



Published in final edited form as:

Cell Rep. 2021 November 09; 37(6): 109952. doi:10.1016/j.celrep.2021.109952.

Quantifying the regulatory role of individual transcription factors in *Escherichia coli*

Sunil Guharajan¹, Shivani Chhabra², Vinuselvi Parisutham¹, Robert C. Brewster^{1,3,4,*}

¹Department of Systems Biology, University of Massachusetts Chan Medical School, Worcester, MA 01605, USA

²Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

³Department of Microbiology and Physiological Systems, University of Massachusetts Chan Medical School, Worcester, MA 01605, USA

⁴Lead contact

SUMMARY

Gene regulation often results from the action of multiple transcription factors (TFs) acting at a promoter, obscuring the individual regulatory effect of each TF on RNA polymerase (RNAP). Here we measure the fundamental regulatory interactions of TFs in *E. coli* by designing synthetic target genes that isolate individual TFs' regulatory effects. Using a thermodynamic model, each TF's regulatory interactions are decoupled from TF occupancy and interpreted as acting through (de)stabilization of RNAP and (de)acceleration of transcription initiation. We find that the contribution of each mechanism depends on TF identity and binding location; regulation immediately downstream of the promoter is insensitive to TF identity, but the same TFs regulate by distinct mechanisms upstream of the promoter. These two mechanisms are uncoupled and can act coherently, to reinforce the observed regulatory role (activation/repression), or incoherently, wherein the TF regulates two distinct steps with opposing effects.

In brief

Guharajan et al. investigate the isolated regulation of 6 *E. coli* transcription factors. The diverse regulatory outcomes are well described by a model wherein TFs act on two different steps of transcription. The degree to which each step is regulated depends on TF identity and binding location.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*Correspondence: robert.brewster@umassmed.edu.

AUTHOR CONTRIBUTIONS

S.C. and R.C.B. designed the initial experiments. S.C., V.P., and S.G. created strains for the experiments. S.G. and S.C. performed experimental measurements. S.C., S.G., and R.C.B. analyzed the data. S.G. and R.C.B. developed the theory. S.G. and R.C.B. wrote the manuscript.

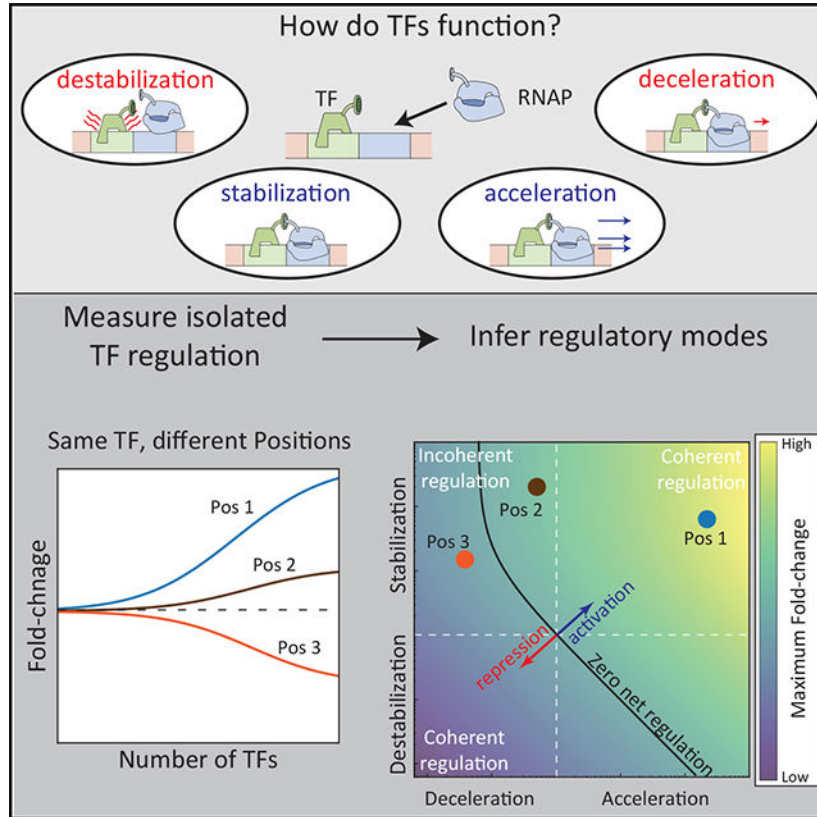
DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.109952>.

Graphical Abstract



INTRODUCTION

Transcriptional regulation of gene expression is one of the major mechanisms by which cells respond to cues and stimuli. Transcription factors (TFs) perform this regulation through binding to the DNA around the promoter to alter the rate of transcription from individual genes (Jacob and Monod, 1961; Ptashne and Gann, 2002). The regulatory DNA of each gene is distinct and can involve several to dozens of TF binding sites arranged in specific architectures to achieve the desired expression level. However, predicting the level of gene expression based on the regulatory architecture of a gene remains a central challenge in the field (Guido et al., 2006; Sprinzak and Elowitz, 2005; Atkinson et al., 2003; Nielsen et al., 2016; Ali et al., 2020).

The genomics era has enabled multiple techniques capable of determining where a TF will bind and with what specificity (Stormo, 2000; Messeguer et al., 2002; Wasserman and Sandelin, 2004; Weirauch et al., 2014). Although this information is crucial for building occupancy-based models of gene regulation, there is still another critical component that is missing; the quantitative regulatory role of a TF, when bound, is often unclear. Historically, measurements of TF function are the result of knocking out the endogenous TF and observing how gene expression changes as a result. This serves a purpose in predicting the specific role of that TF on a given gene but offers less predictive power when

examining regulation of other genes by that same TF or to different binding sites. These measurements of gene regulation are often entangled in indirect regulatory effects such as TF-TF interactions (Vidal et al., 2011; Rolland et al., 2014), feedback (Bateman, 1998; Shen-Orr et al., 2002), and physiological (i.e., growth rate) (Klumpp et al., 2009; Schmidt et al., 2016) and off-target competitive effects of decoy binding sites or other genes in the network (Brewster et al., 2014; Lee and Maheshri, 2012). Because of this, a single TF that binds at the same relative location on two different natural promoters can appear to have opposite regulatory roles. The entanglement between indirect and direct regulation likely contributes to this ambiguity and prevents the field from developing a basic intuition of TF regulatory function. The select few TFs that have arisen as “model TFs,” such as LacI (Oehler et al., 1990; Garcia and Phillips, 2011; Hammar et al., 2014; Daber et al., 2011; Vilar and Saiz, 2013), AraC (Lobell and Schleif, 1990; Schleif, 2010; Egan et al., 2000; Egan, 2002), lambda repressor (Ptashne et al., 1980; Shea and Ackers, 1985; Ptashne, 2004), CRP (Gaston et al., 1989; Ushida and Aiba, 1990; Kolb et al., 1993; Kinney et al., 2010; Forcier et al., 2018), and TetR (Ramos et al., 2005; Deng et al., 2013; Stanton et al., 2014; Gardner et al., 2000), have well-studied regulatory function. Indeed, these TFs have been utilized for design of synthetic circuits with an engineered purpose, such as creation of logic gates (Stanton et al., 2014; Anderson et al., 2007), bistable switches (Gardner et al., 2000; Tan et al., 2009), oscillatory networks (Elowitz and Leibler, 2000; Stricker et al., 2008; Atkinson et al., 2003), synthetic enhancers (Amit et al., 2011; Brunwasser-Meirom et al., 2016), and a host of other dynamic outcomes. Further characterizing the regulatory function of TFs beyond this small subset should provide a more complete toolset for broader synthetic design purposes.

Here we study the isolated regulatory function of a set of *E. coli* TFs in a system designed to remove the typical confounding factors of natural genes and quantify the direct regulatory effect of a TF based on factors such as TF concentration, binding affinity, and binding location. Using a collection of strains where the average copy number of most TFs in the cell can be controlled, we measure the level of regulation of an individual TF acting on a synthetic promoter sequence. This promoter is designed to be regulated only by that TF, and it is targeted to a binding site whose location and sequence we control. To interpret these data, we use a thermodynamic model of gene regulation to parameterize TF regulatory function. In principle, the TF could exert its regulatory effect at any one of the distinct kinetic steps of the transcriptional process (Kontur et al., 2008; Revyakin et al., 2006; Young et al., 2002; Henderson et al., 2017; Jensen and Galburt, 2021) or on several of them, and our model coarse-grains TF activity into two distinct modes of regulation. The first regulatory mode, “stabilization,” corresponds to stabilization (or destabilization) of the polymerase at the promoter by the TF and models the TF’s ability to facilitate the emergence of the closed RNA polymerase (RNAP)-DNA complex. Essentially, this reflects a modification of the off-rate (k_{off}) of bound RNAP when the TF is co-bound at the promoter, resulting in longer or shorter dwell times of RNAP (β greater than or less than 1) (Roy et al., 1998; Liu et al., 2017). The second mode, “acceleration,” corresponds to a TF’s ability to accelerate (or decelerate) initiation of transcription when the TF and polymerase are bound to the promoter. Canonically, the rate of transcription is dictated by the progress of several intermediate steps during promoter melting (Feklistov, 2013; Feklistov et al.,

2017; Boyaci et al., 2019), and the TF's propensity to accelerate or decelerate transcription can be viewed as acting on the ability of RNAP core subunits to initiate this process (Rhodius and Busby, 2000; Feng et al., 2016; Lee and Maheshri, 2012). Using this model, we infer the quantitative contribution from each of these modes in the data. Importantly, this process allows decoupling of properties that are extrinsic to the TF, such as affinity to the operator binding site, the overall concentration of the TF, or feedback in the network from the core regulatory role of the TF in modulating the steps of the transcription process.

We expect the regulatory parameters of a TF to vary based on the identity of the TF and the binding site location on the regulated promoter. In this study, we investigate the role of TF identity by measuring regulation of 6 TFs (AcrR, AgaR, ArsR, AscG, BetI, and CpxR). These 6 TFs were selected based on their diverse *in vivo* functions, which encompass multidrug resistance (AcrR) (Gu et al., 2008), regulation of metabolic homeostasis (AgaR, AscG, and BetI) (Ishida et al., 2009; Lamark et al., 1996; Ray and Larson, 2004; Leyn et al., 2012), tolerance of heavy metal toxicity (ArsR) (Ren et al., 2017), and coordination of the envelope stress response (CpxR) (DiGiuseppe and Silhavy, 2003; Hews et al., 2019). Furthermore, these TFs encompass 5 distinct TF regulatory families and require different co-factors and allosteric configurations to realize their regulatory function (Salgado et al., 2013). We tested each of these TFs at two common binding locations: directly downstream of the promoter, where repression is commonly observed, and 61 bases upstream of the promoter, a site commonly associated with activation (although databases of regulatory interactions record roughly as many TFs repress at -61 as activate). We find that, despite the diverse nature of the TFs tested (five of the TFs are annotated repressors, and one of them, CpxR, is a known activator), the regulation for all TFs immediately downstream is consistent with a form of repression that is set by the degree of occupancy of the TF at the promoter independent of TF identity. This commonality across the TFs disappears when we measure the effect at -61, where the TFs exhibit different degrees of stabilization, with CpxR and AgaR engaging in significant stabilization of RNAP. To compliment this, we took CpxR and systematically quantified the contribution of the regulatory modes as a function of TF binding location and find that CpxR sets the degree of activation by engaging in two distinct regulatory paradigms. Binding locations that see strong activation have CpxR engaging in "coherent" regulation: the activation is enforced by stabilization and acceleration of RNAP. Locations with weak activation, however, have CpxR regulating the two modes oppositely by stabilizing RNAP but slowing the rate of promoter escape, demonstrating that such "incoherent" regulation plays a useful role by allowing a single TF to generate a spectrum of regulatory responses emerging from the relative effects of the TF on these distinct steps.

RESULTS

Thermodynamic model for single TF regulation

To deconvolve the role of TF copy number, binding affinity, and binding location from the intrinsic regulatory interactions of the TF with polymerase, we use a thermodynamic model of gene expression (Ackers et al., 1982; Kuhlman et al., 2007; Bintu et al., 2005; Kinney et al., 2010; Buchler et al., 2003; Vilar and Leibler, 2003; Garcia et al., 2012), where we consider only a single TF acting on an otherwise unregulated gene. Figure 1A

shows the various promoter states considered in the model (left column), along with the relative probability of each state occurring (center column) and the rate of expression from each promoter state (right column); the promoter can be unbound by the TF and RNAP, bound by polymerase only, bound by the TF only, or bound by both. The probability with which these states occur is a function of each molecule's (polymerase and TF) binding affinity with its specific DNA binding sites (ϵ_P and ϵ_{TF}) and the available number of each molecule in the cell (N_P and N_{TF}). For the co-bound state, we consider two distinct mechanistic influences of the TF on gene expression. The first effect represents altered stability of the polymerase at the promoter when TF is bound because of a favorable or unfavorable interaction between the TF and polymerase. As a result, the co-bound state occurs with increased relative probability to the single bound state by a factor β (implying an energetic interaction of $\log(\beta)$ in units of $k_B T$). The second parameter α represents the change in transcription rate when the TF and polymerase are co-bound and is written as a multiplicative factor to the base expression rate of polymerase bound in the absence of the TF; for example, $\alpha = 2$ would imply that the transcription initiation rate is doubled when the TF and polymerase are co-bound. In both cases the parameters represent increases in gene expression when greater than unity and decreases in gene expression when less than unity. Importantly, the parameters are not constrained and can, in principle, have opposing or compounding effects; i.e., this model allows a TF that stabilizes polymerase binding but slows the rate of transcription from that state, resulting in apparent activation or repression, depending on the relative strengths of those effects. The final parameter, N_{NS} , is equated to the size of the genome in base pairs (4.6×10^6) and is not varied in our experiments (for more details, see Phillips et al., 2019, and Bintu et al., 2005). Furthermore, the parameters related to polymerase binding can be simplified into a single parameter in our model as $P = N_P \exp(-\epsilon_P) / N_{NS}$.

The final expression, boxed in Figure 1A (and derived in the STAR Methods), predicts the fold change in gene expression of a target gene. Fold change is defined as the expression level of the target gene in the presence of a number of TFs (N_{TF}) divided by the expression level in the absence of that TF (i.e., $N_{TF} = 0$). A fold change greater than 1 signifies activation, whereas a fold change below 1 signifies repression. The fold change equation is simplified by collecting the regulatory parameters into two effective parameters: FC_{max} and χ . FC_{max} represents the fold change when the number of TFs in the system are saturating; in the case of a repressor, it is the minimum fold change achievable, and in the case of an activator, it is the maximum fold change achievable. Importantly, FC_{max} depends only on a TF's degree of acceleration (α) and stabilization (β) and not TF binding affinity or concentration. The second term, χ , represents the rate at which the fold change approaches FC_{max} . This rate of approach depends on the TF binding affinity (ϵ_{TF}) and the degree to which the TF recruits/stabilizes RNAP ($1 + \beta P$). These factors, together with the number of TFs (N_{TF}), can be thought of as an effective TF concentration.

These effective parameters are useful because they transform this system with many variables (TF and polymerase binding affinity, TF and polymerase number, degree of acceleration, degree of stabilization, etc.) capable of producing a diverse range of response curves in the fold change versus TF copy number space into a very simple system dictated by two fundamental quantities: the maximum fold change, FC_{max} , and the effective TF

concentration, χN_{TF} . This is demonstrated in Figure 1B, where we plot fold change against TF number for an activator (top curves) and a repressor (bottom curves) with $FC_{max} = 100$ and $FC_{max} = 1/100$, respectively. In each scenario, we plot 3 colored curves. The red curve has no contribution from stabilization ($\beta = 1$). The blue curve is identical to the red curve except with slightly stronger TF binding affinity, and the green curve is again identical to the red curve except with significant contribution from stabilization ($\beta = 10$ with α adjusted to keep FC_{max} unchanged). Figure 1C demonstrates the fold change as a function of effective TF concentration, χN_{TF} . When plotted this way, the data from all three curves collapses to a single curve that is determined entirely by the value of FC_{max} , independent of specific values of α and β . Points with identical TF concentrations in Figure 1B now scatter on the collapsed curves, and the green and blue points (which had higher values of χ) are farther along the curve; a large value of χ hastens the approach to FC_{max} for the same TF concentration (N_{TF}). Specifically, two TFs with similar net regulatory effect (similar FC_{max}) that operate through different regulatory mechanisms (for instance, one through strong acceleration [large α , $\beta \approx 1$] and one through strong stabilization [large β , $\alpha \approx 1$]) will trace out exactly the same curve in this space, but the strong stabilizer will have a higher effective TF concentration and, thus, move farther along the curve given the same TF concentrations and binding affinities. In Figure 1D, data collapse curves for a range of FC_{max} values are shown.

Figure 1E provides a hierarchical overview of model parameters and their relationship to controllable biological features of the *in vivo* system. The top level shows the effective parameters FC_{max} and χN_{TF} which define the contour of regulatory curves like those in Figure 1D. These effective parameters are composed of a combination of the “physical” parameters on the second level of the diagram. These parameters correspond to basic features of the system, such as numbers of molecules, affinities, or interactions between molecules. The third level of the diagram shows the biological controls we have available to control the corresponding physical parameters. The approach we take below will be to profile the regulatory function and characterize the inherent regulatory parameters (α and β) of six TFs (AgaR, ArsR, AcrR, AscG, BetI, and CpxR) by controlling the copy number, binding location, and binding sequence of each. A potential challenge in determining α and β stems from their connectedness in the effective parameters χ and FC_{max} . Because χ is proportional to $(1 + \beta P)$, if the stabilization parameter β is much smaller than $1/P \approx 15$ (measured previously for the promoter sequence used in our experiments; Brewster et al., 2012), then χ will no longer strongly depend on β because $(1 + \beta P) \approx 1$. In cases such as this, which we label “weak stabilization,” we are left with only one effective parameter, FC_{max} , to determine the two regulatory parameters α and β , and it is not possible to distinguish between regulation driven by a change in the transcription rate (acceleration/deceleration) or by modulation of polymerase occupancy at the promoter (stabilization/destabilization).

Experimental measurements of individual TF regulatory function

To measure regulation by an individual TF in *E. coli* as a function of TF identity and binding position on the promoter, we utilize synthetic techniques to create simple and controllable gene circuits. In this approach, we have created *E. coli* strains, illustrated

schematically in Figures 2A and 2B, where the endogenous copy of each studied TF is knocked out and reintroduced as a TF-mCherry fusion integrated into the genome at the *ycbN* locus (Figure 2A). Expression of the synthetic TF-mCherry promoter can be induced with anhydrotetracycline (aTc) (for details, see STAR Methods). This system enables precise control of TF copy number (Figure 2D), which is measured by wide-field fluorescence microscopy or flow cytometry (Figure 2C; STAR Methods). In our data, we use a fluctuation-counting method to convert the arbitrary fluorescence from the microscope into TF copy number. We convert the arbitrary fluorescence from the cytometer to TF number through a reference measurement on the microscope measured in parallel on the flow cytometer. This process has been described previously (Rosenfeld et al., 2007; Teng et al., 2010; Brewster et al., 2014) and in more detail in the STAR Methods. Using this system, we are able to induce these TFs from leaky levels (several per cell) up to several thousand per cell; the full induction curve of each strain is shown in Figure 2D. Importantly, although we control the concentration of the TF, many TFs are capable of existing in distinct binding conformations that may alter the active TF number, the regulatory parameters (α and β), or both. For all but one of the TFs studied here, we expect that the TFs will always be active under our growth conditions; however, BetI is inactivated by choline (which we do not control for), and therefore we expect to have some fraction of BetI inactive at low TF concentrations (Lamark et al., 1991; Gama-Castro et al., 2016).

In each of these individually tunable *tf-mCherry* strains, a target promoter that drives YFP expression is integrated into the genome at the *galK* locus (Figure 2B). The basic promoter incorporates a modified *lac* RNAP binding sequence where the RNAP occupancy term, P , was measured previously (Brewster et al., 2012). Otherwise, the promoter is designed to be free of specific known TF binding sequences. To study the regulatory role of a specific TF, we introduce a TF-specific binding site (chosen from an array of known binding sites with strong evidence of that particular TF binding (Salgado et al., 2013; Keseler et al., 2017)) cloned directly downstream of the transcription start site (TSS), annotated as +1 for simplification, or centered at 61 bases upstream of the TSS, annotated as -61. The effect of TF binding to the promoter is then measured in terms of YFP fluorescence protein expression as a function of average number of TFs per cell for a given induction condition (Figure 2C). For this work, our focus is to select a binding site that will bind the TF. The affinity of the site or how the particular choice may influence the regulatory parameters is not something we explore exhaustively here. The specific binding sites chosen for each TF can be found in the STAR Methods.

Regulatory response of six different TFs at +1 and -61

Figures 3A–3F show the fold change in YFP expression (promoter activity) as a function of TF copy number for the TFs examined in this study, measured using single-cell fluorescence microscopy. In these plots, regulation at +1 is shown as red points, regulation at -61 is shown as green points, and a control promoter with no TF binding site is shown in blue. The six TFs display diverse regulatory behavior that depends on the TF identity and TF binding location on the gene. For example, CpxR (Figure 3F) is a repressor when bound at +1 but activates at -61. AscG (Figure 3D) is a strong repressor at +1 but has almost no regulatory role at -61, and BetI (Figure 3E) represses at +1 and -61 but much more weakly

at -61 despite binding to the same sequence at both locations. One commonality between the curves is that, at +1, every TF in this study acts as a repressor. Naively, the “strength” of this repression appears rather diverse; the repression from some TFs reduces YFP expression by 10^{-3} in fold change (BetI), whereas others never drop below 10^{-1} (AgaR). The solid lines in this figure show a fit to the model in Figure 1A (boxed equation). In this case, we are explicitly measuring the number of TFs, N_{TF} , and we fit for the two unknown parameters χ and FC_{max} (Figure 1A). In four of the six curves (Figures 3A and 3D–3F), the +1 regulation data give FC_{max} consistent with zero. This result is consistent with the regulatory mode of perfect repression; i.e., that the TF completely shuts off the gene when bound. Therefore, the difference in regulation at +1 between the TFs must be attributed entirely to binding affinity or differing levels of stabilization (β) between TFs. A typical assumption is that TFs operating at +1 regulate by steric hindrance ($\beta = 0$) so that, when the TF is bound at +1, polymerase binding is occluded. Previous studies (Ptashne and Gann, 1998; Ackers et al., 1982; Brewster et al., 2014; Forcier et al., 2018) support this assumption. The remaining curves (Figures 3B and 3C) show FC_{max} of order 10^{-1} , but in both cases, the binding is weak to the point where we do not see the expected saturation of fold change with TF copy number, and, therefore, the data from these curves are also consistent with FC_{max} of zero with a slightly increased value of χ to compensate. The collapse of all +1 regulation data to the perfect repression contour ($FC_{max} = 0$) is demonstrated in Figure 4A, where we plot the fold change against the effective TF concentration, χN_{TF} , for these six TFs; the +1 data for all six TFs largely fit to a single regulatory contour associated with perfect repression when the extrinsic features, such as TF copy number and binding affinity, are “normalized away.”

Although, at +1, the fold change curve for each TF collapsed on a unifying regulatory profile (with $FC_{max} = 0$), at -61, these TFs operate with a diverse range of regulatory effects; some TFs mirror the function at +1, showing a profile similar to the response function at +1, whereas others show limited repressive capabilities that saturate at specific fold change values (FC_{max}), and still other TFs activate expression at -61. To quantify FC_{max} and χ for each TF acting at -61, we fit the data in Figure 3 with the theory in the fold change equation above. We find that the repressive TFs have FC_{max} values ranging between 0.2 and 0.7, whereas the lone activating TF is around 4. In Figure 4B, we plot the fold change data against the effective TF concentration, χN_{TF} . Now, rather than each TF following the same trajectory, the regulation data for these six TFs follow unique trajectories corresponding to specific values of FC_{max} . Figure 4C demonstrates that FC_{max} at +1 and -61 for these six TFs is not correlated between the two locations.

In Figure 4D, we show the fit value of χ for each TF at -61 against the fit value of χ at +1. Recall that χ is composed of the product of two effects: the binding affinity and the stabilization. We do not expect the TF binding energy, ϵ_{TF} , to depend on where the binding site is located because the binding sequence is not changed (Salgado et al., 2013). The other possible contribution to χ comes from stabilization and takes the form $(1 + \beta P)$, which means that β must be of order $1/P$ or larger for stabilization to affect χ . We expect that the data points will fall into two potential outcomes. Points that lie on the black dashed line indicate that stabilization does not play a significant role in the regulation at -61. However, points that are above the black dashed lines imply that the TF utilizes stabilization of RNAP (shown as red lines in Figure 4D). Many of the data points are consistent with

small or zero β , but two TFs, corresponding to AgaR and CpxR, are significantly above this line, implying that stabilization may play a role in their respective regulation. The implied magnitude of stabilization is shown by the red lines. Interestingly, AgaR in this case is a repressor but appears to impart strong stabilization, suggesting that, even though this TF stabilizes ($\beta > 1$) the polymerase at the promoter, it more strongly decreases the rate of transcription from polymerase bound at the promoter ($\alpha < 1$), resulting in a net repression of gene expression. This highlights a mechanism of repression that is fundamentally distinct from the downstream (+1) position regulation for AgaR and demonstrates an incoherent regulatory strategy where the TF engages RNAP at two distinct steps with opposing effects.

Profiling the spatial regulatory landscape of the CpxR TF

We now examine how the regulatory parameters that quantify stabilization (β) and acceleration (α) vary with binding location for one TF, CpxR. As demonstrated in Figure 5A, CpxR naturally binds to a wide range of promoter locations to regulate dozens of different genes in *E. coli*. However, the regulatory role of CpxR as a function of binding location is unclear from these data; repressive and activating interactions are attributed to many of the locations upstream of the promoter. Furthermore, we have evidence that CpxR is capable of regulating through stabilization at -61 , implying that we may be able to separate regulation through stabilization and regulation through acceleration in a more thorough examination. To measure the isolated regulatory behavior of CpxR as a function of binding location, we take the same synthetic target gene (Figure 2B) and move the TF binding site between positions centered at -48 bp from the TSS to -112 bp from the TSS (Figure 5B). We chose this range because the vast majority of natural CpxR binding sites occur within these limits (Figure 5A). To enable rapid cloning and measurement, the target gene is cloned into a low-copy plasmid (rather than integrated into the genome), and fold change in target expression and TF abundance is measured using flow cytometry (rather than single-cell microscopy) (Table 1). We find consistent results with the target gene on the plasmid or integrated into the genome (STAR Methods).

The data for fold change as a function of CpxR copy number for the binding location sweep is shown in Figure 5D. Regulation of the positions studied here is primarily activation, with 11 positions showing increased expression ranging from 2-fold to over 100-fold and just one upstream position showing moderate repression. We find that the regulatory effect of CpxR depends strongly on binding site location; the -54 binding location shows weak repression but is flanked by activating positions only a few bases away (-50 and -56). As expected, positions that are far from the promoter (in this case, beyond roughly 82 bp) show little to no regulation. In Figure 5C, we show fit values of FC_{max} as a function of binding location on the promoter for these data. We expected to see an 11-base periodicity in FC_{max} corresponding to the helicity of DNA (Müller et al., 1996; Garcia et al., 2012; Sharon et al., 2012); we see this very roughly with maxima in activation around -48 , -60 , -70 , and -80 (± 1 bp), but -64 is a strong outlier that is expected to be close to a minimum but is strongly activating as measured here.

To precisely extract the regulatory parameters α and β from these data, it is helpful to replot the fold change data using a manifold approach. A recent approach used regulatory

manifolds to explore how the CRP TF acted on dozens of different promoters and employed different RNAP binding sequences to trace out the regulatory space of the TF (Forcier et al., 2018). In this case, the power in the manifold approach arises from the ability to eliminate the need to measure (or know) the binding affinity of polymerase to each promoter; by replotting the data, specific data trajectories (or “allelic manifolds”) corresponded to unique values of CRP regulatory parameters. For our study, we use a “concentration manifold” to avoid the need to measure the effective concentration (in particular, the binding affinity of the TF). This approach is demonstrated schematically in Figure 6A. The concentration manifold plots the fold change at one position against the fold change of another position; each data point in this space corresponds to a measurement of both TF binding positions at the same effective TF concentration. In this case, we chose to plot all position data against the corresponding fold change at +1; we chose +1 because, based on Figure 4A, we believe that regulation is “pure steric hindrance” ($FC_{max} = 0$, $\beta = 0$) here. The advantage of this approach is that it enables inference of β at other positions based on the curvature of the data without the need to simultaneously infer the TF binding affinity or TF copy number (see STAR Methods for model assumptions). As seen in Figure 6B, data in the concentration manifold are expected to be lines emanating from (1, 1), which is the defined fold change of both locations when the number of TFs is zero. As TF concentration is increased, the fold change at +1 decreases toward 0, and the fold change at the second position will increase or decrease depending on the regulatory function at that position; therefore, activating positions have curves that rise as you move toward zero on the x axis, whereas repressive positions decrease. The role of β is clear in this formulation. When β is small (compared with $1/P$), the profiles are straight lines. However, larger β will cause the curve to rise or fall more rapidly than linear, reaching FC_{max} at higher values of the corresponding +1 Fold-change.

In Figure 6C, we replot the data from Figure 5D as fold change at an upstream regulatory position against fold change at +1; each data point represents a measurement of fold change at these two different binding locations for a given TF copy number within the cell. The plots are arranged from the highest FC_{max} (strong activation at -64) to the lowest FC_{max} (weak repression at -54). The solid line represents our model curve using the inferred values of α and β . Based on the inferred stabilization values, we find that the activation profiles across the position sweep is driven by varying degrees of stabilization and acceleration. The inferred values for α and β of all measured positions are shown in Table 2. We find strong stabilization in regulation at positions -50, -60, -54, -48, and -64; Figure 6C shows the curvature we expected to see from strong stabilization. Several positions (-58, -56, -70, and -74) have regulation profiles that are approximately straight lines, implying that CpxR destabilizes or regulates through weak stabilization, i.e., $(1 + \beta P) \approx 1$.

Figure 6D shows a heatmap for $\log(FC_{max})$ as a function of the regulation parameters α and β . The dashed black lines, which denote α and β equal to 1, divide the map into four quadrants, each quadrant with a specific qualitative regulatory scheme. The top right and bottom left quadrants represent coherent regulation strategies where α and β contribute to activate (top right) or repress (bottom left) gene expression. On the other hand, the top left and bottom right quadrant are incoherent in the sense that α and β have opposing regulatory effects: TFs in these quadrant slow the initiation rate of transcription while stabilizing polymerase at the promoter (top left) or increase the initiation rate of

transcription while destabilizing the polymerases' presence at the promoter (bottom right). The solid white contour in this plot shows where these two effects balance and the net fold change is 1; left of this line represents TFs that repress, and right of this line represents TFs that activate. Contours of constant FC_{max} are drawn as white dashed lines, marking 10-fold increases/decreases in FC_{max} . On this plot, we also show the inferred probability distribution of the parameters α and β for each position in our data. The black points are lower probability, with lighter points representing higher probability values for the α and β parameters. One notable phenomenon is that, for positions with β less than roughly 10, the inference begins to fail for α and β . This results in inference clouds with "tails" that stretch across quadrants and precludes assessment of the mode of regulation (see position -74). The alignment between the inference clouds of these positions and the constant FC_{max} contours, however, shows that, although we make very precise estimations of FC_{max} , the values of α and β are less certain and correlated. This is an unfortunate consequence of the weak stabilization limit in our model resulting in parameter combinations that encompass coherent and incoherent regulatory regimes, which could explain the data well (Transtrum et al., 2015).

At some positions (-54, -50, -80, and -82), we see the incoherent behavior discussed above, where the TF stabilizes polymerase at the promoter and also slows the rate of initiation, essentially serving opposing functions in influencing gene expression. For -54, the net effect of these opposing mechanisms is repression, whereas at -80, -82, and -50, the result is activation. However, the positions with strong activation signatures (-48, -60, and -64) as well as some intermediate ones (-56, -58, and -70) have stabilization and acceleration values that impart a coherent strategy of regulation where RNAP recruitment and acceleration of transcription work together. Interestingly, all but one of the regulatory positions studied here show clear positive stabilization ($\beta > 1$), even the lone repressive position (-54). In our data, upstream regulation by CpxR typically involves stabilizing RNAP, regardless of the net regulatory function (repression or activation). However, the level of acceleration/deceleration varies more significantly between positions from a roughly 20-fold deceleration, which results in overall repression of expression, up to a 25-fold acceleration, which results in strong activation.

Finally, combining the inferred regulatory parameters, α and β , determined through inference in the concentration manifold space, with the measured extrinsic features (TF copy number and binding affinity) of gene regulation produces model curves using the effective parameterization that fit our data well. The effective parameters FC_{max} and χ for the CpxR TF data (see STAR Methods for details) are used to plot the fold change as a function of χN_{TF} for the 13 regulatory positions in Figure 6E. Crucially, we demonstrate that the two key variables, χ and FC_{max} , are effective in capturing the fold change position sweep profiles similar to the plots in Figures 4A and 4B.

DISCUSSION

To build a predictive understanding of gene regulation, we need to understand not just where and when TFs bind but also learn the function and magnitude of the mechanisms of regulation at work by each TF (Weingarten-Gabbay and Segal, 2014; Gertz et al.,

2009; Bylino et al., 2020). Often, it is difficult to separate the contributing factors of regulation, such as the TF binding affinity, copy number, and interactions with other TFs, from the regulatory role of the TF that is characterized by its interactions with RNAP at the promoter (Ross and Gourse, 2009). Here we use a synthetic biology approach to measure the isolated regulatory effect of a TF on an otherwise constitutive promoter. These data are interpreted through a thermodynamic model of gene expression that treats the regulatory role of TFs as a combination of interactions that stabilize (or destabilize) the polymerase at the promoter and interactions that accelerate (or decelerate) the rate of transcription when the TF is cobound with polymerase. The model used here allows both modes simultaneously and, importantly, enables us to quantify TF regulatory function continuously rather than categorically as “activators” or “repressors.” Using this model, we are able to characterize the wide range of regulation we see from the TFs in this study, which ranges from 10,000-fold repression up to 100-fold activation and everything in between with the same model. We believe that this fluid classification of TF function can be a useful tool for characterizing TFs for the purpose of model building and predictive design of gene regulation.

We found that, for TFs operating immediately downstream of the promoter, the regulation of each TF was consistent with strong repression (i.e., with $FC_{max} \approx 0$). Despite the large range in magnitude of regulation at this location, the same intrinsic regulatory mechanism seems to be conserved; differences in the magnitude of regulation were primarily due to differences in TF binding affinities rather than in the fundamental regulatory mechanisms of the TFs. In contrast, when these same TFs bind 61 bp upstream of the promoter, the regulatory function of the TFs varied more substantially. We find that some TFs remained strong repressors (similar to their function at +1), but other TFs only weakly repress expression regardless of TF copy number. We attribute this to intrinsic properties of the TF, polymerase (de)stabilization and (de)acceleration of transcription initiation by the TF, which depend on TF identity. Furthermore, when profiling the regulation of CpxR at upstream binding locations, we find that this TF can regulate multiple steps of the transcriptional process and joins a growing body of evidence for TFs engaging in complex regulatory maneuvers at the promoter (Smith et al., 1996; Monsalve et al., 1997; Jensen et al., 2019). This insight into how activation is actually brought about by the independent contributions of acceleration and stabilization demonstrates the applicability of the model to *in vivo* data and complements previous work investigating the regulatory profile of the CRP TF (Forcier et al., 2018) as well as *in vitro* biochemical and structural considerations probing the kinetics of activation (Rhodius et al., 1997). We find that the contributions of these two mechanisms do not correlate with position; in some locations, we found stabilization and acceleration acting together to produce strong activation, and in other locations, deceleration and stabilization worked incoherently, resulting in weaker activation and repression. A startling feature of this incoherent regulation was its presence in the regulatory response of an activator (CpxR) and a repressor (AgaR) and demonstrates that this type of regulation is not just accessible to TFs but may be a pervasive aspect of TF-RNAP regulation.

The concept of stabilization and acceleration working antagonistically, with the step carrying the larger effect size ultimately determining the status of expression (activation or repression), has implications for the current paradigm of viewing TF regulation, particularly activation in the context of class I and class II promoters. This delineation of promoter

class is based on the type of molecular contacts the activator makes with RNAP (Li et al., 1997; Niu et al., 1996; Lee et al., 2012; Zhou et al., 2014; Savery et al., 2002). For specific TFs, mapping between contacts and regulatory mode has been established (Li et al., 1997; Savery et al., 2002) However, addressing how these contacts shape the relative effects of α and β across a wide range of TFs would fill a vital gap in elucidating the molecular determinants that give rise to coherent and incoherent regulatory regimens. Such information would allow more complete characterization of TFs, and, in conjunction with methods of profiling TFs through genome-wide occupancy techniques, provides an edge in the challenge of uncovering an “expression code”: i.e., a set of rules that govern the magnitude and duration of gene expression from natural promoters (Kinkhabwala and Guet, 2008; Beer and Tavazoie, 2004; Ireland et al., 2020). Realizing this goal requires systematic characterization of position-dependent TF regulatory profiles to determine the spatial landscape of α and β for TFs across different families, with the aim of generating a “regulatory compendium” that classifies TFs according to their regulatory mode. This argument, however, is predicated on our ability to infer regulation driven by stabilization and acceleration with sufficient precision. One unfortunate feature of our experiments, as designed, is our inability to measure destabilization with good precision ($\beta \approx 1$) and limits the characterization of TFs that engage through this mode. In these experiments, FC_{max} acts as a “stiff” parameter that completely determines the expected regulatory outcome, whereas the individual values of α and β are “sloppy” (Machta et al., 2013; Transtrum et al., 2015). One potential way to overcome this is to select a stronger promoter sequence for the target gene. The selected sequence we used in our synthetic circuit was an attenuated form of the *lacUV5* promoter, which was selected to increase the dynamic range of the promoter. Therefore, measuring the isolated regulatory function of TFs may require a range of promoters to fully characterize the wide array of possible regulatory behaviors.

Here we focus entirely on the isolated regulatory role of each TF, but it is clear that one of the next steps is to probe how quantified TFs regulate together. Extending the simple thermodynamic model to incorporate regulation directed by multiple TFs will play a crucial role in untangling elaborate regulatory architectures (Buchler et al., 2003), especially those found in eukaryotes. Indeed, thermodynamic models have been employed to interrogate the regulatory function of eukaryotic promoters involved in key processes such as cellular differentiation, body patterning, and a host of other biological roles (Segal et al., 2006, 2008; Sayal et al., 2016; Chen et al., 2008; Ay and Arnosti, 2011; Fakhouri et al., 2010; Bashor et al., 2019). The ability to distinguish and quantify the different modes of regulation (stabilization and acceleration) and characterize TFs based on them is important for developing general theories of regulation that include multiple TFs that act on different kinetic steps of the transcription process (Scholes et al., 2017; Martinez-Corral et al., 2020; Wong and Gunawardena, 2020); predictions for the combined regulatory effect of two stabilizing TFs should be different than predictions for a stabilizing TF acting together with an accelerating TF (Scholes et al., 2017). With each characterized TF, we can develop an empirical baseline or null hypothesis for what a TF should do on a gene; departures from this expectation, because of emergence of complex regulatory phenomenon brought about by TF-TF interactions (Weingarten-Gabbay and Segal, 2014), allosteric interactions

(Rosenblum et al., 2020; Kim et al., 2013), or other effects indicate surprises that warrant testing in these expanded models.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—All request for information regarding datasets, materials, and reagents as well as questions pertaining to the manuscript should be directed to Robert Brewster (Robert.Brewster@umassmed.edu).

Materials availability—All *E. coli* strains and plasmids generated in this study are available on request by contacting the lead contact.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Microbial strain and culture conditions—*E. coli* strain MG1655 was the base strain used for all synthetic regulatory circuit strains constructed and measured in this work. For a comprehensive list of engineered strains, see Tables S2 and S4. All strains were cultured at 37° C with 250rpm in a shaking incubator in an initial culture of LB and appropriate antibiotics until saturation. Cultures were then diluted 10⁴- to 10⁵-fold into 1 mL of fresh M9 minimal media supplemented with 0.5 percent of glucose at different aTc concentrations. We then assessed when the cultures were at steady state using OD measurements, and performed our quantitative measurements of TF copy number and YFP expression on either microscopy or flow cytometry immediately. More details on these procedures can be found in Culture conditions and Data acquisition procedures for Microscopy and Flow Cytometry Data in the Method details. The comprehensive list of engineered strains are found in Tables S2 and S4.

METHOD DETAILS

Thermodynamic model for single TF regulation—In the work here we use the standard form of the thermodynamic model as derived elsewhere (Ackers et al., 1982; Bintu et al., 2005; Buchler et al., 2003). In our framework, we include an additional consideration of the TF altering the rate of transcription as detailed in Garcia et al. (2012). The partition function for this system is:

$$Z = \left(1 + (N_p/N_{NS})e^{-(\Delta\epsilon_p)} + (N_{TF}/N_{NS})e^{-(\Delta\epsilon_{TF})} + (N_p N_{TF}/N_{NS}^2)e^{-(\Delta\epsilon_p + \Delta\epsilon_{TF} + \Delta\epsilon_I)} \right), \quad (\text{Equation 1})$$

where each term (in order) represents the weight of the unbound, bound by polymerase, bound by TF and cobound state. The terms N_p and N_{TF} are the total number of polymerase or TF molecules, with the term N_{NS} in the denominator scaling the respective terms with the total number of potential binding sites on the genome to give an effective concentration on the chromosome. The energy terms ϵ_p and ϵ_{TF} are the binding affinities of polymerase or TF to their promoter or operator site. The stabilization term (β) as discussed in Figure 1 is represented by the exponentiation of ϵ_I . The probability to find polymerase bound as a function of TF number is then,

$$P_{bound}(N_{TF}) = \frac{\left((N_p/N_{NS})e^{-(\Delta\epsilon_p)} + (N_p N_{TF}/N_{NS}^2)e^{-(\Delta\epsilon_p + \Delta\epsilon_{TF} + \Delta\epsilon_I)} \right)}{Z} \quad (\text{Equation 2})$$

To compare with experimental measurements, we model YFP expression from our synthetic gene circuit as the convolution of the state specific transcription rates and the states enumerated in $P_{bound}(N_{TF})$. For the state in which RNAP is solely bound, we give a rate of expression as r (a course-grained parameter representing the basal rate of YFP production). For the TF-RNAP co-bound state, we assign a scaling factor α that represents the change in transcription rate when the TF is bound (acceleration):

$$YFP_{expression} = r \left(\frac{(N_p/N_{NS})e^{-(\Delta\epsilon_p)}}{Z} + \frac{\alpha(N_p N_{TF}/N_{NS}^2)e^{-(\Delta\epsilon_p + \Delta\epsilon_{TF} + \Delta\epsilon_I)}}{Z} \right) \quad (\text{Equation 3})$$

In practice, what we seek to model is the fold change in gene expression, which is the change in expression level relative to the unregulated gene. Based on the partition function and the state specific transcription rates, the fold change in expression then assumes the following form:

$$\begin{aligned} \text{fold - change} &= \frac{YFP_{expression}(N_{TF} \neq 0)}{YFP_{expression}(N_{TF} = 0)} \\ &= \frac{1 + \alpha(N_{TF}/N_{NS})e^{-(\Delta\epsilon_{TF} + \Delta\epsilon_I)}}{1 + \frac{\left((N_{TF}/N_{NS})e^{-(\Delta\epsilon_{TF})} \right) \left(1 + (N_p/N_{NS})e^{-(\Delta\epsilon_p + \Delta\epsilon_I)} \right)}{\left(1 + (N_p/N_{NS})e^{-(\Delta\epsilon_p)} \right)}} \end{aligned} \quad (\text{Equation 4})$$

We then define the following terms $\beta = e^{-\Delta\epsilon_I}$ and $P = (N_p/N_{NS})e^{-\Delta\epsilon_p}$. As we have measured the value of P in our synthetic circuit to be 6.6×10^{-2} (Brewster et al., 2012), we safely assume the weak promoter limit simplifies the expression $\frac{1 + P\beta}{1 + P} \sim 1 + P\beta$. The fold change is then written as:

$$fold - change = \frac{1 + (N_{TF}/N_{NS})e^{-(\Delta\epsilon_{TF})}\alpha\beta}{1 + (N_{TF}/N_{NS})e^{-(\Delta\epsilon_{TF})}(1 + P\beta)}. \quad (\text{Equation 5})$$

We now define the final term in our derivation: $\chi = (e^{-\Delta\epsilon_{TF}/N_{NS}})(1 + P\beta)$ which represents the effective component that modifies the TF copy number with the product χN_{TF} acting in our model as the effective TF concentration. We now re-write the fold change in terms of the effective TF number (χN_{TF}), and the maximal fold ($FC_{max} = \frac{\alpha\beta}{1 + P\beta}$) as presented in Figure 1 and the main text.

$$fold - change = \frac{1 + FC_{max}\chi N_{TF}}{1 + \chi N_{TF}}. \quad (\text{Equation 6})$$

Choice of the core-promoter sequence used in the synthetic circuit—Previously we have found that the weak promoter approximation describes *in vivo* measurements of repression of the *lacUV5* promoter by LacI (Garcia and Phillips, 2011; Brewster et al., 2014). For this study, where we expect to find both activation and repression, we decided to use a weaker promoter for the target gene. This promoter was designed such that $e^{-\Delta\epsilon_p}$ is roughly $1k_B T$ lower than that of *lacUV5* (Brewster et al., 2012). We have confirmed that the basal expression of this promoter decreases as expected and that regulation follows the same quantitative response to LacI as for *lacUV5* (Brewster et al., 2012). Given that we have previously measured the core-promoter strength used in our synthetic circuit to be at $P = 6.6 \times 10^{-2}$ (Brewster et al., 2012), we expect that the approximation $1 + P \approx 1$ used in deriving the thermodynamic predictions for the fold change in gene regulation is justified in our work. The choice of the promoter sequence comes with its trade-offs: a weaker promoter sequence, while allowing for a larger window of measurement for activation, potentially limits the ability to measure smaller β values for activation (the weaker the promoter, the larger the range of beta that is constrained to measurable dependence with α - note the weak stabilization limit discussed in the Main Text). Taking all points into consideration, we feel that our choice of the promoter sequence adequately balances competing objectives of detecting activation, and inferring the contributions of α and β , and allowing for strong enough expression to measure 100- to 1,000-fold repression at +1.

Concentration manifold derivation—As our primary motivation in this study rests on changing the binding location to explore the regulatory properties of a particular TF, we looked for a way to reformulate the thermodynamic model in such a way as to remove the binding affinity parameter from our consideration (under the assumption that the binding affinity is set primarily by the TF binding sequence which is invariant). This would allow us to infer the intrinsic regulatory features for a TF (the acceleration and stabilization parameters). To do this, consider the following reformulation of the thermodynamic model where the TF concentration is written as a function of the fold change (abbreviated FC below) for a given binding regulatory location (designated by the superscript x or y):

$$N_{TF} = \left(\frac{1}{\varepsilon^{(y)} K^{(y)}} \right) \frac{FC^{(y)} - 1}{FC_{max}^{(y)} - FC^{(y)}} = \left(\frac{1}{\varepsilon^{(x)} K^{(x)}} \right) \frac{FC^{(x)} - 1}{FC_{max}^{(x)} - FC^{(x)}}. \quad (\text{Equation 7})$$

Given that we measure TF abundance, we can essentially couple the fold change in regulation between two different positions by allowing the TF concentration to trace out a manifold that specifies the fold change at positions y as a function of the fold change at position x .

$$FC^{(y)} = \frac{1 + FC_{max}^{(y)} \left(\frac{\varepsilon^{(y)} K^{(y)}}{\varepsilon^{(x)} K^{(x)}} \right) \left(\frac{FC^x - 1}{FC_{max}^{(x)} - FC^{(x)}} \right)}{1 + \left(\frac{\varepsilon^{(y)} K^{(y)}}{\varepsilon^{(x)} K^{(x)}} \right) \left(\frac{FC^{(x)} - 1}{FC_{max}^{(x)} - FC^{(x)}} \right)}. \quad (\text{Equation 8})$$

As in the first section of the Methods, we define $P = (N_p/N_{NS})e^{-\Delta\varepsilon_p}$ and $FC_{max}^p = \frac{\alpha^p \beta^p}{1 + P\beta^p}$ where the superscript p represents the regulatory position. We also introduce two new terms for compactness: $\varepsilon = e^{-\Delta\varepsilon_{TF}}$ and $K^p = 1 + P\beta^p$. Assuming the binding affinity is constant between the two locations ($\varepsilon = \varepsilon^x = \varepsilon^y$) we achieve a reduction in the manifold:

$$FC^{(y)} = \frac{1 + FC_{max}^y \left(\frac{K^{(y)}}{K^{(x)}} \right) \left(\frac{FC^x - 1}{FC_{max}^x - FC^x} \right)}{1 + \left(\frac{K^{(y)}}{K^{(x)}} \right) \left(\frac{FC^x - 1}{FC_{max}^x - FC^x} \right)}. \quad (\text{Equation 9})$$

Based on this reformulation, the key parameters to consider are the acceleration parameters (couched in the FC_{max} term as described in Figure 1) and the stabilization parameters at positions x and y (a total of 4 parameters). In a sense, the ‘‘concentration manifold’’ allows us to remove what we consider to be the extrinsic feature of TF regulation (the binding affinity and TF copy number) from the intrinsic features of the TF regulatory response (the regulatory activity as of the TF on RNAP through stabilization and acceleration). We can further simplify the model by taking into account a judicious binding location for the position x . Taking the +1 binding location, where the assumption of steric hindrance in our thermodynamic model for all the TFs surveyed in this work is justified, we set $\beta^x \sim 0$ (which makes $K = 1$) and $FC_{max}^x = 0$. This leads us to the final form of the concentration manifold for a given regulatory position (as a function of the fold change at the +1 position - $FC^{(+1)}$):

$$FC^{(y)} = \frac{1 + FC_{max}^y (1 + P\beta^{(y)}) \left(\frac{1 - FC^{(+1)}}{FC^{(+1)}} \right)}{1 + (1 + P\beta^{(y)}) \left(\frac{1 - FC^{(+1)}}{FC^{(+1)}} \right)}. \quad (\text{Equation 10})$$

Now we see that the only parameters that remain in the model are the acceleration and stabilization at position y (the regulatory position under consideration).

Culture conditions and data acquisition procedures for microscopy and flow cytometry data—The position dependent regulatory profiles for the 6 TFs -AcrR, AgaR, ArsR, AscG, BetI, CpxR - evaluated at +1 and -61 positions were measured using fluorescence microscopy. At every microscopy session, the TF titration strains (harboring the integrated TF-mCherry fusions in the *ycbN* locus and the synthetic circuit in the *galK* locus) were cultured with companion strains. These include the TF-mCherry fusions lacking the *galK* synthetic circuit integration (necessary to derive the calibration factor to convert the arbitrary fluorescence signal into TF copy number) and TF-mCherry fusion strains with the TF binding site missing from the integrated synthetic circuit (necessary to account for TF titration effect on gene expression). Furthermore, constitutive strains lacking the TF-mCherry fusion (integration in the *ycbN* locus) expressing the integrated synthetic circuit were necessary to compute the fold change in gene expression.

Single colonies of bacterial cultures from freshly streaked LB-Agar plates with appropriate antibiotics are grown overnight in 1 mL of LB in a 37° C incubator shaking at 250 rpm. Cultures are diluted 10⁴- to 10⁵-fold into 1 mL of fresh M9 minimal media supplemented with 0.5 percent of glucose at different aTc concentrations (0, 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, 8, and 10 ng/mL) and allowed to grow at 37° C until they reach an OD600 of 0.1 to 0.2 and harvested for microscopy. 1 μ L of cells is spotted on a 2 percent low melting agarose pad (Invitrogen 16520050) made with 1X PBS. An automated fluorescent microscope (Nikon TI-E) with a heating chamber set at 37°C is used to record multiple fields per sample (between 6–12 unique fields of view) resulting in roughly 100 to 500 individual cells per sample.

The calibration factor for the conversion of mCherry fluorescence to TF copy number is quantified by measuring the fluctuations in fluorescence partitioning during cell division (Brewster et al., 2014). Briefly, cells expressing the TF-mCherry fusion protein are grown as described above, and just before imaging 100 μ L of cells from different aTc concentrations are pooled together and washed twice with M9-glucose minimal media containing no aTc. Cells are then spotted on 2% low melting agarose pad made with M9-glucose minimal media. Phase images are captured for roughly 150 to 200 fields and their positions are saved for later. These phase images (named as Lineage tracker) will serve as a source file for lineage tracking of the mother-daughter pair. After one doubling time (roughly 1 hour or depending on the doubling time for different TF strains), the microscope stage was returned to the same field of view using the saved position matrix and are imaged again (and named as daughter finder) using both phase and mCherry channels

To measure the regulatory profiles for the CpxR TF position sweep constructs, we used flow cytometry for rapid and reproducible data acquisition. The CpxR-titration strains harboring the position regulation plasmids (See Strains) were cultured in LB + Kanamycin media from single colony inoculates until saturation. A 1:10,000 dilution for each strain was then made in M9 minimal media supplemented with glucose along with the appropriate amount of aTc to titrate CpxR-mCherry levels. We found that the following aTc concentrations 0,

0.5, 1, 1.5, 3, 4, 6 and 8 ng/ml provided a good dynamic range of TF expression while maintaining viability of the CpxR strains. The aTc dilution solutions were made from a stock solution of 1 mg/mL suspended in ethanol and were made fresh prior to the application of the aTc for the M9 culture. After M9 dilution, the strains were grown in 96 well plates to steady state (OD₆₀₀ of 0.1 – 0.2). Similar to microscopy acquisition procedure, we had constitutive (CpxR-KO) strains transformed with the binding position plasmids along with the CpxR-titration strain transformed with the plasmid having the TF binding sequence removed from the promoter to account for physiological effects of CpxR-mCherry titration to calculate the Fold-change. Cells were diluted between 1:2 to 1:3 fold in PBS media in a 96 well cytometer plate prior to data acquisition and cytometry was performed on a MacsQuant VYB. At the beginning of each run, an initial gating strategy involving the Forward Scatter and Side Scatter area information was used to eliminate background events and samples were run to achieve ~ 60,000 gated events for each position strain at a given aTc concentration.

Engineering the titratable TF-mCherry fusion strains—All strains used in this study are constructed from the parent strain *E. coli* MG1655. The TFs investigated in this study include AcrR, AgaR, ArsR, AscG, BetI, and CpxR. Each TF gene is deleted from its wild-type locus and expressed from the *ybcN* locus under the regulation of the *P_{tet}* promoter. The autofluorescence strain for each experiment is *E. coli* MG1655 with the corresponding TF knocked out from the wild-type locus. We used the KEIO library (Baba et al., 2006) as the starting point for the construction of the 6 TFs with the titratable TF-mCherry fusion cassette. We selected the corresponding clone from the KEIO library with the TF-knockout (the coding and upstream regions of the TF gene are replaced by a constitutive promoter expressing Kanamycin), and deletion of TF gene from the wild-type locus was performed by P1 transduction of the corresponding knockout from the KEIO collection to the MG1655 *E. coli* strain. The kanamycin cassette was cured using the *frt* flippase expressed from *pCP20* plasmid. Primers listed in Table S1 were then used to amplify the coding sequence (without the stop codon) of the 6 TFs profiled from the MG1655 genome. The amplified coding regions had overhangs for the pTet-AEK-mcherry plasmid that contains the *P_{Tet}* promoter in frame with the flexible AEK linker sequence (GCAGAAGCAGCAGCAAAGGAAGCAGCAGCAAAGGCA) and mCherry, and were subsequently cloned into the plasmid via Gibson Assembly to make the respective pTet-TF-AEK-mCherry plasmids. The pTet-TF-AEK-mCherry fusion cassette were then amplified with *ybcN* integration primers for chromosomal insertion using lambda red recombineering assisted by plasmid *pKM208* as described previously (Murphy and Campellone, 2003). We sequenced the regions surrounding the *P_{Tet}* promoter and the TF-Linker-mCherry cassette to confirm the regions were free of any mutations. For the final step, the *ybcN* locus was moved using P1 transduction to the TF knockout strain harboring a constitutively expressed TetR integrated at the *gspI* locus. These strains (Table S2) allow for inducible control of TF copy number by titrating TetR repression with aTC (anhydrous tetracycline) and were used for all synthetic circuit measurements presented in the Main Text.

Cloning the TF specific binding location gene circuits—The upstream promoter sequence in our synthetic gene circuit was derived from the *P_{DL5}* plasmid that contains a modified version of the *lacUV5* promoter sequence as used previously (Brewster et al., 2014). Binding sequences for AcrR, AgaR, ArsR, AscG, BetI, CpxR TFs listed in Table S3 were cloned at the +1 and the –61 locations (relative to the Transcription Start Site) on the plasmid using Gibson Assembly with primers having the TF binding sequences as 5′ overhangs to the priming sequences of the *P_{DL5}* vector at the respective binding locations. Primers harboring 40bp homology to the *galK* locus were then used to amplify the promoter region of the plasmid for intergration into the locus using the *pKM208* recombination *E. coli* strains (Murphy and Campellone, 2003). The strains were then sequenced and verified to contain the appropriate regulatory and promoter elements before using P1 transduction to transfer the *galK* locus into the appropriate TF titration strains to make the TF inducible, synthetic circuit strains listed in Table S4.

To clone the synthetic target promoters for profiling the regulatory activity CpxR at multiple binding locations, we designed an approach to make fast and efficient cloning of any TF binding sequence at defined locations ranging from +1 to –112bp relative to the TSS on the unregulated DNA circuit (*P_{DL5}*). We designed forward and reverse primers to amplify the *P_{DL5}* plasmid at defined locations in the promoter sequence (see Table S5). These primers were used to insert the *ccdB* cassette at the precise location upstream of the gene circuit and had overhangs for the typeIIIs BbsI restriction site. This allowed for excision of the *ccdB* cassette and cloning of BbsI digested TF binding sequences that had complementary overhangs to the excised region. The *P_{DL5}–ccdB* plasmids (Table S6) were assembled and transformed into the *Escherichia coli* DB3.1 strain that harbors key mutations in DNA gyrase that tolerates the *ccdB* toxin (Bernard, 1996) for stock curation and sequencing. The plasmids were then incubated with double stranded oligos that had the CpxR *ppiA* binding sequence flanked with the BbsI restriction sites and the 4 bp complementary sequence to the digested plasmid. Digestion and ligation of the *ppiA* binding sequence to the *P_{DL5}–ccdB* position plasmids were done in a single incubation step. The cloned *P_{DL5}–ppiA* plasmids in Table S7 were then sequenced and transformed into the CpxR titration strain.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis and statistics—Information pertaining to the data plotted in Figures 2, 3, 4, 5, and 6 including the number of replicates and meaning of the error bars can be found in the captions of the respective figures.

Data processing steps for microscopy and cytometry data—To process the regulatory data for the 6 TFs profiled at +1 and –61, we took the microscopy images and segmented individual cells using a modified version of the MATLAB code Schnitzcells (Rosenfeld et al., 2007). We use this code to segment the phase images of each sample to identify single cells. Mean pixel intensities of YFP and mCherry signals are extracted from the segmented phase mask for each cell. The autofluorescence is calculated by averaging the mean intensity of the autofluorescence strain in both mCherry and yfp channels and is subtracted from each measured YFP or mCherry value. Total fluorescence for each channel is obtained by multiplying the mean pixel-intensity with the area of the cell. Fold change

in expression for a given binding site is calculated by the ratio of total fluorescence of strains expressing the TF to the strains with no TF. For partitioning statistics to estimate the calibration factor, mother-daughter pairs are first automatically identified and verified manually to ensure cells made exactly one division. The mean pixel intensity and area of the mother-daughter pairs are obtained. The background fluorescence is estimated as described previously (Ali et al., 2020) using the inverse mask of individual frames. The sum and difference in fluorescence of the two daughters were then used to find the conversion factor, ν , between fluorescence and number of TFs using the equation $(I_1 - I_2)^2 = \nu(I_1 + I_2)$, which stems from the assumption of binomial partitioning of TFs at cell division (Rosenfeld et al., 2007).

For the CpxR position cytometry data, we adapted a robust data analysis procedure (Razo-Mejia et al., 2018) to computationally gate events to ensure reproducible Fold-change measurements for a given position across replicates. For a given position strain replicate measurement, we collected the data across all the aTC concentrations and proportionally binned the single flow cytometry events into 16 RFP intervals (intervals were of unequal size in RFP space with the number of cells in each bin more or less constant). We then took these binned events and gated them using the \log_{10} values of the Forward Scatter and Side Scatter area profiles for each event (referred to as FSC and SSC respectively) to improve the likelihood that the final retained events were single cell measurements. To construct this gate, we computed the mean and covariance matrix for each dataset for every RFP bin and used these statistics to fit an ellipsoid to the full dataset according to the following formula:

$$\begin{bmatrix} FSC \\ SSC \end{bmatrix}^T \Sigma^{-\frac{1}{2}} \begin{bmatrix} FSC \\ SSC \end{bmatrix} \leq \alpha, \quad (\text{Equation 11})$$

$$\text{with } \Sigma = \begin{bmatrix} \text{Var}(FSC) & \text{Cov}(SSC, FSC) \\ \text{Cov}(FSC, SSC) & \text{Var}(SSC) \end{bmatrix}. \quad (\text{Equation 12})$$

This step retains events that are within a particular distance from the center of the ellipsoid using an appropriate value for the cut-off (α). We based the value of the cut-off on the following rationale: As each event is essentially a vector of \log_{10} values for the FSC (Forward Scatter) and SSC (Side Scatter), we assume the joint values are normally distributed, which translates to a distance metric that is a chi-square random variable (the sum of two normally distributed entities is chi-square with 2 degrees of freedom). We selected α as the 5th percentile of values from the cumulative distribution, and events within the cutoff were taken to be single cell measurements used to compute the fold change values presented in the results section. The resulting gated events for each of the 16 RFP intervals were pooled from all position strains and replicates, and we excluded events with RFP measurements below a certain threshold determined by visually assessing the the fold change profile of the “control” (*DeIBS*) circuit. The YFP signal for these events had large fluctuations and we reasoned that the flow-cytometry approach probably fails in measuring cells at lower TF copy number. The retained events were then binned proportionally into 22 intervals, and the median RFP and Fold change values for each interval was reported

as the representative measurements for that bin. The choice of intervals at this step (with the exception for very large bin sizes) does not seem to appreciably alter the main findings from our inference into the acceleration and stabilization for the CpxR binding locations (see Figures S5 and S6 for details).

Measurement of TF abundance in flow cytometry experiments: Converting the mCherry signal to TF copy number—To measure the regulatory response of the CpxR regulated promoter at 22 binding locations, we used flow-cytometry to measure TF abundance and target gene expression for cells at steady state. Our goal was to look at the regulatory level of the target gene as a function of TF copy number (Brewster et al., 2014), which required TF abundance measurements to be converted from arbitrary fluorescent units to copy number of the TF-mCherry fusion molecules in the cell. To convert the arbitrary mCherry signal from the flow cytometer to TF copy number, we took the flow cytometer measurements of the Fold change response curve as a function of mCherry levels at the +1 position and compared it to the measurements we made using microscopy (see Figure S1). As we show in the Main Text this regulation is consistent with steric hindrance ($FC_{max} = 0, \beta = 0$ (Bintu et al., 2005)). We then extracted the parameter, χ , from both the microscopy and cytometry curves using the thermodynamic model we present in the Main text and compared them. Details on this approach can be found in the STAR Methods (Quantification and statistical analysis) in the subsections Parameter fitting and inference for position dependent fold change regulation data and Using the concentration manifold parameters to generate the thermodynamic model in FC versus N_{TF} space.

We reasoned that while the value inferred for the cytometry curve (χ_{RFP}) and microscopy curve (χ) are different, the relation $\chi_{RFP}RFP = \chi N_{TF}$ will be true according to the thermodynamic model of simple repression:

$$fold - change_{+1} = \frac{1}{1 + \chi_{RFP}RFP} = \frac{1}{1 + \chi N_{TF}}. \quad (\text{Equation 13})$$

This leads to the following interpretation of the quantity $\chi_{RFP}RFP$.

$$\chi_{RFP}RFP = \lambda_{RFP} \frac{1 + P\beta}{1 + P} RFP = \lambda_{mic} \frac{1 + P\beta}{1 + P} \mu RFP = \chi N_{TF}. \quad (\text{Equation 14})$$

Here the parameter λ represents the binding affinity of CpxR with the subscript denoting the units it was measured in. In Figure S1, we show the regulatory curves from the +1 parallel measurements for the cytometry (magenta curve and data points) and microscopy (black curve). The difference between the curves is expected as the units of χ (χ_{RFP} for cytometry and χ_{mic} for microscopy) are different, and the scaling factor μ (in units of TF per RFP signal) is required to make the curves the same. We find $\chi_{RFP} = 2.33 \times 10^{-4}$ (magenta curve) and $\chi = 1.3 \times 10^{-4}$ with the value of $\mu = 1.8$. We used this value of μ and multiplied the RFP signal in the cytometry data to scale the x axis in Figures 5C and 6E.

Parameter fitting and Inference for position dependent fold change regulation

data—We interpret the promoter regulatory data from the 6 TFs surveyed at the +1 and –61 binding locations through the thermodynamic model as specified in the results section. The fold change data for the TF-position strain (See Data processing steps for microscopy and cytometry data section for details) as a function of N_{TF} was fit to Equation 6 with the aim of extracting the best-fit value of the FC_{max} and χ (the product of the stabilization effect and the TF binding affinity). We used a bootstrapping procedure to generate confidence intervals for the both the FC_{max} and binding affinity parameters, and for each of the TFs surveyed we fit the +1 and –61 parameter sets independently. The bootstrapping procedure resampled the data points from the fold change versus RFP curve for a given TF-binding location across all replicate datasets 1000 times. For each iteration, fold change replicate data points from a given induction condition were sampled to generate a possible regulatory response as a function of TF-copy number. Each of these resampled curves were then fit to the thermodynamic model outlined in Figure 1A using a non-linear least-squares fitting procedure to determine the optimal fit for the values of FC_{max} and χ parameters. As seen in Figures 4C and Figures 4D, we report the means and confidence intervals for these two parameters and plot the curve generated by taking the expected value of the thermodynamic model conditioned on the model parameters along with the 95% confidence interval.

For the CpxR position sweep data, we used the concentration manifold formalism to delineate the values of the acceleration and stabilization parameters. For positions that showed discernible regulation (12 out of the 22 upstream positions profiled), we assume that the binding affinity is constant between the regulatory positions as the only changing variable is the binding location (the binding sequence is the same) and recast the binned data from the Fold-change versus RFP replicates using the concentration manifold formalism as detailed in the STAR Methods (see Concentration Manifold Derivation). To sample the probability space of the acceleration and stabilization parameters for a given binding location, we started by inferring the joint posterior distribution FC_{max} and $K = 1 + P\beta$ using a Bayesian approach that relates the parameters underlying our thermodynamic model to the data according to the following relation:

$$p(FC_{max}, K | FC) = \frac{L(FC | FC_{max}, K)p(FC_{max}, K)}{p(FC)}, \quad (\text{Equation 15})$$

where the term on the left hand side is the posterior distribution of the parameters (FC_{max} and K). The terms on the right hand side represent the likelihood of the data given the parameters ($L(FC | FC_{max}, K)$), the prior distribution of the parameters ($p(FC_{max}, K)$), and lastly the distribution of the data itself $p(FC)$. Each of the 12 regulatory positions were fit separately using the Bayesian inference procedure, we specified our likelihood function as a normal distribution of the form:

$$L^K(FC^k | FC_{max}^k, K^k) = \prod_i^{\text{DataPoints}} \text{Normal} \left(FC_i^k | u = FC_{prop}^k(FC_i^{+1}; FC_{max}^k, K^k), sd = \theta^k \right). \quad (\text{Equation 16})$$

The superscript k represents an upstream binding location for the CpxR TF and the subscript i represents the data points for a given position regulatory dataset. The proposed Fold-change value FC_{prop} from the model takes the form:

$$FC_{prop}(FC_i^{+1}; FC_{max}, K) = \frac{1 + FC_{max}K \left(\frac{1 - FC_i^{+1}}{FC_i^{+1}} \right)}{1 + K \left(\frac{1 - FC_i^{+1}}{FC_i^{+1}} \right)}. \quad (\text{Equation 17})$$

The crux of the Bayesian approach to model inference is to simulate candidate draws of the joint posterior distribution of the FC_{max} and K parameters by proposing candidate values from the prior distribution, generating the thermodynamic curve, and evaluating the likelihood function. A transition in the jointly sampled parameter space from one set of parameter values to another is based on the premise that parameter sets will be sampled in proportion to the probability of the posterior distribution (as long as the sampling chain is drawing from the stationary distribution). This process is repeated until a given number of draws have been made from the joint posterior distribution. The results of this inference procedure are used to draw the model curves in Figure 5C and we use the relation between the sampled parameters (FC_{max} and K) and the acceleration and stabilization parameters as detailed previously in the methods section. The sampling procedure was implemented with the PyMC3 package that utilizes the NUTS sampler, a particular implementation of the Hamiltonian Monte-Carlo algorithm, to sample the joint posterior distribution (Hoffman and Gelman, 2014).

We used a uniform distribution as the priors for both the FC_{max} and K model parameters with appropriate bounds for each parameter. For K , we ensured that the lowest potential value is 1, in line with the assumptions from the derivation of the concentration manifold formalism. We checked the inferences from each position to ensure that the bounds we enforced on both parameters were appropriate and that the sampled values were not tending toward the edge of the sample space. Furthermore, we cast the σ parameter (the standard deviation) of the position specific likelihood function as a hyper-parameter in our sampling procedure and set the prior distribution as uniform over a defined interval with the lower bound = 0. Overall, our inference approach allowed us to ensure precise inference of α and β and for 11 out of the 12 regulatory positions (See Table 2).

To get the stabilization values from this inference procedure, we simply used the following relation between the K and β and the fact that the polymerase occupancy has a measured value of 6.65×10^{-2} in our synthetic promoter,

$$\beta = \frac{K - 1}{P}. \quad (\text{Equation 18})$$

Given the stabilization value for a draw in the chain, we then find the corresponding acceleration value using the jointly sampled FC_{max} value and the relation:

$$\alpha = \frac{FC_{max}}{\beta} \left(1 + \left(\frac{P}{1+P} \right) (\beta - 1) \right). \quad (\text{Equation 19})$$

Table 2 lists the inferred acceleration and stabilization parameters. We report the median values of the inference chain along with the bounds that encompass the 68%th percent Bayesian credible interval of the parameters. For position -74 , the inference estimates are not precise due to the phenomenon of the “weak stabilization” limit as discussed in the Main text. Examining the posterior distribution of the FC_{max} and K parameters, we find the values of the 68% credible interval for this positions (both α and β) to encompass both coherent and incoherent regimes.

Inference plots for CpxR position sweep data - posterior distributions for the FC_{max} and K —As detailed in the Methods section, we used a particular formulation of the thermodynamic model (Forcier et al., 2018) that used the TF abundance measurements from cytometry to rewrite the fold change data for a given regulatory position as a function of the fold change at the $+1$ position (see the STAR Methods section on the Concentration Manifold Derivative). The benefit of this re-formulation was the ability to remove the binding affinity parameter and find the “intrinsic” regulatory parameters of the TF (α and β).

To extract α and β for a given regulatory position, we began by inferring the FC_{max} and the $K = 1 + P\beta$ parameters using a Bayesian Sampling approach ((Hoffman and Gelman, 2014)) as detailed in the inference procedure (see STAR Methods Section for details). Figure S2 shows the result of the inference of the posterior distribution for the two parameters for each of the 12 regulatory positions. The first and third column shows the outcome of the Bayesian approach to sampling the posterior distribution for FC_{max} and K parameters, and the second and fourth column presents the resulting transformation into the α and β joint distribution space. The plots are arranged in order from highest β at the top left of the figure to smallest β at the bottom right. As seen, there is tight coupling for the joint distributions between the FC_{max} and the K with an inverse dependence between the two parameters for activation (high values of FC_{max} are sampled jointly with low values of K and vice versa) and opposite for the single repressing position -54 (low values of FC_{max} are sampled jointly with high values of K and vice versa). Arranging them in order of decreasing β serves to demonstrate an important point: the correlation between inference of α and β becomes stronger as β gets smaller (i.e., as K approaches 1). In all cases, we can infer FC_{max} effectively but the individual value of one regulatory parameter depends strongly on the other because when βP is small, $FC_{max} \approx \alpha\beta$ and $K \approx 1$, and thus β and α have a strong inverse relationship. When plotted on log axis this appears as a straight line (of slope -1) and the domain of β from the inference sampling becomes less constrained. For one of the 12 regulatory positions (-74), we are unable to separate α and β with any certainty and the inferred values encompass both incoherent and coherent regulatory outcomes for the appropriate FC_{max} (Transtrum et al., 2015).

Using the concentration manifold parameters to generate the thermodynamic model in fold change versus N_{TF} space—The parameters inferred using the

concentration manifold approach were used to construct the thermodynamic model curves in Figures 5C and 6E. To accomplish this, we used the $K = 1 + P\beta$ and FC_{max} values inferred for each of the 12 regulatory positions (see STAR Methods: Parameter fitting and Inference for position dependent fold change regulation data). To transform the Markov chain of K values to the χ for the main thermodynamic model (Figure 1, and STAR Methods) requires the binding affinity of the TF (the final remaining term in χ), and we inferred the binding affinity from the datasets encompassing all 13 regulatory positions for the CpxR TF (the 12 upstream regulatory positions and the one immediate downstream positions – the +1 position). The binding affinity was treated as a global parameter in our Bayesian inference scheme and was inferred according to the following model:

$$\prod_k^{Position} \prod_j^{Datapoints} Normal(FC_j^k | u = FC_{Thermodynamic(j)}^{(k)}(\lambda), sd = \sigma^{(k)}) \quad (\text{Equation 20})$$

with the mean of the likelihood function specified by the thermodynamic model outlined in the STAR Methods section Thermodynamic model for single TF regulation and takes the following form:

$$FC_{Thermodynamic(j)}^k = \frac{1 + FC_{max}^k \chi^k N_{TF(j)}^k}{1 + \chi^k N_{TF(j)}^k}, \quad (\text{Equation 21})$$

where $\chi^k = \lambda(1/1 + P)K^k$. The parameter λ , is the global parameter representing the scaled binding affinity of the CpxR TF ($\lambda = (1/N_{NS})e^{-\Delta\epsilon_{TF}}$) and is assumed to be constant across the regulatory positions assessed in this work. FC^k is the general thermodynamic model specified in Equation 6 with the object of our inference to infer λ . For the 12 upstream regulatory positions, as the chains of FC_{max} and K were available from the inference of the concentration manifold dataset, we inferred the global λ parameter to the datasets for each of those positions with the values of these two parameters determined from the mean of their respective Markov chains. For the +1 dataset, the parameter FC_{max} was set to 0 as determined in Figure 4A, and the K was set to 1 in keeping with the assumption of steric hindrance. The mean value of the Markov chain λ inferred from this model, along with the scaling factor (see SI section - Converting the mCherry signal to TF copy number) was used to generate the thermodynamic model curves ($mean \pm 2\sigma$) as seen in Figures 5C and 6E.

Robustness of the concentration manifold results—To analyze the fold change data of our experiments we bin the single-cell fluorescence measurements to find the average fold change of cells with similar TF concentrations (mCherry levels). To accomplish this we divided the data into a specified number of bins based on the proportion of the total data points and calculated the ensemble fold change from the cells in each bin. In Figures S5 and S6 we show how the determination of the parameters β and α depends on this choice. In these plots the inferred value of alpha and beta for each of the 12 positions are shown for 14 different bin numbers (6, 8, 10, 12, 14, 16, 18, 20, 24, 26, 28, 30, 32, 36) number of bins and plotted against the value found with 22 bins (used in the main text). The quantity on the y axis is a measurement of the degree to which the values of β (or α) differ from

the reference bin across all regulatory positions, and is computed by taking the mean of the \log_{10} ratios between the reference bin and the bin under consideration for each of the 12 regulatory positions. In Figure S5, we see that the inferred value of β for other bin sizes is tight across most of the regulatory positions with more substantial deviations from 1 as the number of bins becomes very small (<12). This phenomenon is an indication to the degree the larger bin sizes (smaller bin numbers) inadequately convey the curvature inherent in the data, pushing more of the regulatory positions to overestimate the degree of curvature for certain regulatory positions. Figure S6 shows this same measure for the inferred values of α , where we once again see consistency in the inferred value across most of the regulatory positions except for the small number of bins as seen in Figure S5. Crucially, the inference of α and β is not critically sensitive to the choice of bin size above 12 bins.

Testing alternative models of transcriptional regulation for the CpxR position sweep data—In this section, we present an alternate interpretation of the CpxR position sweep data. Specifically we explore if the data can be explained by a simpler model with only one unique regulatory parameter for a given concentration manifold dataset. Specifically, the models we will evaluate in this section will assume the TF operates only through (de)stabilization of RNAP ($\alpha = 1$).

For the model inference runs, we set the energy of the promoter used in our synthetic circuit (DL5 promoter sequence (Brewster et al., 2012)) as a global parameter across the 12 regulatory positions and allowed each position to infer its own stabilization (β) value, with FC_{max} and the K terms as defined in the concentration manifold section:

$$FC^y = \frac{1 + FC_{max}^{(y)} K \left(\frac{1 - FC^{(+1)}}{FC^{(+1)}} \right)}{1 + K \left(\frac{1 - FC^{(+1)}}{FC^{(+1)}} \right)}, \quad (\text{Equation 22})$$

$$FC_{max}^{(y)} = \frac{\beta^{(y)}}{1 + \left(\frac{P}{1+P} \right) (\beta^{(y)} - 1)}, \quad (\text{Equation 23})$$

$$K = 1 + P\beta^{(y)}, \quad 24$$

where $FC^{(y)}$ and $FC_{max}^{(y)}$ represent the fold change and FC_{max} effective parameter when the binding site is introduced at position y on the promoter. Note that FC_{max} no longer has the acceleration parameter. For the Bayesian inference scheme, we set the prior of the DL5 promoter sequence energy as a uniform distribution between the values $-10 k_B T$ and $-2 k_B T$. The sequence energetics have been measured in prior work as $-6.5 k_B T$, as such we believed this was a reasonable interval for the chain to sample. For the position dependent stabilization parameters, we used a uniform distribution with acceptable bounds for all activation positions. We modeled the likelihood function as product of normal distributions (assuming the global likelihood function is a product of the individual position specific likelihood functions) with the mean of these position specific distributions as the theoretical

fold change value generated from the thermodynamic model conditioned on the parameters. To ensure that differences in the fold change profiles between strong activation positions and weak activation positions was adequately conveyed in the likelihood function, we made the standard deviation for the respective position specific normal distributions a hyperparameter in our model, with the final form of the likelihood function as follows:

$$\prod_i^{\text{Position Data Points}} \prod_j^{\text{Points}} \text{Normal}(FC_j^i | u = FC_{\text{stabilization}(j)}^{(i)}(P, \beta^i), sd = \sigma^{(i)}), \quad (\text{Equation 25})$$

and:

$$FC_{\text{stabilization}(j)}^{(i)} = \frac{1 + FC_{\text{max}}^{(i)}(P, \beta^{(i)})K(P, \beta^{(i)})\left(\frac{1 - FC_j^{(+1)}}{FC_j^{(+1)}}\right)}{1 + K(P, \beta^{(i)})\left(\frac{1 - FC_j^{(+1)}}{FC_j^{(+1)}}\right)}. \quad (\text{Equation 26})$$

We initialized the inference procedure as sampling from a global vector, θ , that contained the global and position specific parameters in our model,

$$\theta = [P^{\text{energy}}, \beta^{-48}, \beta^{-50}, \beta^{-54}, \dots, \beta^{-80}, \beta^{-82}]. \quad (\text{Equation 27})$$

The inference procedure was run for 50000 runs initialized on 4 different chains to ensure adequate sampling of the joint parameter space. The fits for the “stabilization only” model are shown in Figure S3 with the values of the 68%th percent Bayesian credible interval reported in Table S8. In this figure, we plot the results of the thermodynamic model from the inference sampling. It is clear that the stabilization only model fails in capturing the highest activation position (−64) and the curvature seen in some positions (−50, −54). This failure to explain strong activation is expected, as in a model without acceleration ($\alpha = 1$) the maximum possible fold change is constrained by the individual occupancy of the promoter by RNAP (the maximal possible fold change is roughly $1/P$ for weak promoters or more precisely $(1 + P)/P$ if the weak promoter assumption is lifted (Phillips and Milo, 2009)); intuitively, in this model without acceleration if the constitutive promoter has polymerase occupancy 10% of the time, the largest fold change possible is 10 (corresponding to 100% occupancy). To reach the fold change values obtained in the CpxR position sweep data, the inferred value of P needs to be at least ~ 10 fold lower than the value of P measured in previous work (Brewster et al., 2012). As such, we feel confident that this model can not sufficiently describe our data.

It is worth noting that even if this assumption is incorrect and in reality P is much smaller than we expect from previous measurements, the model without acceleration still does not describe the data well; the theory does not match the curvature of the data seen in some positions such as −50, −54, and −60 (Figure S3). This feature highlights the importance of acceleration (α) in explaining these positions and the regulation data at large.

Physiological effects of TF titration on synthetic circuit expression—One concern for the gene expression measurements was separating the fundamental regulatory role of a TF from the apparent expression changes due to potential physiological effects such as slowing growth rates from, for instance, high inducer concentrations or changing TF concentration in the cell (Berthoumieux et al., 2013; Klumpp and Hwa, 2014; Keren et al., 2016) brought about by inducing the TFs to different levels. Increasing concentrations of the TF in the cell could potentially alter YFP expression by turning on or off genes involved in global regulation of translation or through a host of post-transcriptional events. In Figure S4, we do not see a major perturbation in synthetic circuit expression for most of the TF titration strains where YFP expression hovers at the FC=1 line as the TF concentration increases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We wish to thank Zufikar Ali, Mike Lee, and Amir Mitchell for stimulating discussions. We thank Marian Walhout, Job Dekker, and Rob Phillips for helpful comments on the manuscript. The research reported in this publication was supported by NIGMS of the National Institutes of Health under award R35GM128797.

REFERENCES

- Ackers GK, Johnson AD, and Shea MA (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl. Acad. Sci. USA* 79, 1129–1133. [PubMed: 6461856]
- Ali MZ, Parisutham V, Choubey S, and Brewster RC (2020). Inherent regulatory asymmetry emanating from network architecture in a prevalent autoregulatory motif. *eLife* 9, e56517. [PubMed: 32808926]
- Amit R, Garcia HG, Phillips R, and Fraser SE (2011). Building enhancers from the ground up: a synthetic biology approach. *Cell* 146, 105–118. [PubMed: 21729783]
- Anderson JC, Voigt CA, and Arkin AP (2007). Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* 3, 133. [PubMed: 17700541]
- Atkinson MR, Savageau MA, Myers JT, and Ninfa AJ (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* 113, 597–607. [PubMed: 12787501]
- Ay A, and Arnosti DN (2011). Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit. Rev. Biochem. Mol. Biol.* 46, 137–151. [PubMed: 21417596]
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, and Mori H (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008.
- Bashor CJ, Patel N, Choubey S, Beyzavi A, Kondev J, Collins JJ, and Khalil AS (2019). Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science* 364, 593–597. [PubMed: 31000590]
- Bateman E (1998). Autoregulation of eukaryotic transcription factors. *Prog. Nucleic Acid Res. Mol. Biol.* 60, 133–168. [PubMed: 9594574]
- Beer MA, and Tavazoie S (2004). Predicting gene expression from sequence. *Cell* 117, 185–198. [PubMed: 15084257]
- Bernard P (1996). Positive selection of recombinant DNA by CcdB. *Biotechniques* 21, 320–323. [PubMed: 8862819]
- Berthoumieux S, de Jong H, Baptist G, Pinel C, Ranquet C, Ropers D, and Geiselmann J (2013). Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.* 9, 634. [PubMed: 23340840]

- Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, and Phillips R (2005). Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* 15, 116–124. [PubMed: 15797194]
- Boyaci H, Chen J, Jansen R, Darst SA, and Campbell EA (2019). Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* 565, 382–385. [PubMed: 30626968]
- Brewster RC, Jones DL, and Phillips R (2012). Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. *PLoS Comput. Biol.* 8, e1002811. [PubMed: 23271961]
- Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, and Phillips R (2014). The transcription factor titration effect dictates level of gene expression. *Cell* 156, 1312–1323. [PubMed: 24612990]
- Brunwasser-Meirom M, Pollak Y, Goldberg S, Levy L, Atar O, and Amit R (2016). Using synthetic bacterial enhancers to reveal a looping-based mechanism for quenching-like repression. *Nat. Commun.* 7, 10407. [PubMed: 26832446]
- Buchler NE, Gerland U, and Hwa T (2003). On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. USA* 100, 5136–5141. [PubMed: 12702751]
- Bylino OV, Ibragimov AN, and Shidlovskii YV (2020). Evolution of Regulated Transcription. *Cells* 9, 1675.
- Chen CC, Zhu XG, and Zhong S (2008). Selection of thermodynamic models for combinatorial control of multiple transcription factors in early differentiation of embryonic stem cells. *BMC Genomics* 9 (Suppl 1), S18.
- Daber R, Sochor MA, and Lewis M (2011). Thermodynamic analysis of mutant lac repressors. *J. Mol. Biol.* 409, 76–87. [PubMed: 21459098]
- Deng W, Li C, and Xie J (2013). The underlying mechanism of bacterial TetR/AcrR family transcriptional repressors. *Cell. Signal.* 25, 1608–1613. [PubMed: 23602932]
- DiGiuseppe PA, and Silhavy TJ (2003). Signal detection and target gene induction by the CpxRA two-component system. *J. Bacteriol.* 185, 2432–2440. [PubMed: 12670966]
- Egan SM (2002). Growing repertoire of AraC/XylS activators. *J. Bacteriol.* 184, 5529–5532. [PubMed: 12270809]
- Egan SM, Pease AJ, Lang J, Li X, Rao V, Gillette WK, Ruiz R, Ramos JL, and Wolf RE Jr. (2000). Transcription activation by a variety of AraC/XylS family activators does not depend on the class II-specific activation determinant in the N-terminal domain of the RNA polymerase alpha subunit. *J. Bacteriol.* 182, 7075–7077. [PubMed: 11092872]
- Elowitz MB, and Leibler S (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. [PubMed: 10659856]
- Fakhouri WD, Ay A, Sayal R, Dresch J, Dayringer E, and Arnosti DN (2010). Deciphering a transcriptional regulatory code: modeling short-range repression in the *Drosophila* embryo. *Mol. Syst. Biol.* 6, 341. [PubMed: 20087339]
- Feklistov A (2013). RNA polymerase: in search of promoters. *Ann. N Y Acad. Sci.* 1293, 25–32. [PubMed: 23855603]
- Feklistov A, Bae B, Hauver J, Lass-Napiorkowska A, Kalesse M, Glaus F, Altmann KH, Heyduk T, Landick R, and Darst SA (2017). RNA polymerase motions during promoter melting. *Science* 356, 863–866. [PubMed: 28546214]
- Feng Y, Zhang Y, and Ebricht RH (2016). Structural basis of transcription activation. *Science* 352, 1330–1333. [PubMed: 27284196]
- Forcier TL, Ayaz A, Gill MS, Jones D, Phillips R, and Kinney JB (2018). Measuring cis-regulatory energetics in living cells using allelic manifolds. *eLife* 7, e40618. [PubMed: 30570483]
- Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeida D, Muñiz-Rascado L, García-Sotelo JS, Alquicira-Hernández K, Martínez-Flores I, Pannier L, Castro-Mondragón JA, et al. (2016). RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.* 44 (D1), D133–D143. [PubMed: 26527724]
- Garcia HG, and Phillips R (2011). Quantitative dissection of the simple repression input-output function. *Proc. Natl. Acad. Sci. USA* 108, 12173–12178. [PubMed: 21730194]
- Garcia HG, Sanchez A, Boedicker JQ, Osborne M, Gelles J, Kondev J, and Phillips R (2012). Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep.* 2, 150–161. [PubMed: 22840405]

- Gardner TS, Cantor CR, and Collins JJ (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342. [PubMed: 10659857]
- Gaston K, Kolb A, and Busby S (1989). Binding of the *Escherichia coli* cyclic AMP receptor protein to DNA fragments containing consensus nucleotide sequences. *Biochem. J.* 261, 649–653. [PubMed: 2673223]
- Gertz J, Siggia ED, and Cohen BA (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* 457, 215–218. [PubMed: 19029883]
- Gu R, Li M, Su CC, Long F, Routh MD, Yang F, McDermott G, and Yu EW (2008). Conformational change of the AcrR regulator reveals a possible mechanism of induction. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 64, 584–588.
- Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, and Collins JJ (2006). A bottom-up approach to gene regulation. *Nature* 439, 856–860. [PubMed: 16482159]
- Hammar P, Walldén M, Fange D, Persson F, Baltekin O, Ullman G, Leroy P, and Elf J (2014). Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nat. Genet.* 46, 405–408. [PubMed: 24562187]
- Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, Ruff EF, Bieter L, Kraft JE, and Record MT Jr. (2017). Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc. Natl. Acad. Sci. USA* 114, E3032–E3040. [PubMed: 28348246]
- Hews CL, Cho T, Rowley G, and Raivio TL (2019). Maintaining Integrity Under Stress: Envelope Stress Response Regulation of Pathogenesis in Gram-Negative Bacteria. *Front. Cell. Infect. Microbiol.* 9, 313. [PubMed: 31552196]
- Hoffman MD, and Gelman A (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* 15, 1351–1381.
- Ireland WT, Beeler SM, Flores-Bautista E, McCarty NS, Röschinger T, Belliveau NM, Sweredoski MJ, Moradian A, Kinney JB, and Phillips R (2020). Deciphering the regulatory genome of *Escherichia coli*, one hundred promoters at a time. *eLife* 9, e55308. [PubMed: 32955440]
- Ishida Y, Kori A, and Ishihama A (2009). Participation of regulator AscG of the beta-glucoside utilization operon in regulation of the propionate catabolism operon. *J. Bacteriol.* 191, 6136–6144. [PubMed: 19633077]
- Jacob F, and Monod J (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356. [PubMed: 13718526]
- Jensen D, and Galburt EA (2021). The Context-Dependent Influence of Promoter Sequence Motifs on Transcription Initiation Kinetics and Regulation. *J. Bacteriol.* 203, e00512–20. [PubMed: 33139481]
- Jensen D, Manzano AR, Rammohan J, Stallings CL, and Galburt EA (2019). CarD and RbpA modify the kinetics of initial transcription and slow promoter escape of the *Mycobacterium tuberculosis* RNA polymerase. *Nucleic Acids Res.* 47, 6685–6698. [PubMed: 31127308]
- Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I., Alisar H, Kaminski S, Weinberger A, Alon U, Milo R, and Segal E (2016). Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* 166, 1282–1294.e18. [PubMed: 27545349]
- Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, et al. (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 45 (D1), D543–D550. [PubMed: 27899573]
- Kim S, Broströmer E, Xing D, Jin J, Chong S, Ge H, Wang S, Gu C, Yang L, Gao YQ, et al. (2013). Probing allostery through DNA. *Science* 339, 816–819. [PubMed: 23413354]
- Kinkhabwala A, and Guet CC (2008). Uncovering cis regulatory codes using synthetic promoter shuffling. *PLoS ONE* 3, e2030. [PubMed: 18446205]
- Kinney JB, Murugan A, Callan CG Jr., and Cox EC (2010). Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci. USA* 107, 9158–9163. [PubMed: 20439748]
- Klumpp S, and Hwa T (2014). Bacterial growth: global effects on gene expression, growth feedback and proteome partition. *Curr. Opin. Biotechnol.* 28, 96–102. [PubMed: 24495512]

- Klumpp S, Zhang Z, and Hwa T (2009). Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139, 1366–1375. [PubMed: 20064380]
- Kolb A, Busby S, Buc H, Garges S, and Adhya S (1993). Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* 62, 749–795. [PubMed: 8394684]
- Kontur WS, Saecker RM, Capp MW, and Record MT Jr. (2008). Late steps in the formation of *E. coli* RNA polymerase-lambda P R promoter open complexes: characterization of conformational changes by rapid [perturbant] upshift experiments. *J. Mol. Biol.* 376, 1034–1047. [PubMed: 18191943]
- Kuhlman T, Zhang Z, Saier MHJ Jr., and Hwa T (2007). Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 104, 6043–6048. [PubMed: 17376875]
- Lamark T, Kaasen I, Eshoo MW, Falkenberg P, McDougall J, and Strøm AR (1991). DNA sequence and analysis of the bet genes encoding the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. *Mol. Microbiol.* 5, 1049–1064. [PubMed: 1956285]
- Lamark T, Røkenes TP, McDougall J, and Strøm AR (1996). The complex bet promoters of *Escherichia coli*: regulation by oxygen (ArcA), choline (BetI), and osmotic stress. *J. Bacteriol.* 178, 1655–1662. [PubMed: 8626294]
- Lee TH, and Maheshri N (2012). A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol. Syst. Biol.* 8, 576. [PubMed: 22453733]
- Lee DJ, Minchin SD, and Busby SJ (2012). Activating transcription in bacteria. *Annu. Rev. Microbiol.* 66, 125–152. [PubMed: 22726217]
- Leyn SA, Gao F, Yang C, and Rodionov DA (2012). N-acetylgalactosamine utilization pathway and regulon in proteobacteria: genomic reconstruction and experimental characterization in *Shewanella*. *J. Biol. Chem.* 287, 28047–28056. [PubMed: 22711537]
- Li M, McClure WR, and Susskind MM (1997). Changing the mechanism of transcriptional activation by phage lambda repressor. *Proc. Natl. Acad. Sci. USA* 94, 3691–3696. [PubMed: 9108039]
- Liu B, Hong C, Huang RK, Yu Z, and Steitz TA (2017). Structural basis of bacterial transcription activation. *Science* 358, 947–951. [PubMed: 29146813]
- Lobell RB, and Schleif RF (1990). DNA looping and unlooping by AraC protein. *Science* 250, 528–532. [PubMed: 2237403]
- Machta BB, Chachra R, Transtrum MK, and Sethna JP (2013). Parameter space compression underlies emergent theories and predictive models. *Science* 342, 604–607. [PubMed: 24179222]
- Martinez-Corral R, Park M, Biette K, Friedrich D, Scholes C, Khalil AS, Gunawardena J, and DePace AH (2020). Transcriptional kinetic synergy: a complex landscape revealed by integrating modelling and synthetic biology. *bioRxiv*. 10.1101/2020.08.31.276261.
- Messeguer X, Escudero R, Farré D, Núñez O, Martínez J, and Albà MM (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 18, 333–334. [PubMed: 11847087]
- Monsalve M, Calles B, Mencía M, Salas M, and Rojo F (1997). Transcription activation or repression by phage psi 29 protein p4 depends on the strength of the RNA polymerase-promoter interactions. *Mol. Cell* 1, 99–107. [PubMed: 9659907]
- Müller J, Oehler S, and Müller-Hill B (1996). Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *J. Mol. Biol.* 257, 21–29. [PubMed: 8632456]
- Murphy KC, and Campellone KG (2003). Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli*. *BMC Mol. Biol.* 4, 11. [PubMed: 14672541]
- Nielsen AA, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, and Voigt CA (2016). Genetic circuit design automation. *Science* 352, aac7341. [PubMed: 27034378]
- Niu W, Kim Y, Tau G, Heyduk T, and Ebright RH (1996). Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell* 87, 1123–1134. [PubMed: 8978616]
- Oehler S, Eismann ER, Krämer H, and Müller-Hill B (1990). The three operators of the lac operon cooperate in repression. *EMBO J.* 9, 973–979. [PubMed: 2182324]
- Phillips R, and Milo R (2009). A feeling for the numbers in biology. *Proc. Natl. Acad. Sci. USA* 106, 21465–21471. [PubMed: 20018695]

- Phillips R, Belliveau NM, Chure G, Garcia HG, Razo-Mejia M, and Scholes C (2019). Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. *Annu. Rev. Biophys.* 48, 121–163. [PubMed: 31084583]
- Ptashne M (2004). *A genetic switch: phage lambda revisited*, Third Edition (Cold Spring Harbor Laboratory Press).
- Ptashne M, and Gann A (1998). Imposing specificity by localization: mechanism and evolvability. *Curr. Biol.* 8, R812–R822. [PubMed: 9818164]
- Ptashne M, and Gann A (2002). *Genes and Signals* (Cold Spring Harbor Laboratory Press).
- Ptashne M, Jeffrey A, Johnson AD, Maurer R, Meyer BJ, Pabo CO, Roberts TM, and Sauer RT (1980). How the lambda repressor and cro work. *Cell* 19, 1–11. [PubMed: 6444544]
- Ramos JL, Martínez-Bueno M, Molina-Henares AJ, Terán W, Watanabe K, Zhang X, Gallegos MT, Brennan R, and Tobes R (2005). The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* 69, 326–356. [PubMed: 15944459]
- Ray WK, and Larson TJ (2004). Application of AgaR repressor and dominant repressor variants for verification of a gene cluster involved in N-acetylgalactosamine metabolism in *Escherichia coli* K-12. *Mol. Microbiol.* 51, 813–826. [PubMed: 14731281]
- Razo-Mejia M, Barnes SL, Belliveau NM, Chure G, Einav T, Lewis M, and Phillips R (2018). Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. *Cell Syst.* 6, 456–469.e10. [PubMed: 29574055]
- Ren S, Li Q, Xie L, and Xie J (2017). Molecular Mechanisms Underlying the Function Diversity of ArsR Family Metalloregulator. *Crit. Rev. Eukaryot. Gene Expr.* 27, 19–35. [PubMed: 28436329]
- Revyakin A, Liu C, Ebright RH, and Strick TR (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314, 1139–1143. [PubMed: 17110577]
- Rhodium VA, and Busby SJ (2000). Transcription activation by the *Escherichia coli* cyclic AMP receptor protein: determinants within activating region 3. *J. Mol. Biol.* 299, 295–310. [PubMed: 10860739]
- Rhodium VA, West DM, Webster CL, Busby SJ, and Savery NJ (1997). Transcription activation at class II CRP-dependent promoters: the role of different activating regions. *Nucleic Acids Res.* 25, 326–332. [PubMed: 9016561]
- Rolland T, Tasxan M, Charlotiaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, et al. (2014). A proteome-scale map of the human interactome network. *Cell* 159, 1212–1226. [PubMed: 25416956]
- Rosenblum G, Elad N, Rozenberg H, Wiggers F, and Hofmann H (2020). Allostery through DNA drives phenotype switching. *Nat. Commun.* 12, 2967.
- Rosenfeld N, Young JW, Alon U, Swain PS, and Elowitz MB (2007). Accurate prediction of gene feedback circuit behavior from component properties. *Mol. Syst. Biol.* 3, 143. [PubMed: 18004276]
- Ross W, and Gourse RL (2009). Analysis of RNA polymerase-promoter complex formation. *Methods* 47, 13–24. [PubMed: 18952176]
- Roy S, Garges S, and Adhya S (1998). Activation and repression of transcription by differential contact: two sides of a coin. *J. Biol. Chem.* 273, 14059–14062. [PubMed: 9603899]
- Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muñoz-Rascado L, García-Sotelo JS, Weiss V, Solano-Lira H, Martínez-Flores I, Medina-Rivera A, et al. (2013). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* 41, D203–D213. [PubMed: 23203884]
- Savery NJ, Lloyd GS, Busby SJ, Thomas MS, Ebright RH, and Gourse RL (2002). Determinants of the C-terminal domain of the *Escherichia coli* RNA polymerase alpha subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters. *J. Bacteriol.* 184, 2273–2280. [PubMed: 11914359]
- Sayal R, Dresch JM, Pushel I, Taylor BR, and Arnosti DN (2016). Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early *Drosophila* embryo. *eLife* 5, e08445. [PubMed: 27152947]
- Schleif R (2010). AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* 34, 779–796. [PubMed: 20491933]

- Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, and Heinemann M (2016). The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* 34, 104–110. [PubMed: 26641532]
- Scholes C, DePace AH, and Sánchez Á (2017). Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle. *Cell Syst.* 4, 97–108.e9. [PubMed: 28041762]
- Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang JP, and Widom J (2006). A genomic code for nucleosome positioning. *Nature* 442, 772–778. [PubMed: 16862119]
- Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, and Gaul U (2008). Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 451, 535–540. [PubMed: 18172436]
- Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, Keren L, Yakhini Z, Weinberger A, and Segal E (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* 30, 521–530. [PubMed: 22609971]
- Shea MA, and Ackers GK (1985). The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J. Mol. Biol.* 181, 211–230. [PubMed: 3157005]
- Shen-Orr SS, Milo R, Mangan S, and Alon U (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31, 64–68. [PubMed: 11967538]
- Smith SB, Cui Y, and Bustamante C (1996). Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 271, 795–799. [PubMed: 8628994]
- Sprinzak D, and Elowitz MB (2005). Reconstruction of genetic circuits. *Nature* 438, 443–448. [PubMed: 16306982]
- Stanton BC, Nielsen AA, Tamsir A, Clancy K, Peterson T, and Voigt CA (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* 10, 99–105. [PubMed: 24316737]
- Stormo GD (2000). DNA binding sites: representation and discovery. *Bioinformatics* 16, 16–23. [PubMed: 10812473]
- Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, and Hasty J (2008). A fast, robust and tunable synthetic gene oscillator. *Nature* 456, 516–519. [PubMed: 18971928]
- Tan C, Marguet P, and You L (2009). Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* 5, 842–848. [PubMed: 19801994]
- Teng SW, Wang Y, Tu KC, Long T, Mehta P, Wingreen NS, Bassler BL, and Ong NP (2010). Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. *Biophys. J.* 98, 2024–2031. [PubMed: 20441767]
- Transtrum MK, Machta BB, Brown KS, Daniels BC, Myers CR, and Sethna JP (2015). Perspective: Slowness and emergent theories in physics, biology, and beyond. *J. Chem. Phys.* 143, 010901. [PubMed: 26156455]
- Ushida C, and Aiba H (1990). Helical phase dependent action of CRP: effect of the distance between the CRP site and the –35 region on promoter activity. *Nucleic Acids Res.* 18, 6325–6330. [PubMed: 2173826]
- Vidal M, Cusick ME, and Barabási AL (2011). Interactome networks and human disease. *Cell* 144, 986–998. [PubMed: 21414488]
- Vilar JM, and Leibler S (2003). DNA looping and physical constraints on transcription regulation. *J. Mol. Biol.* 331, 981–989. [PubMed: 12927535]
- Vilar JMG, and Saiz L (2013). Reliable prediction of complex phenotypes from a modular design in free energy space: an extensive exploration of the lac operon. *ACS Synth. Biol.* 2, 576–586. [PubMed: 23654358]
- Wasserman WW, and Sandelin A (2004). Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* 5, 276–287. [PubMed: 15131651]
- Weingarten-Gabbay S, and Segal E (2014). The grammar of transcriptional regulation. *Hum. Genet.* 133, 701–711. [PubMed: 24390306]

- Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443. [PubMed: 25215497]
- Wong F, and Gunawardena J (2020). Gene Regulation in and out of Equilibrium. *Annu. Rev. Biophys.* 49, 199–226. [PubMed: 32375018]
- Young BA, Gruber TM, and Gross CA (2002). Views of transcription initiation. *Cell* 109, 417–420. [PubMed: 12086598]
- Zhou Y, Kolb A, Busby SJ, and Wang YP (2014). Spacing requirements for Class I transcription activation in bacteria are set by promoter elements. *Nucleic Acids Res.* 42, 9209–9216. [PubMed: 25034698]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- Synthetic biology approach to dissect the fundamental role of TFs
- Model infers the role of two regulatory mechanisms: acceleration and stabilization
- TFs use different degrees of each mechanism based on binding location

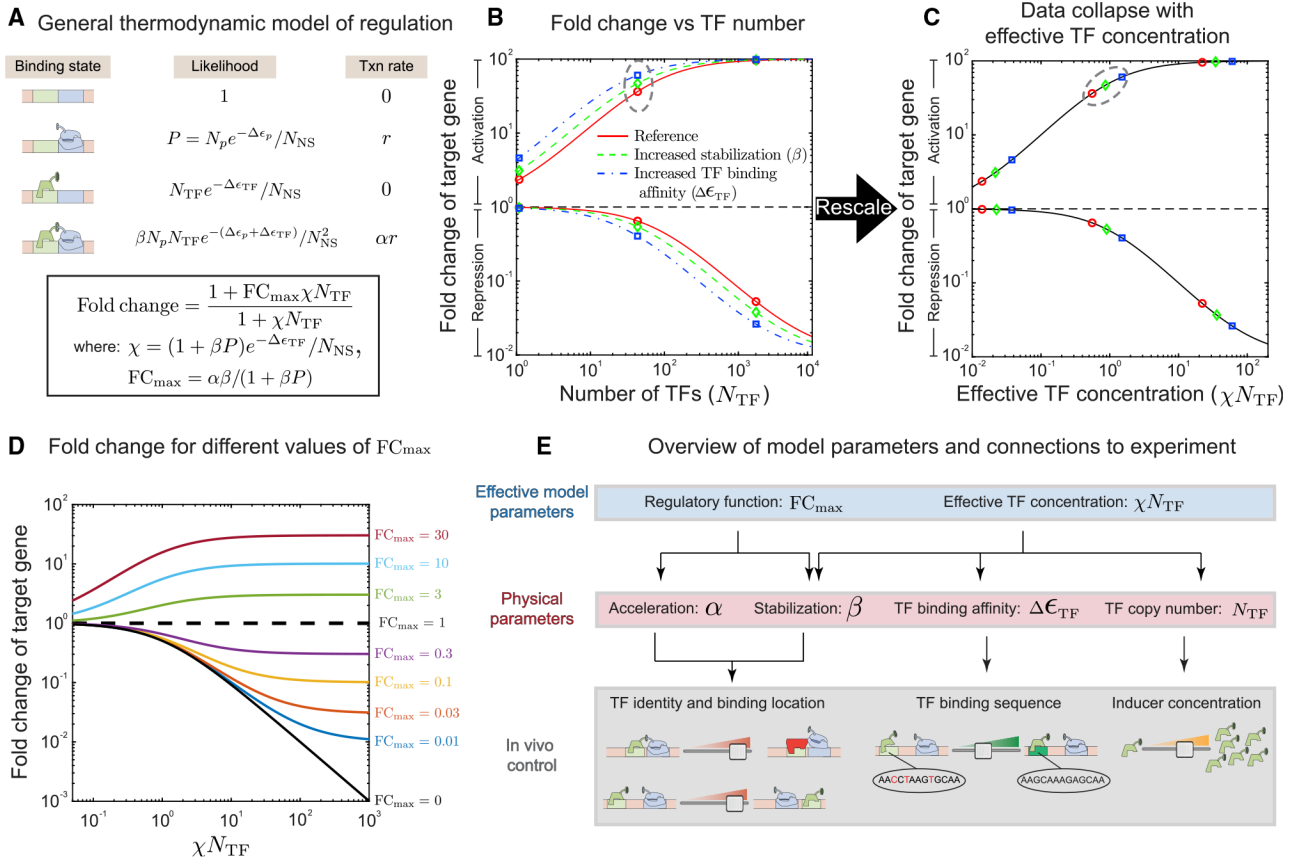


Figure 1. Thermodynamic modeling for measuring TF regulatory features

(A) The thermodynamic model for the target gene allows for four states: unbound by TF or RNAP, bound by RNAP, bound by TF, or bound by both. The probability of each of these states occurring is listed in the center column. The rightmost column shows the transcription rates in these states.

(B) Fold change versus TF copy number (N_{TF}) for a gene regulated by an activator (top set of curves) or a repressor (bottom set of curves). The blue and green curves have the same FC_{max} as the red curve but with increased stability (β , green curve) or TF binding affinity (ϵ_{TF} , blue curve).

(C) Replotting the curves in (B) as a function of effective TF concentration (χN_{TF}) demonstrates that each of the curves now falls onto a single “collapsed” curve defined by the effective parameters FC_{max} and χN_{TF} .

(D) Data collapse curve for a range of different FC_{max} values.

(E) The relationship between theoretical parameters of the model and *in vivo* molecular details of the regulation.

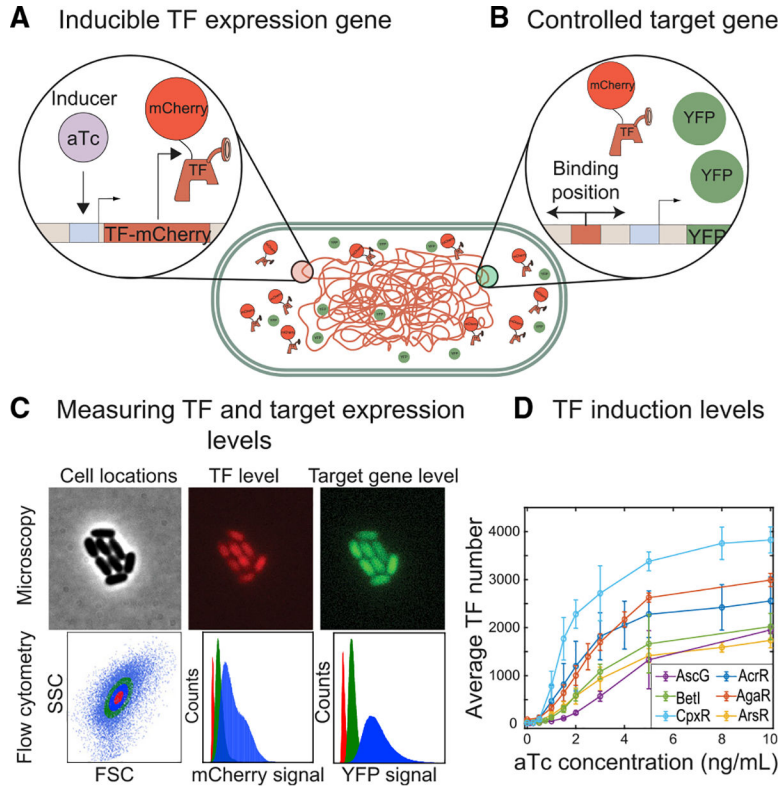


Figure 2. Experimental setup for measuring the TF position-dependent regulatory profile
 (A) The inducible TF expression strains consist of a set of base strains where the endogenous copy of any one of our 6 TFs is knocked out and reintroduced as a TF-mCherry fusion at the *ycbN* locus expressed from an inducible *tet* promoter.
 (B) Regulation by the controlled TF is measured using a synthetic target promoter driving YFP expression integrated to the *galK* locus. The target promoter is designed to be unregulated except by a single binding site for the controlled TF. The sequence and location of this binding site can be controlled systematically.
 (C) The quantitative regulation is measured as the fold change in YFP expression as a function of mCherry signal.
 (D) The range of TF concentrations explored for each TF is shown. Data points for each TF represent the mean number of TFs across 3 replicate measurements ($n = 3$), with the error bars representing the standard error of the mean. Here we converted the arbitrary mCherry fluorescence signal into number of TFs using a fluctuation counting method detailed further in the STAR Methods.

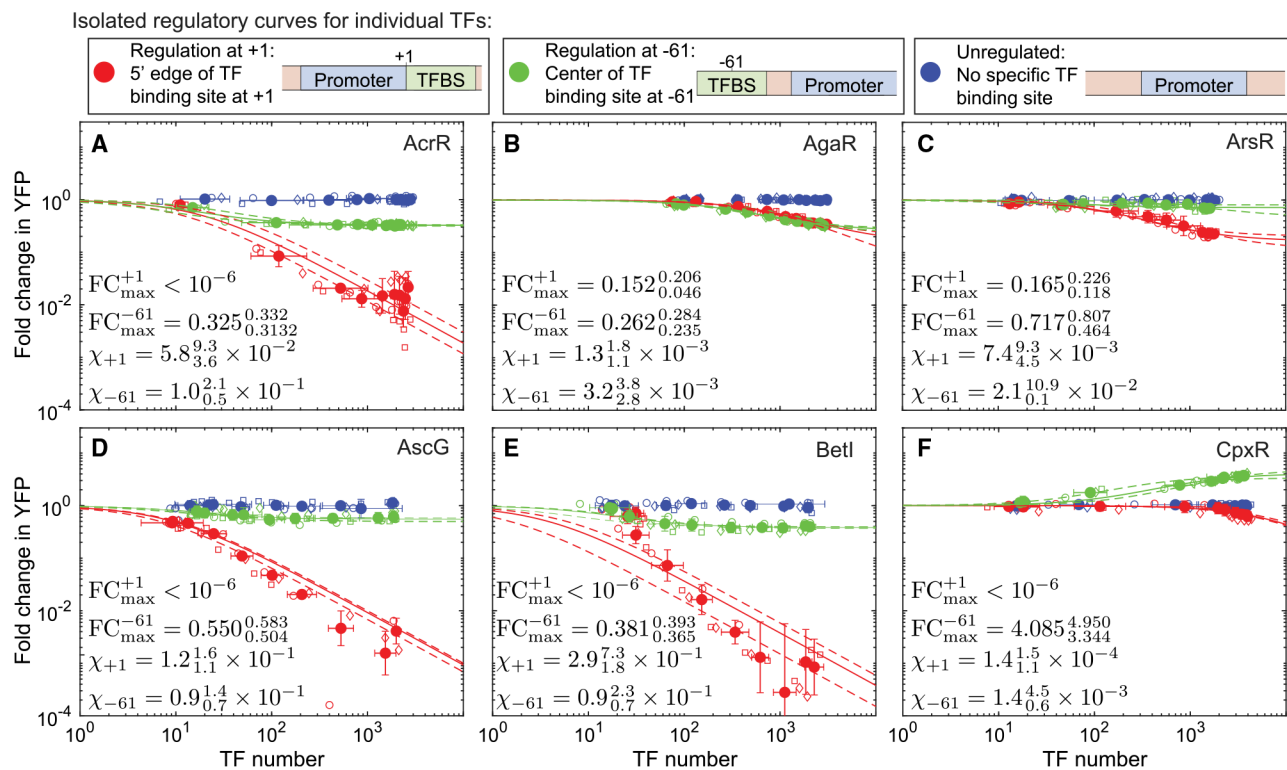


Figure 3. Regulatory curves for individual TFs

(A–F) Each curve shows the response of a gene regulated only by the controlled TF to a measured level of TF. In all cases, the average number of TFs from each induction condition is found by converting the arbitrary fluorescence signal of each TF-mCherry fusion to TF number through a fluctuation counting method. For the “no binding site” control data (blue points), the fold change is typically 1 for all TF concentrations; in other words, there is no regulatory response to the TF in the absence of a binding site. The empty data points on the plots represent the sample means of TF number and fold change in each of three replicates. The filled data points represent the mean of the 3 replicates, and the error bars represent the standard error of the mean. When the binding site is inserted just downstream at +1 (red points), the observed regulatory function is always repression. However, at -61 (green points), the response can vary between repression that is as strong as +1 (i.e., AgaR in B), repressive but weaker than at +1 (i.e., AcrR or BetI in A and E), or it can have the opposite role and activate (i.e., CpxR in F). The fits represent least-square optimization of the theory presented in Figure 1, with dashed lines representing the 95% confidence interval generated by bootstrap sampling.

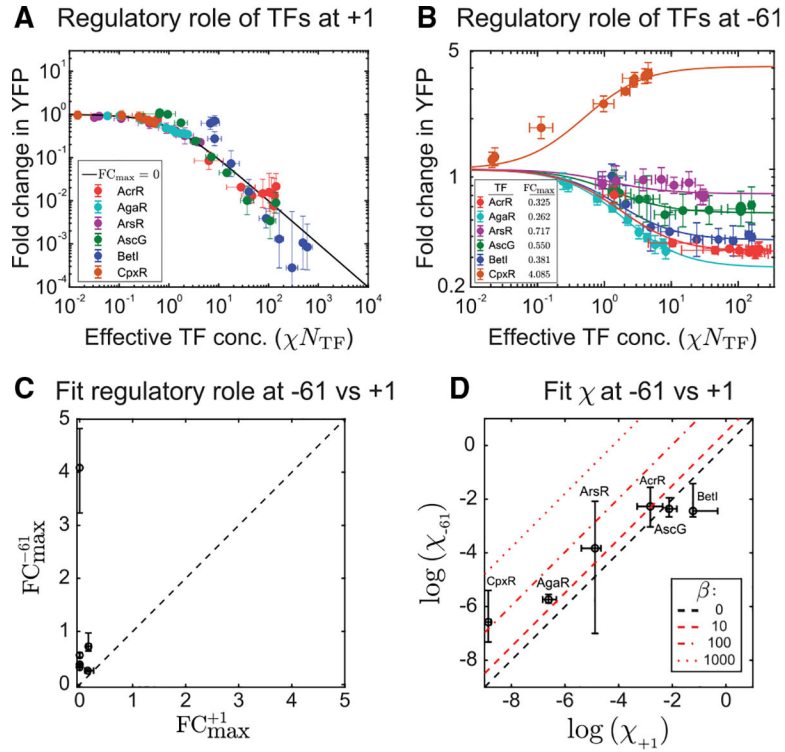


Figure 4. Fold change for a given regulatory location versus N_{eff}

(A and B) The regulatory curve for all TFs when acting at (A) +1 and (B) -61 plotted against $N_{\text{eff}} = \chi N_{\text{TF}}$. Filled data points for each TF along with the error bars represent the mean and standard error of the sampling mean as in Figure 3. In all cases, the binding energy and FC_{\max} are determined from fitting the equation in Figure 1A to the +1 and -61 data independently. Although the data for +1 are well described by a single regulatory behavior for every TF (pure repression; i.e., $FC_{\max} \approx 0$), the same TFs at -61 have a spectrum of quantitatively distinct regulatory behaviors.

(C) There is no correlation for the overall regulatory role of the TF between +1 and -61, indicating position dependence for the regulatory role of these TFs.

(D) The inferred TF binding affinity is consistent between +1 and -61 for all but two TFs corresponding to AgaR and CpxR, possibly indicating a contribution from TF stabilization ($P\beta > 1$).

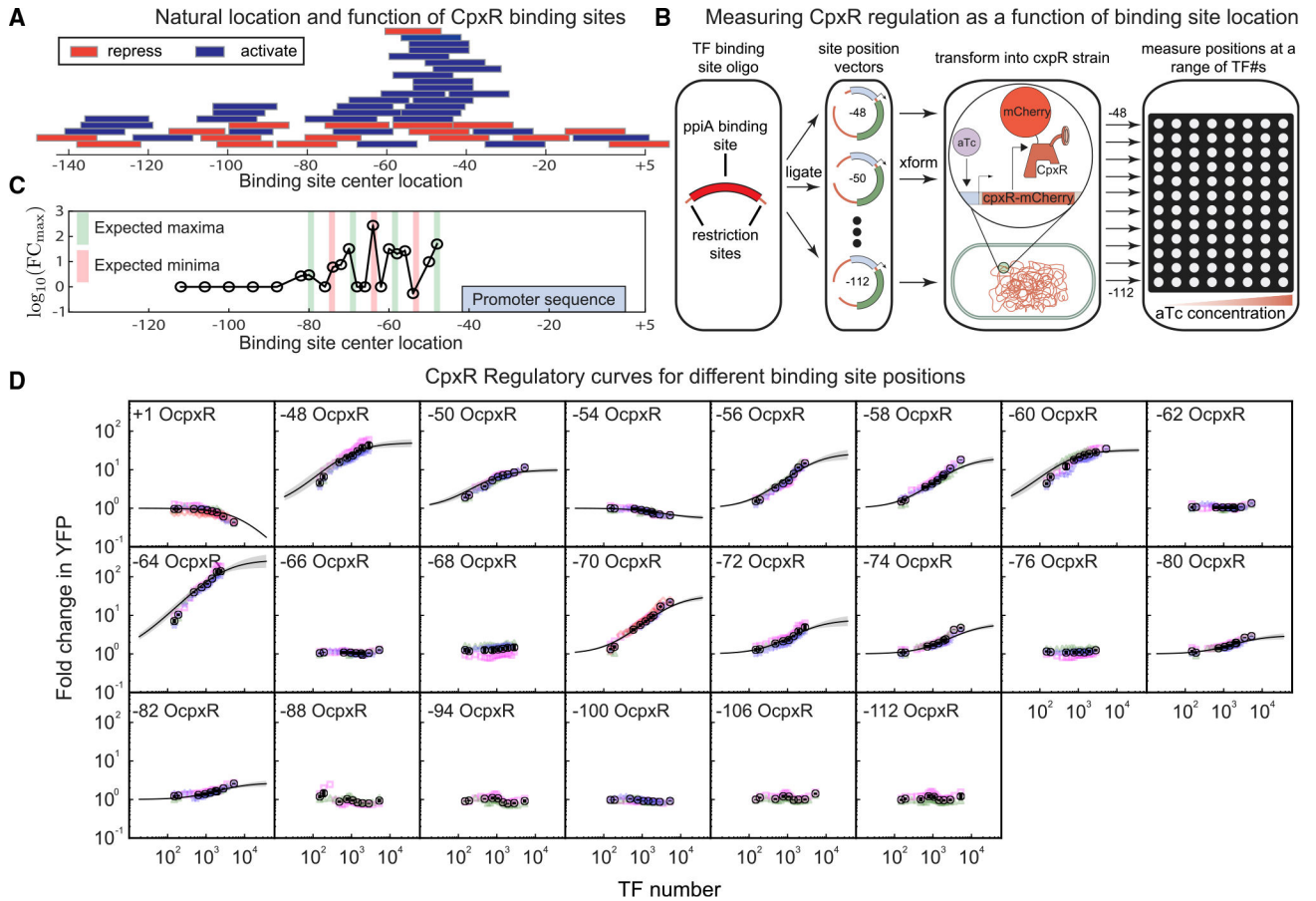


Figure 5. Position-specific regulatory profiles of the CpxR TF

(A) The distribution of CpxR binding sites across all naturally occurring genes in *E. coli*. The length of the rectangles represents the span of the binding sequence, and the border color represents activation (blue) or repression (red). The majority of the binding sites are centered between the -40 to -80 positions.

(B) Schematic of the strategy for constructing and measuring CpxR acting at a specified binding site (the *ppiA* binding sequence) inserted at 21 upstream positions and 1 downstream position on the promoter.

(C) The mean of the inferred maximal fold change (FC_{max}) for each of the 21 upstream binding locations as a function of the binding location at the promoter. The centers of the red and green shaded areas denote the presumed locations for the minima and maxima of the regulatory response (anchored on -48) based on the 10.5-bp periodicity of B-form DNA.

(D) The regulatory profile of CpxR as a function of TF copy number for the 22 binding locations. Each panel shows the TF copy number on the x axis and the fold change in YFP on the y axis of individual replicates (colored points). For all positions that show regulation (plots with the model curves), $n \geq 3$, with n as the number of replicates. The black points represent the mean and standard error of these replicates. For convenience, this is not shown for every TF number in the plot. The dashed line running through the data points is the theory prediction based on inference of the model parameters detailed in Figure 6. The

shaded regions represent ± 2 standard deviations of the thermodynamic model (Figure 1A) conditioned on the inferred model parameters.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

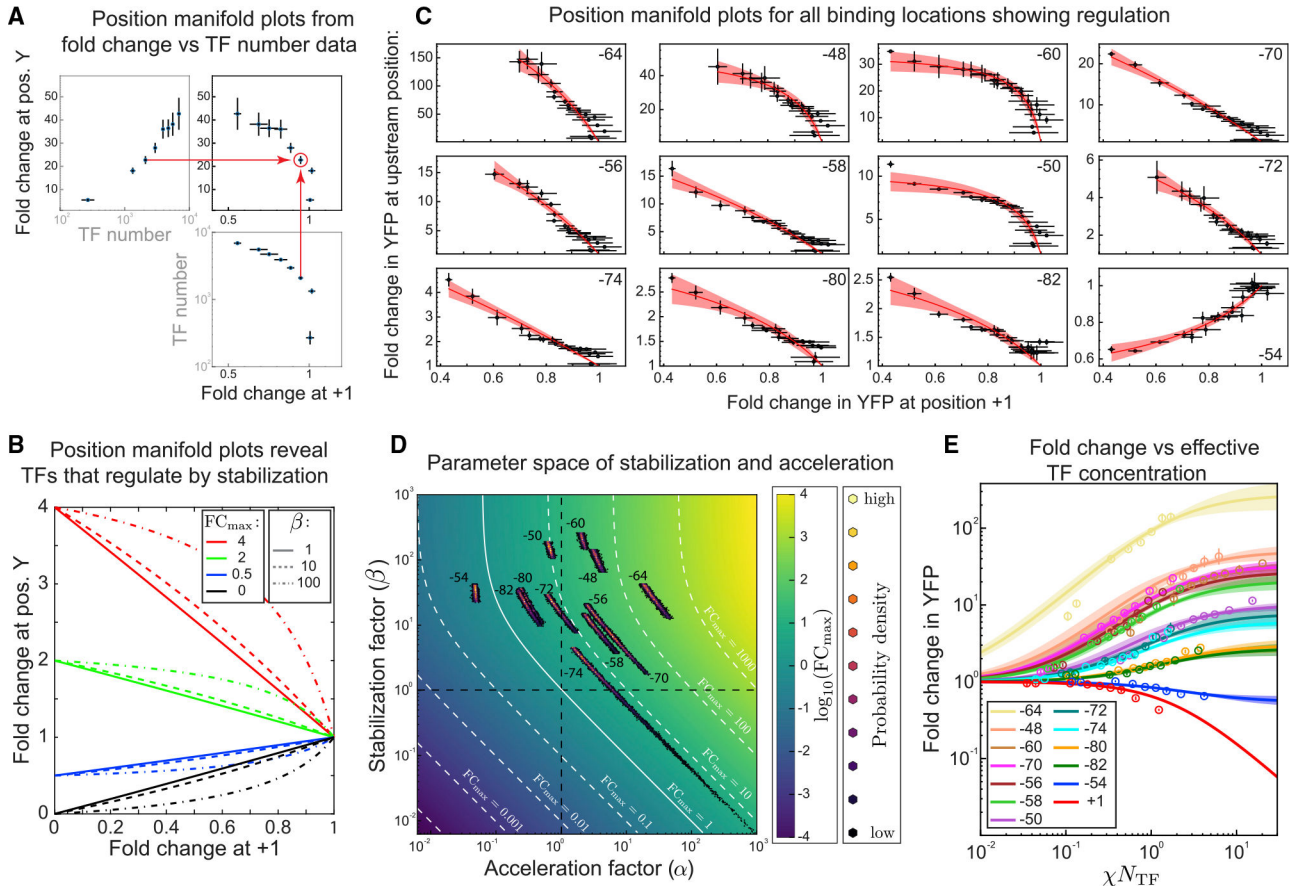


Figure 6. CpxR binding location determines the mode of TF-RNAP regulation

(A) Tracing out the position regulatory manifold using TF abundance from two different binding locations.

(B) The concentration manifold curves are predicted to be straight lines when β is small but curved for larger values of β .

(C) Concentration manifold plots for the 12 upstream regulatory positions as a function of the fold change at +1. The solid red line represents the mean of the model regulatory profile (thermodynamic model) generated from the inferred acceleration and stabilization parameters, and the shaded regions represent ± 2 standard deviations from the mean. Data points represent the mean and standard error of the fold change as seen in the plot of Figure 5D for each respective position.

(D) Phase plot of FC_{max} as a function of α and β parameters. The horizontal black line marks $\beta = 1$ (no stabilization or destabilization), and the vertical black dashed line marks $\alpha = 1$ (no acceleration or deceleration). The white lines represent contours of constant FC_{max} . The colored points represent the parameter inference of α and β for each of the regulatory positions. The median as well as the 68% credible intervals of the inferred parameters for each position are reported in Table 2.

(E) Plot of fold change against the effective TF concentration χN_{TF} for the 13 regulatory positions (12 upstream and 1 downstream) using parameters derived from the concentration manifold plots in (C).

Table 1.

Channel settings used for the cytometry acquisition

Emission filter	Channel	Voltage
525/50 nm	B1	510 V
615/20 nm	Y2	524 V
561/4 nm	FSC	405 V
561/4 nm	SSC	315 V

All cytometry measurements were done on a MacsQuant VYB with the listed channel settings. A threshold using the forward scatter (FSC) and side scatter (SSC) area measurements was used initially to gate the event data. Subsequent data processing steps to convert these cytometry measurements to the fold change plots in Figures 5 and 6 can be found in the STAR Methods.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Inferred acceleration and stabilization parameters

Binding position	α (alpha)	β (beta)
-48	3.506 ^{4.008} _{2.898}	116.257 ^{150.131} _{66.453}
-50	0.675 ^{0.736} _{0.612}	161.766 ^{199.140} _{108.281}
-54	0.049 ^{0.051} _{0.047}	34.496 ^{43.447} _{23.154}
-56	3.38 ^{4.496} _{1.745}	14.723 ^{21.647} _{5.304}
-58	3.107 ^{4.127} _{1.635}	10.519 ^{15.386} _{3.796}
-60	2.161 ^{2.33} _{1.974}	219.757 ^{270.481} _{151.499}
-64	24.911 ^{31.849} _{14.911}	29.545 ^{44.169} _{12.984}
-70	6.995 ^{9.748} _{3.101}	6.224 ^{9.316} _{1.776}
-72	0.813 ^{1.053} _{0.451}	20.905 ^{30.744} _{7.678}
-74	2.329 ^{3.944} _{0.467}	2.982 ^{4.685} _{4.789 × 10⁻⁶}
-80	0.289 ^{0.355} _{0.197}	27.025 ^{37.948} _{10.944}
-82	0.271 ^{0.334} _{0.182}	24.811 ^{35.501} _{9.835}

Shown are median values of the inference chain along with the bounds that encompass the 68% Bayesian credible interval of the parameters α and β inferred from the data presented in Figure 6C. The “weak stabilization limit” limits the precision estimates for position -74.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
<i>Escherichia coli</i> , strain BW135112	KEIO collection	NZ_CP037857
Chemicals, peptides, and recombinant proteins		
M9 minimal media	BDDiagnostics	DF0485-17
Critical commercial assays		
ZR Plasmid Miniprep	Zymogen	Cat#D4015
Experimental models: Organisms/strains		
<i>Escherichia coli</i> , strain MG1655 – Complete list of modified <i>E. coli</i> strains can be found in Tables S2 and S4	This Paper	CGSC#6300
Oligonucleotides		
Position Sweep Primers for <i>ccdB</i> cassette insertion into pZS25LongUPDL5-delbs-YFP (see Table S5 – Supplement Section)	Primer design using custom-made python script, synthesized by Genewiz	N/A
Primers for amplifying the TF gene cassettes from the MG1655 genome – See Table S1	Genewiz	N/A
Recombinant DNA		
pDONR P4-P1r	Thermo Fisher (Invitrogen-Life Technologies)	N/A
<i>ccdB</i> Position cloning strains (Table S6)	This Paper	N/A
Software and algorithms		
MATLAB	https://www.mathworks.com/products/matlab.html	N/A
Python	https://www.python.org/downloads/	N/A
PyMC3	https://docs.pymc.io/en/stable/about.html	N/A