein complex with an effective persistence length longer than that of bare DNA. Thus, for $L<L_t$ looping is far less likely to take place and the RNAP will remain poised.

For long loop lengths, where the weak repression regime is observed, the rigidification effect observed for lower lengths is diminished. In this regime, the data indicates that repression levels are weakly dependent on the loop length and the synthetic 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 dashed lines that demarks each of colored data sets representing the repression functions $r_1(L)$, $r_2(L)$, and $r_3(L)$ (see SI for the functional form of these terms) on the graph. Finally, at very long loop lengths these repression values (as indicated by the dashed lines) seem to converge on a particular constant for each cassette configuration, rather than approach the non-repressed value of 100%.

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 Fig. 1 and Box 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 (see Box 2). As shown in Fig. 2C for the 1-Tet case, all of the states come in two broad categories: unlooped and inactive, and looped and active. 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 Box 2, we can derive an expression for long-distance repression that is a ratio of the repressed to the unrepressed looping capacity functions (eqn. (5)), which converges to a fixed value and gives a sense of the theoretical underpinnings for r(L).

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 (see SI and fits in Fig. 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 which appear to be correlated to the number of TetR binding sites and the distance between the beginning of the NRI#1 site through the last TetR binding site.

Multiple TetR binding sites generate step functions from a variable input

The long-range regulatory capability of our synthetic enhancer system discussed above has further regulatory potential. Our measurements reveal that in the long looping length regime, each cassette yields a repression state

characterized by weak repression levels (Fig. 2B). 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 (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 (Fig. 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 Fig. 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 containing a single binding site for TetR which we hypothesize correspond to the two partial occupancy configurations for the two-binding site architecture, and with identical

length to the 2-Tet cassette (Fig. 3B). Examination of Fig. 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 Fig. 3A, and with the weak repression regime for the 1-Tet cassette repression data (Fig. 2B).

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 Fig. 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 (Fig. 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 Fig. 3C reflect these partial occupancy states, we measured the repression values for six additional cassettes that account for all possible occupancy configurations (Fig. 3D). We found that only the triply occupied configuration is strongly repressed, while 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 – Fig. 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 (Fig. 3F) does not exhibit an increase in the number of intermediates, but instead is characterized by two intermediates with more evenly spaced repression-ratio values, with sharper transitions that produce a more distinct step-like function than for the 2- and 3-Tet cassettes (see also Fig. 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.

Synthetic Enhancer-type Computation and Code

Examining the data for the 1,2,3 and 6- binding site cassettes closer we find additional regulatory features, which likely would not have been guessed *a priori*. The dose response for each TetR cassette type indicates that the transition (Fig. 4A-D) between the low repressed state and the first intermediate are characterized by a Hill coefficient greater than one, interestingly, roughly

equal to the number of TetR binding sites. This result seems to imply that the regulatory function reflects an effective interaction in the factors that bind to the cassette, which can be interpreted as a type of computation.

To further examine the mechanistic underpinnings of our measurements, we examined the output function of the cassettes while altering their position within the spacer region. Fig. 4E shows that for the 3-Tet binding site cassette, the output function keeps its elementary characteristics (i.e. strongly repressed state, and a transition to one or two weakly repressed or unrepressed states) regardless of where the cassette is positioned within the spacer region. The results shown in Fig. 4E, and the different response functions for the 1,2, and 6-Tet cassettes (see Fig. 4A, 3A, and 3E respectively) suggest that each cassette type encodes an output function, whose characteristic dose response output is "encoded" or programmed by the geometry and binding site arrangement of the various TetR binding cassettes.

Modeling the regulatory code output functions

Given the modeling framework discussed in Box 1 and 2, which were used to model the looping and the bi-modal repression data, is it possible to generalize this scheme to reproduce the output functions shown in Fig. 3 and 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 (Hillen and Berens, 1994; Lederer et al., 1995) that up to two aTc ligand molecules can bind the TetR

dimers, and that TetR can bind its DNA binding site in two forms: unoccupied and occupied by a single aTc ligand, but with different Kd's (see Table S4). These assumptions are based on crystal structure analysis (Orth et al., 2000) and *in vitro* binding experiments (Lederer et al., 1996; Lederer et al., 1995). 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 Fig. 5A and SI) which generate mathematical expressions (eqn. S14-S16) 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 install those results into our states and weight schemes for the various cassette occupancies, which in turn allowed us to formulate a model for the repression-ratio data (see SI for details, and Fig. 5B-C for generalized model schematics) which accounts not only 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.

First, we consider a model for the 1-Tet cassette. Fig. 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 (Fig. S5A), as does the empirical fit given by a Hill function with Hill coefficient one as shown in Fig. 4A.

In order to extend the model to the 2-Tet case (see SI for detailed derivation and model fits), we incorporate an additional parameter (ω_s) that accounts for the free energy cost of 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 anti-cooperative behavior, which leads to increased stability for the single-occupancy configurations as compared with the double occupancy one. On the other hand, if (ω_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 (not shown).

In Figure 6 (B-C) we plot the individual probabilities (eqn. S29) for the cassette sub-occupancies 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 (ω_s)~1 (Fig. 6B) it overlaps significantly with the other two probabilities, leading to a relatively small over-all contribution from the single occupancy configurations, which results in an output function that lacks an intermediate step

(Fig. 6B – black line). However, for values of (ω_s) that promote anti-cooperativity in the protein-protein interaction, the overlap of the probabilities is significantly reduced (Fig. 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 (Fig. 6D-F and S4C 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 ~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 (Fig. 6E) that looks similar to the data in Fig. 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 (Fig. 6F), one has to introduce a second weaker anti-cooperativity term (ω_l) for the next to nearest neighbor interaction. As a result, we conclude that the existence of anti-cooperativity interaction parameters seem 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 both of the DNA mechanics and the ways in which the repressors interact both with each other and their DNA substrate.

Discussion

We explored transcriptional and regulatory characteristics of an enhancer-based transcriptional system by constructing increasingly complex enhancer elements from the ground up with small DNA cassettes, and subsequently characterizing the output expression function using a set of thermodynamics models. Our results show that expression in our system is characterized by a looping capacity function ($\chi(L)$), an extension of a model (Box 1) that is typically used to describe DNA cyclization (J-factor). According to our model and experimental data, the dependence of transcription on the looping capacity leads to complex regulatory behavior when additional transcription factor binding sites are present.

Unlike the conventional model for repression, where a repressor inhibits transcription by competing for the RNAP binding site, or by interfering with RNAP initiation (Alberts et al., 2002), our synthetic enhancers exhibit repression by a modification of the looping capacity function. This leads to a repression function 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. 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

(Box 2 and Fig. 2B). Furthermore, in the very long length limit, our data indicate that the level of repression trends towards a constant value. Thus, these long-distance repression effects together with our thermodynamic model (Box 2) suggest that when the distance separating the enhancer and promoter is large *($L>>L_t$)*, regulatory action-at-a-distance depends only on local enhancer characteristics and is independent of the looping length. Interestingly, some enhancer elements in eukaryotes have been experimentally defined based on the observation that their regulatory output was conserved with respect to the candidate enhancer's position and orientation relative to the promoter (Driever and Nusslein-Volhard, 1989; Driever et al., 1989). Therefore, it is tempting to speculate that the regulatory action-at-a-distance scenario elucidated here may be applicable to eukaryotic enhancers as well.

The most striking transcriptional 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, we showed that the intermediate steps that form in the response for the 2-Tet and 3-Tet cases can be explained by repression levels of preferred cassette occupancy states. The preferred states, in turn, are determined by an anti-cooperativity parameter, which is used to model a destabilizing interaction between two TetR proteins that are too closely bound to one another.

Extending the model further to the 6-Tet case is difficult due to the many possible enhancer occupancies and associated repression levels $r_n(L)$ that have not been accessed directly by the experiment. Given this limitation, do the

models for the 1, 2-Tet and 3-Tet cassettes provide an insight as to why the 6-Tet cassettes exhibit only a four-state regulatory function? The models indicate that the most important parameters in tuning the repression-ratio response to produce step functions are the anti-cooperativity interaction parameters: ω_s and ω_l. In Fig. 6D-E, we show that even though it is possible to model our 3-Tet data adequately with just a nearest-neighbor interaction, none of those curves exhibits two intermediate states regardless of the value of ω_s. In Fig. 6F, we show that in order to obtain two intermediates, we needed to introduce an additional next-to nearest neighbor interaction ω_l. Thus, we conclude the characteristics of the 6-Tet repression-ratio data (Fig. 3E) can be understood in terms of our model's general framework, provided we introduce appropriate nearest (ω_s) and next nearest (ω_l) neighbor interactions. However, to really test this understanding requires a strategy of systematic experimental characterization in which the number of binding sites, their affinities and relative positions are all controlled.

Regulatory Computation and Code

The significance of the discretization effect observed is further enhanced by that the observation that a form of "computation" may take place at the synthetic enhancer. Traditionally (Alberts et al., 2002), regulation has been used to explain the phenomenon of gene switching from "on" to "off" and vice versa, which motivated the binary or logical description of regulatory computation in models that were used to describe simple synthetic gene regulatory networks (Basu et al., 2005; Elowitz and S., 2000; Gardner et al., 2000; Hasty et al., 2002).

But what about cases like that described here where there are apparently more than two discrete regulatory states that can be accessed?

What type of computation do we observe in our data? Since the output functions are non-binary, we conclude that the type of computation that we seem to be observing is not well described via standard two-state logic gates or binary operations. Rather, these networks seem to exhibit a type of analog computation. These characteristics can be identified in two features exhibited by the data: the three or four discrete expression levels in the output step functions that signify particular cassette occupancy configurations, and in the ability to "count" or "add" the number of bound TetR proteins in order to transition to the strong repression regime.

We view our work as providing new insights into the nature of the regulatory code. In other contexts, it has been suggested that conserved regulatory regions between distantly related genomes encode for a particular regulatory outcome (Davidson, 2006), which in principle can be read directly off genomic sequences provided that we have the key. The complex regulatory functions studied in this work are completely synthetic and were constructed from a standard σ^{70} repressor (TetR). Their characteristics as specific and independent DNA-binders were co-opted in this case to engineer new and complex regulatory functions with a distinct signature. Since there seems to be a stereotyped relationship between a particular cassette design and its output function, the output functions obtained here can be interpreted to be "encoded" within the binding site cassettes.

Experimental Procedures

Synthetic Enhancer Cassette Design

See supplemental experimental procedures

Strain Construction

See supplemental experimental procedures

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 Peng, 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 Supplementary Information). 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 Fig. 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.

Looping Length Dependence Assay

Bulk expression levels of strains with varying spacer length were carried out as follows. 20 mL of fresh LB with appropriate antibiotics were inoculated in 125 ml flasks with overnight starters of appropriate synthetic enhancer strains. Cultures were then grown, and fluorescence measurements were taken at 30 minute intervals for roughly 5 hours to cover the mid log growth range. For each measurement, 200 µl of culture was dispensed in each of 4 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 auto-fluorescence 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. Synthetic enhancer plasmids were transformed with either pACT-Tet (Fig. S1A) or pACT-Tra plasmids in 3.300LG cells (where the *traR* gene replaces the *tetR* gene), to construct synthetic enhancer strains with and without endogenous TetR proteins.

To carry out the assay, synthetic enhancer strains were grown in fresh LB with appropriate antibiotics (Kan/Amp) to mid log range as measured by a spectrophotometer (Pharmacia Biotech) OD600 of ~0.6, and resuspended in low growth/low auto-fluorescence 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 – 10x 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 culure 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 Tecan 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 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

Bulk assays were carried out as follows. Synthetic enhancers strains containing the pACT-Tet plasmid were grown as described above. In this case, appropriate concentrations of aTc were dispensed in each well of a 48-well plate, spanning 4-6 orders of magnitude. For each strain, we used two 48-well plates to allow for 94 different readings of fluorescence as a function of aTc concentration (2 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 concentration, 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, where the averaging is carried over adjacent aTc concentrations. This algorithm is used to smooth out short-range fluctuations, and highlights the large-scale features that span wide concentration ranges.

Supplemental Information

Supplemental Information includes five figures, five tables, Supplemental Experimental Procedures, Importance of NRI#3,4,5 discussion, and a detailed derivation of our cascade of models.

Acknowledgements

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. RA was supported by a NIH Ruth L. Kirschstein fellowship, a Caltech CBCD grant, and the NIH through award NIH ARRA R01 GM085286-01S. RP and HG gratefully acknowledge awards NIH ARRA R01 GM085286-01S, R01 GM085286, and the NIH Director's PIONEER Award DP1 OD000217.

Figure Captions

Figure 1: Enhancer activation depend strongly on looping

(A) Schematic for synthetic enhancer circuit. For a detailed description of circuit construction and operation see Experimental Procedures. 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 mutant NRII2302. The synthetic enhancer circuit was transformed into a *ΔGlnL:ΔGlnG:3.300 E. coli* strain (3.300LG (Atkinson et al., 2003)) on a low-copy plasmid (~10/cell). (B) Relative fluorescence level $p_o(L)/p_o(70)$ vs. 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 wild type looping length). The fits correspond to our expression model with (blue dashed line) and without (red line) the periodic modulation (see Box 1 for more details). To confirm the results from the bulk measurements, we carried out a similar measurement on a subset of strains at the single cell level and found no detectable difference from the results presented in panel B (See Fig. S1C-D). (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.

Box 1 – 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. 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 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 protein-protein interactions that are mediated by DNA looping for *in vitro* and *in vivo* repression measurements (Bintu et al., 2005; 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

$$\frac{d[mRNA]}{dt} = \alpha P_{loop}(L) - \beta[mRNA], \tag{1}$$

where 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:

$$P_{loop} \cong \frac{\frac{J(L)}{K_{nr}}}{1+\frac{J(L)}{K_{nr}}}, \tag{2}$$

To develop intuition for the significance of the $J(L)/K_{nr}$ ratio and simplify the discussion below, we define the "looping capacity" as,

$$\chi(L) \equiv \frac{J(L)}{K_{nr}}. \tag{3}$$

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

$$[P] \propto [mRNA] \cong \frac{\alpha}{\beta}\left(\frac{\chi(L)}{1+\chi(L)}\right) \tag{4}$$

where *[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., 2005). In particular, 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 Fig. 1C using eqn. (4) and the expression for the looping J-factor used in (Bintu et al., 2005), modulated by a term that accounts for the torsional flexibility (Becker et al., 2005; Bloomfield et al., 1999)

$$\chi\left(L\right)=\exp\left(-\frac{d}{L}+eL-b\log(L)+f+\log\left(\sum_{n}\exp\left(-\frac{4\pi^{2}C_{twist}}{2LP^{2}}(L-L_{opt}+nP)^{2}\right)\right)\right), \qquad (5)$$

where *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 Fig. 1 and S2.

Figure 2: Bimodal repression of enhancer-based transcription

(A) Schematic showing the constructs used to study repression in the context of enhancer-based activation 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 t_{s3}) is hypothesized to be 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), r_3(L) levels correspond to the repression functions as defined in Box 2 – eqn (4) and eqn. S31 in the SI. The values of these functions at particular lengths are used as input to the model and fits (see. Fig. 3,4,6, and S4). The colored curved double lines for each data set correspond to model fits (see Fig. S4 for additional detail), and the dashed lined correspond to the length-independent repression value that each cassette seems to approach. (C) States and weights schematic for the model used to describe the 1-Tet repression data. Note, that in this case there are a pair of additional states above and beyond those associated with the model presented in Fig. 1C. The two additional states correspond to the looped and unlooped configurations of the DNA with TetR bound to the enhancer. The looping capacity for the looped states with TetR bound is designated as χ_{short}.

Box 2: 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):

$$\frac{p_r}{p_o} = \frac{1}{1 + \frac{[R]}{K_d}} \quad (1)$$

where 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 Fig. 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 Fig. 1C and described in Box 1, 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 of Box I, the looped state with TetR bound can also

generate transcription, but with a different looping capacity function we term $\chi_{short}(L)$ (see associated Fig. 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:

$$p_{r,1} \propto \frac{\chi_{long}(L) + \frac{[T]}{K_{TD}}\chi_{short}(L)}{1 + \chi_{long}(L) + \frac{[T]}{K_{TD}}\left(1 + \chi_{short}(L)\right)} \tag{2}$$

where *[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 Fig. 2B, we divide eqn. (2) by the level of expression in the absence of repressor (eqn. (4) – Box I) to get:

$$\frac{p_{r,1}}{p_o} = \frac{1 + \frac{[T]}{K_{TD}}\frac{\chi_{short}(L)}{\chi_{long}(L)}}{1 + \frac{[T]}{K_{TD}}\left(\frac{1 + \chi_{short}(L)}{1 + \chi_{long}(L)}\right)}. \tag{3}$$

At saturating concentrations of TetR this expression reduces to

$$r_1(L) \equiv \lim_{[T]\to\infty} \frac{p_{r,1}}{p_0} = \frac{\frac{\chi_{short}(L)}{\chi_{long}(L)}}{\frac{1 + \chi_{short}(L)}{1 + \chi_{long}(L)}}. \tag{4}$$

This corresponds to the magnitudes shown in Fig. 2B and Fig. 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

$$\begin{aligned} \lim_{L\to\infty} \chi_{long}(L) &= \gamma_{long} L^{-\frac{3}{2}} \\ \lim_{L\to\infty} \chi_{short}(L) &= \gamma_{short} L^{-\frac{3}{2}} \end{aligned} \tag{5}$$

where γ_{short} and γ_{long} are constants that are dependent on the particular physical properties of the loop. These looping capacities will approach $\chi \ll 1$ for an appropriate large length, which implied that in this limit eqn. (4) approaches the constant value:

$$\lim_{L\to\infty} r_1(L) = \frac{\gamma_{short}}{\gamma_{long}}. \tag{6}$$

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 Fig. 2B by the dashed lines. However, a more rigorous treatment of this limit would require knowledge of the absolute values of the looping capacities.

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 shows 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=115bp. For this length only the 2-Tet cassette is found in the fully repressed state. 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 approximately 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 are identical to 3-Tet cassette partial occupancy 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,E) correspond to two (A) or three (C,E) piecewise continuous empirical Hill function fits that highlight the transitions and levels observed in the data.

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 (1,2,3,6) 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. These transitions indicate that the looping process generates a type of analog computation that is carried out by the regulatory region, counting binding sites in order to achieve complete repression for a particular spacer length. (E) Regulatory code. By shifting the cassette towards the σ^{54} promoter and away from the driver NRI#1,2 sites, a similarly shaped regulatory function (top data) is observed. Namely, a relatively strong repressed state at low aTc concentrations is followed by a sharp transition, to a stable intermediate weak repression level and a subsequent shallow increase in repression-ratio to unrepressed levels at 100%. This indicates that the regulatory output function is encoded within the shape and structure of the binding cassette.

Figure 5: Generalized Model schematic for repression-ratio data

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-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 2-3 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.

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 1-Tet case exhibiting a simple two state dose-response function. (B) States of the two TetR binding and associated two-state dose-response function (i.e lacking an intermediate step) constructed using the parameters from (A) and $\omega_s=1$ implying no anticooperativity between adjacent TetR molecules. (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. The effect of $\omega_s<1$ is to reduce the overlap between the double and no occupancy state distributions, leading to a wider single-occupancy state, which generates a stable intermediate level in the dose-response function. (D-F) Probability of each state and associated dose-response functions for three TetR binding sites for the cases (D-F) Probability of each state and associated dose-response functions for three TetR binding sites for the cases (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 $\omega_l=10^{-2}$, where ω_s and ω_l correspond to the anti-cooperativity terms used to model the nearest and next-to-nearest neighbors interactions respectively. Sample fits of the model to the data sets in Fig. 3 and 4 are shown in Fig. S5.

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).

Arnosti, D.N., Barolo, S., Levine, M., and Small, S. (1996). The eve stripe 2 enhancer employs multiple mode of transcriptional synergy. Development *122*, 205-214.

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 III, L.J. (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. (2005). Transcriptional regulation by the numbers: applications. Curr Opin Genet Dev *15*, 125-135.

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

Bolouri, H., and Davidson, E.H. (2002). Modeling DNA Sequence-Based cis-Regulatory Gene Networks. Dev Biol *246*, 2-13.

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.

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

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

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.

Driever, W., and Nusslein-Volhard, C. (1989). The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. Nature *337*, 138-143.

Driever, W., Thoma, G., and Nusslein-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 S., L. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335-338.

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

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

Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007). Probing the limits to positional information. Cell *130*, 153-164.

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.

Hasty, J., McMillen, D., and Collins, J.J. (2002). Engineered gene circuits. Nature *420*, 224-230.

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.

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.

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

Korzh, V. (2007). Transposons as tools for enhancer trap screens in vertebrates. Genome Biol *8 Suppl 1*, S8.

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., 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.

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

Lewis, D.E., and Adhya, S. (2002). *In vitro* repression of the gal promoters by GalR and HU depends on the proper helical phasing of the two operators. J Biol Chem *277*, 2498-2504.

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.

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

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.

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.

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

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

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).

Rappas, M., Bose, D., and Zhang, X. (2007). Bacterial enhancer-binding proteins: unlocking sigma-54 dependent gene transcription. Curr Op Struct Biol *17*, 110-116.

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.

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.

Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophilla embryo. EMBO J *11*, 4047-4057.

Stathopoulos, A., and Levine, M. (2005). Genomic regulatory networks and animal development. Dev Cell *9*, 449-462.

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

Swigon, D., Coleman, B.D., and Olson, W.K. (2006). Modelling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. Proc Natl Acad Sci U S A *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*, 25001.

Yuh, C.-H., Bolouri, H., and Davidson, E.H. (2001). Cis-regulatory logic in the endo16 gene: switching from specification to a differentiation mode of control. Development *128*, 617-629.

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.

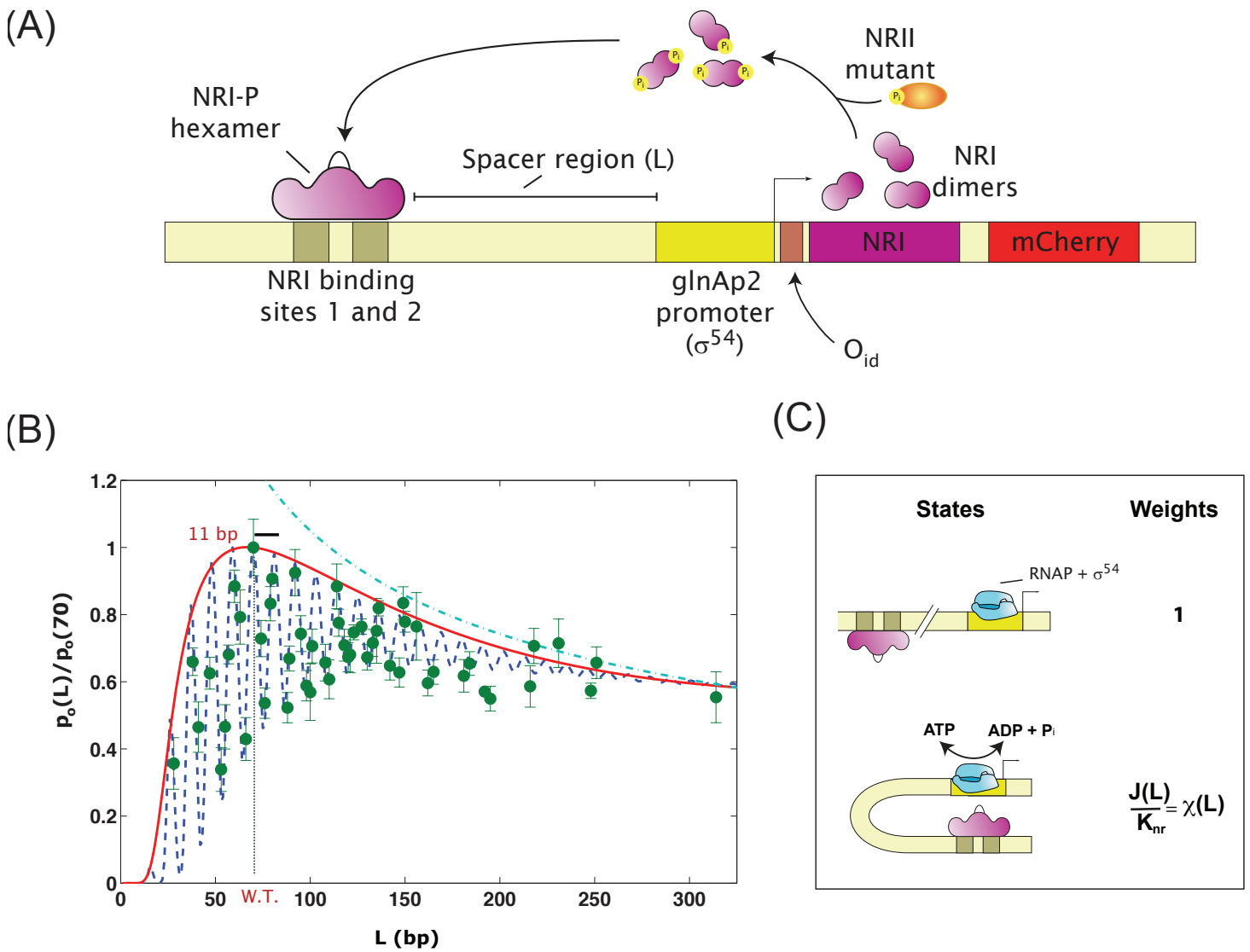

(A)
NRII
mutant
NRI-P
hexamer
Spacer region (L)
NRI
dimers
NRI
mCherry
NRI binding
sites 1 and 2
glnAp2
promoter
(σ54)
Oid
(B)
11 bp
W.T.
po(L)/po(70)
L (bp)
(C)
States
Weights
RNAP + σ54
1
ATP
ADP + Pi
J(L)/Knr = χ(L)

Figure 1

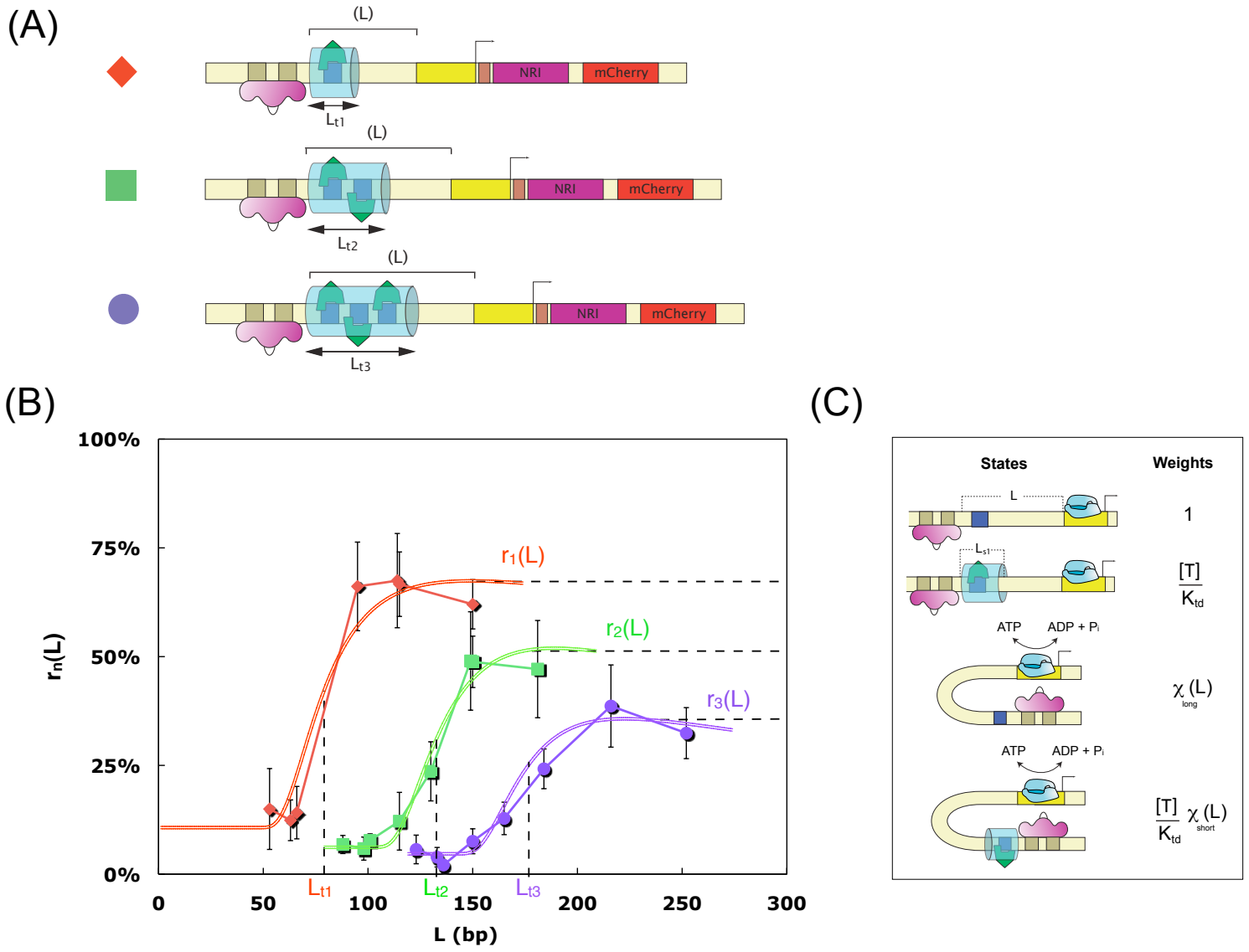

(A)
(L)
NRI
mCherry
Lt1
Lt2
Lt3
(B)
100%
75%
50%
25%
0%
rn(L)
r1(L)
r2(L)
r3(L)
0
50
100
150
200
250
300
L (bp)
(C)
States
Weights
L
Ls1
1
[T]/Ktd
ATP
ADP + Pi
χ (L) long
χ (L) short

Figure 2

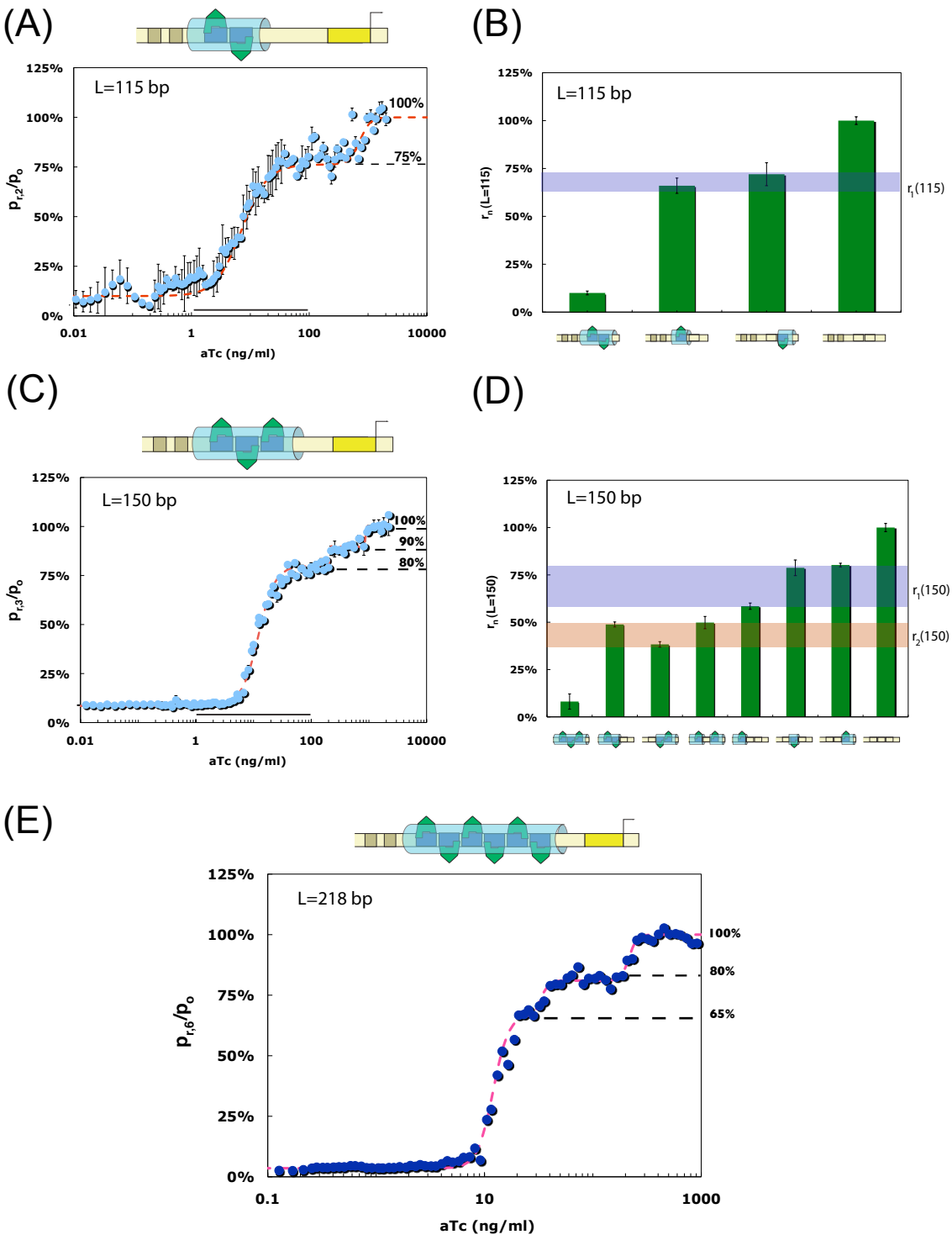

(A)
L=115 bp
100%
75%
125%
100%
75%
50%
25%
0%
0.01
1
100
10000
aTc (ng/ml)
(B)
L=115 bp
r1(115)
(C)
L=150 bp
100%
90%
80%
aTc (ng/ml)
(D)
L=150 bp
r1(150)
r2(150)
(E)
L=218 bp
100%
80%
65%
0.1
10
1000
aTc (ng/ml)

Figure 3

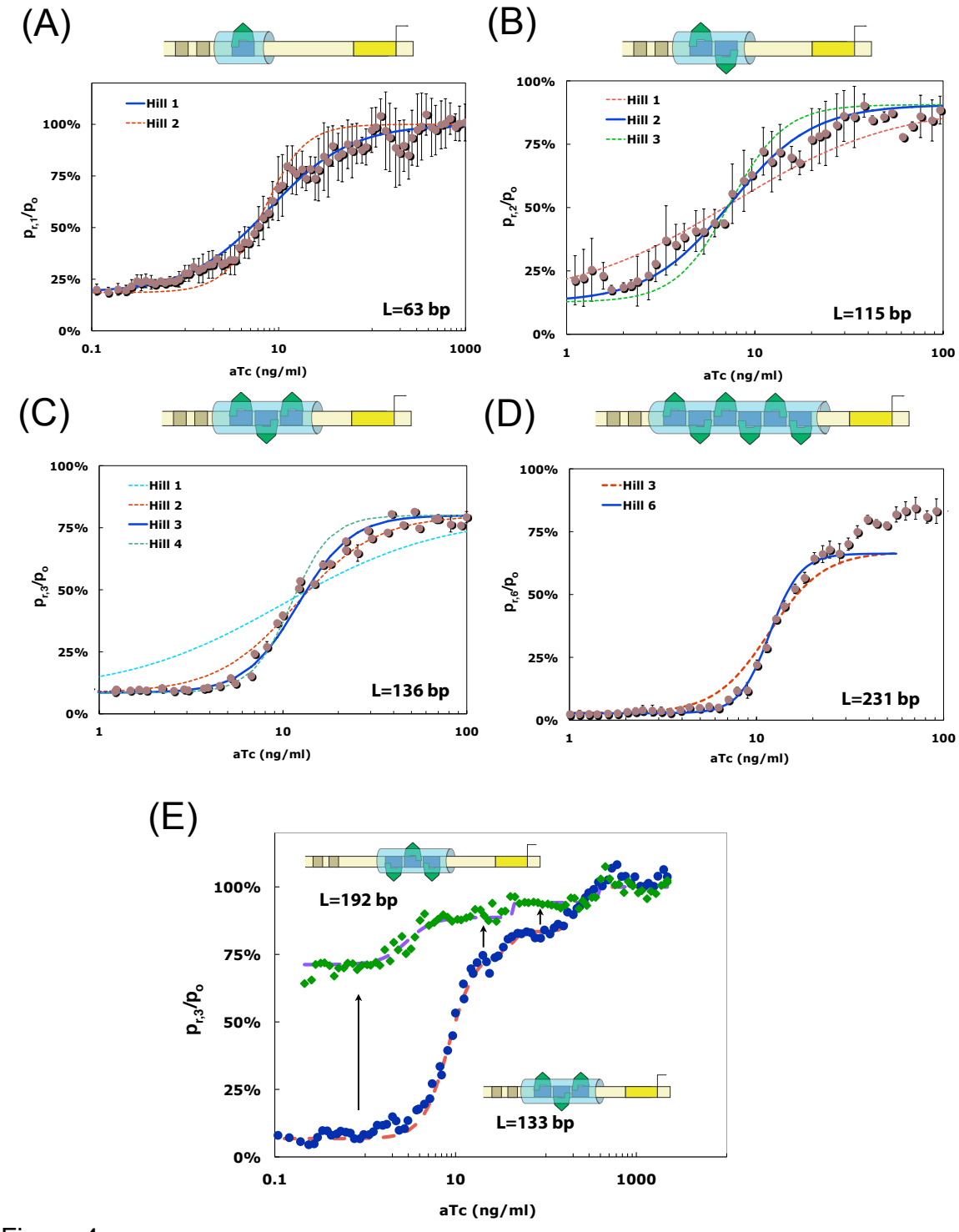
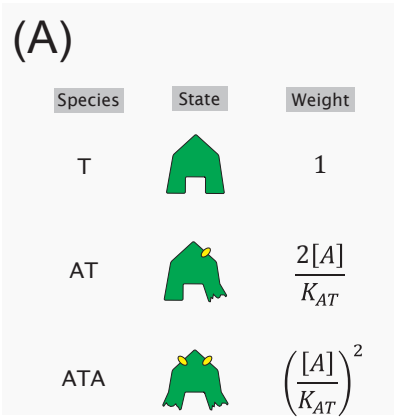
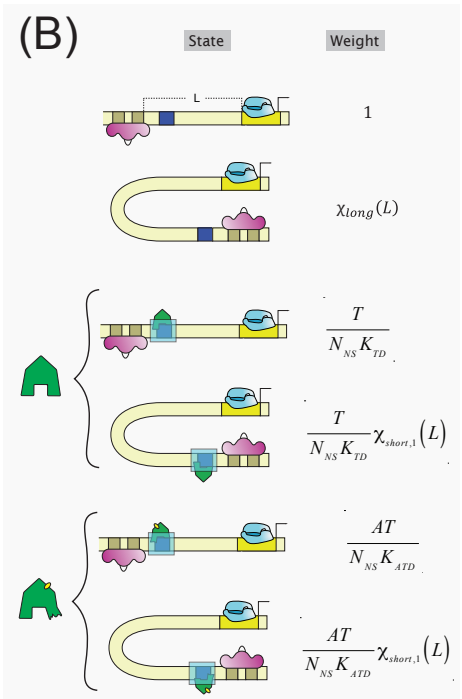

(A)
Hill 1
Hill 2
100%
75%
50%
25%
0%
$p_{r,1}/p_o$
L=63 bp
0.1
10
1000
aTc (ng/ml)
(B)
Hill 1
Hill 2
Hill 3
$p_{r,2}/p_o$
L=115 bp
1
10
100
aTc (ng/ml)
(C)
Hill 1
Hill 2
Hill 3
Hill 4
$p_{r,3}/p_o$
L=136 bp
aTc (ng/ml)
(D)
Hill 3
Hill 6
$p_{r,6}/p_o$
L=231 bp
aTc (ng/ml)
(E)
L=192 bp
L=133 bp
$p_{r,3}/p_o$
aTc (ng/ml)

Figure 4


(A)
Species
State
Weight
T
1
AT
$\frac{2[A]}{K_{AT}}$
ATA
$\left(\frac{[A]}{K_{AT}}\right)^2$


(B)
State
Weight
L
1
$\chi_{long}(L)$
$\frac{T}{N_{NS}K_{TD}}$
$\frac{T}{N_{NS}K_{TD}}\chi_{short,1}(L)$
$\frac{AT}{N_{NS}K_{ATD}}$
$\frac{AT}{N_{NS}K_{ATD}}\chi_{short,1}(L)$

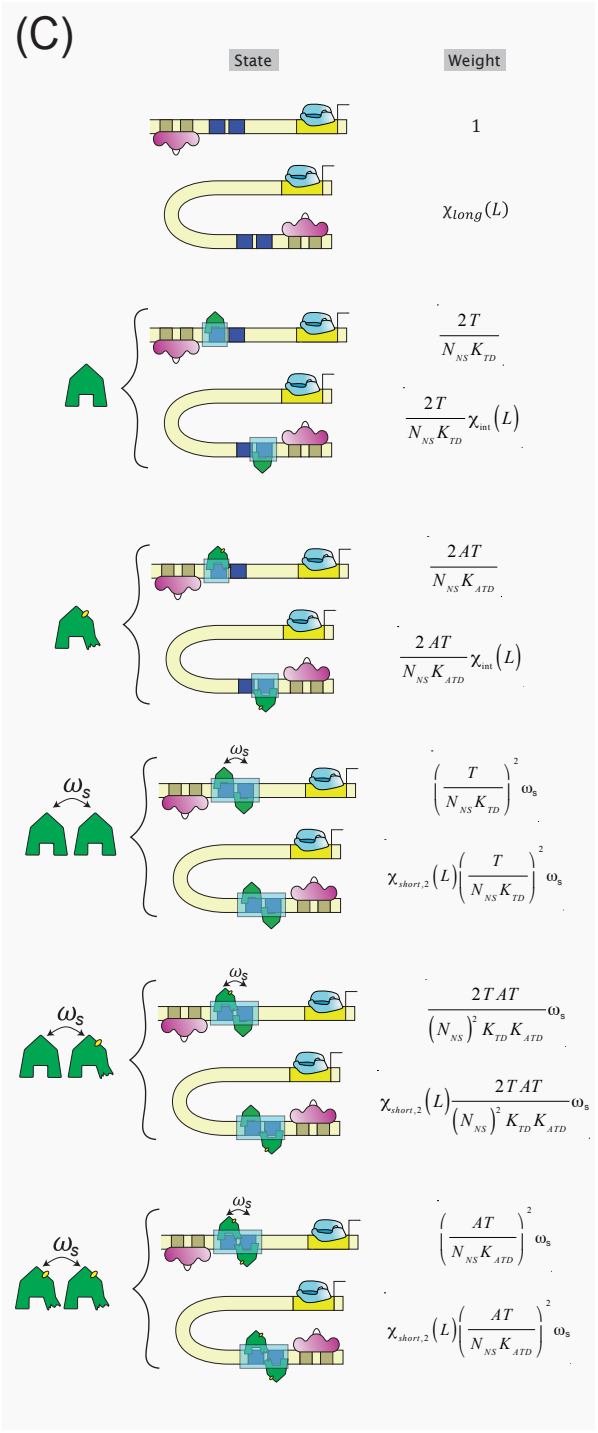

(C)
State
Weight
1
$\chi_{long}(L)$
$\frac{2T}{N_{NS}K_{TD}}$
$\frac{2T}{N_{NS}K_{TD}}\chi_{int}(L)$
$\frac{2AT}{N_{NS}K_{ATD}}$
$\frac{2AT}{N_{NS}K_{ATD}}\chi_{int}(L)$
ω_s
$\left(\frac{T}{N_{NS}K_{TD}}\right)^2\omega_s$
$\chi_{short,2}(L)\left(\frac{T}{N_{NS}K_{TD}}\right)^2\omega_s$
$\frac{2TAT}{(N_{NS})^2K_{TD}K_{ATD}}\omega_s$
$\chi_{short,2}(L)\frac{2TAT}{(N_{NS})^2K_{TD}K_{ATD}}\omega_s$
$\left(\frac{AT}{N_{NS}K_{ATD}}\right)^2\omega_s$
$\chi_{short,2}(L)\left(\frac{AT}{N_{NS}K_{ATD}}\right)^2\omega_s$

Figure 5

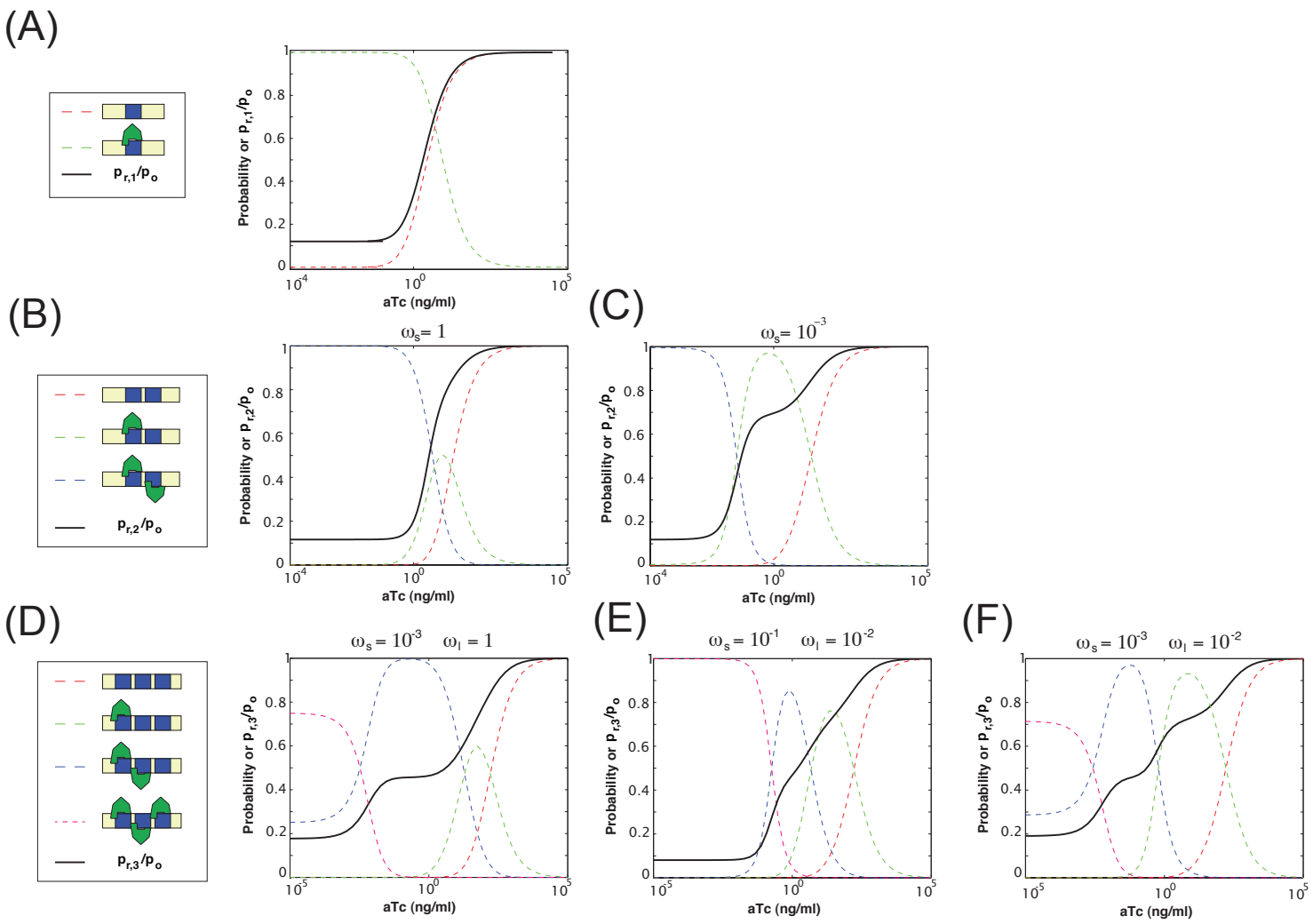

(A)
$p_{r,1}/p_o$
Probability or $P_{r,1}/P_o$
aTc (ng/ml)
(B)
$p_{r,2}/p_o$
$\omega_s = 1$
Probability or $P_{r,2}/P_o$
aTc (ng/ml)
(C)
$\omega_s = 10^{-3}$
Probability or $P_{r,2}/P_o$
aTc (ng/ml)
(D)
$p_{r,3}/p_o$
$\omega_s = 10^{-3}$ $\omega_l = 1$
Probability or $P_{r,3}/P_o$
aTc (ng/ml)
(E)
$\omega_s = 10^{-1}$ $\omega_l = 10^{-2}$
Probability or $P_{r,3}/P_o$
aTc (ng/ml)
(F)
$\omega_s = 10^{-3}$ $\omega_l = 10^{-2}$
Probability or $P_{r,3}/P_o$
aTc (ng/ml)

Figure 6