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

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

Abstract

Gene regulation often results from the action of multiple transcription factors (TFs) acting at a promoter, with a net regulation that depends on both the direct interactions of TFs with RNA polymerase (RNAP) and the indirect interactions with each other. Here we measure the fundamental regulatory interactions of TFs in *E. coli* by designing synthetic target genes that isolate the individual TFs regulatory effect. Using a thermodynamic model, the direct regulatory impact of the TF on RNAP is decoupled from TF occupancy and interpreted as acting through two mechanisms: (de)stabilization of RNAP and (de)acceleration of transcription initiation. We find the contributions of each mechanism depends on TF identity and binding location; for the set of TFs profiled, regulation immediately downstream of the promoter is insensitive to TF identity, yet these same TFs regulate by distinct mechanisms upstream of the promoter. Strikingly, we observe two fundamental regulatory paradigms with these two mechanisms acting coherently, to reinforce the observed regulatory role (activation or repression), or incoherently, where the TF regulates two distinct steps with opposing effect. This insight provides critical information on the scope of TF-RNAP regulation allowing for a stronger approach to characterize the endogenous regulatory function of TFs.

Keywords: Systems biology, Synthetic biology, Transcription regulation

¹Program in Systems Biology, University of Massachusetts Medical School, Worcester, MA 01605

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

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

1 Introduction

2 Transcriptional regulation of gene expression is one of the major mech-
 3 anisms by which cells respond to cues and stimuli. Transcription factors
 4 (TFs) perform this regulation through binding to the DNA around the pro-
 5 moter to alter the rate of transcription from individual genes [1, 2]. The
 6 regulatory DNA of each gene is distinct and can involve several to dozens
 7 of TF binding sites arranged in specific architectures to achieve the desired
 8 expression level. However, predicting the level of gene expression based on
 9 the regulatory architecture of a gene remains a central challenge in the field
 10 [3, 4, 5, 6, 7].

11 The genomics era has enabled multiple techniques capable of determining
 12 where a TF will bind and with what specificity [8, 9, 10, 11]. Although this in-
 13 formation is crucial for building occupancy-based models of gene regulation,
 14 there is still another critical component which is missing; the quantitative
 15 regulatory role of a TF, once bound, is often unclear. Historically, measure-
 16 ments of TF function are the result of knocking out the endogenous TF and
 17 observing how gene expression changes as a result. This serves its purpose
 18 in predicting the specific role of that TF on a given gene but offers less pre-
 19 dictive power when examining regulation of other genes by that same TF or
 20 to different binding sites. These measurements of gene regulation are often
 21 entangled in indirect regulatory effects such as TF-TF interactions [12, 13],
 22 feedback [14, 15], physiological (*i.e.* growth rate) [16, 17] and off-target com-
 23 petitive effects of decoy binding sites or other genes in the network [18, 19].
 24 Due to this, a single TF that binds at the same relative location on two dif-
 25 ferent natural promoters can appear to have opposite regulatory roles. The
 26 entanglement between indirect and direct regulation likely contributes to this
 27 ambiguity and prevents the field from developing a basic intuition of TF reg-
 28 ulatory function. The select few TFs that have arisen as “model TFs” such
 29 as LacI [20, 21, 22, 23, 24], AraC [25, 26, 27, 28], lambda repressor [29, 30, 31]
 30 and TetR [32, 33, 34, 35] have well studied regulatory function. Indeed, these
 31 TFs have been utilized for design of synthetic circuits with engineered pur-
 32 pose such as the creation of logic gates [34, 36], bistable switches [35, 37],
 33 oscillatory networks [38, 39, 5], synthetic enhancers [40, 41] and a host of
 34 other dynamic outcomes. Further characterizing the regulatory function of
 35 TFs beyond this small subset should provide a more complete toolset for

36 broader synthetic design purposes.

37 Here we study the isolated regulatory function of a set of *E. coli* TFs in
 38 a system designed to remove the typical confounding factors of natural genes
 39 and quantify the direct regulatory effect of a TF based on factors such as
 40 TF concentration, binding affinity and binding location. Using a collection
 41 of strains where the average copy number of most TFs in the cell can be
 42 controlled, we measure the level of regulation of an individual TF acting on
 43 a synthetic promoter sequence. This promoter is designed to be regulated
 44 only by that TF, and it is targeted to a binding site whose location and
 45 sequence we control. To interpret this data, we use a thermodynamic model
 46 of gene regulation to parameterize TF regulatory function. In principle,
 47 the TF could exert its regulatory effect at any one of the distinct kinetic
 48 steps of the transcriptional process [42, 43, 44, 45] or on several of them
 49 simultaneously, and our model course-grains TF activity into two distinct
 50 modes of regulation. The first regulatory mode “stabilization” corresponds
 51 to stabilization (or destabilization) of the polymerase at the promoter by
 52 the TF. The second mode “acceleration” corresponds to a TFs ability to
 53 accelerate (or decelerate) the initiation of transcription when both the TF
 54 and polymerase are bound to the promoter. Using this model we infer the
 55 quantitative contribution from each of these modes in the data. Importantly,
 56 this process allows for the decoupling of properties that are extrinsic to the
 57 TF such as affinity to the operator binding site, the overall concentration of
 58 the TF or feedback in the network from the core regulatory role of the TF
 59 in modulating the steps of the transcription process.

60 We expect the regulatory parameters of a TF to vary based on the iden-
 61 tity of the TF and the binding site location on the regulated promoter. In
 62 this study, we investigate the role of TF identity by measuring regulation of 6
 63 TFs (AcrR, AgaR, ArsR, AscG, BetI, and CpxR) at two common binding lo-
 64 cations: directly downstream of the promoter, where repression is commonly
 65 observed, and 61 bases upstream of the promoter, a site commonly associ-
 66 ated with activation (although databases of regulatory interactions record
 67 roughly as many TFs repress at -61 as activate). We find that, despite the
 68 diverse nature of the TFs tested (five of the TFs are annotated repressors
 69 and one of them, CpxR, is a known activator), the regulation for all TFs
 70 immediately downstream is consistent with a form of repression that is set
 71 by the degree of occupancy of the TF at the promoter - independent of TF
 72 identity. This commonality across the TFs disappears when we measure the
 73 effect at -61, where the TFs exhibit different degrees of stabilization with

both CpxR and AgaR engaging in significant stabilization of RNAP. To complement 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 yet slowing the rate of promoter escape - demonstrating that such “incoherent” regulation plays a useful role in 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

In order 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 [46, 47, 48, 49, 50, 51, 52] 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 (middle column) and the rate of expression from each promoter state (right column); the promoter can be unbound by TF and RNAP, bound by polymerase only, bound by TF only, or bound by both. The probability with which these states occur is a function of each molecules (polymerase and TF) binding affinity with their specific DNA binding sites ($\Delta\epsilon_P$ and $\Delta\epsilon_{TF}$) and the available number of each molecule in the cell (N_p and N_{TF}). For the cobound 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 due to a favorable (activating) or unfavorable (repressive) interaction between TF and polymerase. As a result, the cobound 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 cobound and is written as a multiplicative factor to the base expression rate of polymerase bound in the absence of TF;

for example, $\alpha = 2$ would imply that transcription initiation rate is doubled when the TF and polymerase are co-bound. In both cases the parameters represent increases to gene expression when greater than unity and decreases of 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 for a TF that both 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 [53, 48]. Furthermore, the parameters related to polymerase binding can be simplified into a single parameters in our model as $P = N_p \exp(-\Delta\epsilon_P)/N_{NS}$.

The final expression, boxed below in Figure 1A (and derived in the Methods section), 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.* $N_{TF} = 0$). Fold-change greater than 1 signifies activation, while 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 TFs 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 ($\Delta\epsilon_{TF}$) and the degree to which the TF recruits/stabilizes RNA polymerase ($1 + \beta P$). 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 vs 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 (upper curves) and a repressor (lower 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 is once 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 further 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, however the strong stabilizer will have a higher effective TF concentration and thus move further along the curve given the same TF concentrations and binding affinities. In figure 1D, data collapse curves for a range of $FC_{\max}$ values is shown.

Figure 1E provides a hierarchical overview of model parameters and their relationship to controllable biological features of the *in vivo* system. The upper 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 $FC_{\max}$. Due to the fact that 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 [54]) 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

185 $FC_{\max}$ to determine the two regulatory parameters α and β and it is not pos-
 186 sible to distinguish between regulation driven by a change in the transcription
 187 rate (acceleration/deceleration) or by modulation of polymerase occupancy
 188 at the promoter (stabilization/destabilization).

189 **Experimental measurements of individual TF regulatory function**

190 To measure regulation by an individual TF in *E. coli* as a function of
 191 TF identity and binding position on the promoter we utilize synthetic tech-
 192 niques to create simple and controllable gene circuits. In this approach we
 193 have created *E. coli* strains, schematically illustrated in Figure 2A-B, where
 194 the endogenous copy of each studied TF is knocked out and reintroduced
 195 as a TF-mCherry fusion integrated into the genome at the *ybcN* locus (Fig-
 196 ure 2A). Expression of the synthetic TF-mCherry promoter can be induced
 197 with anhydrotetracycline (aTc) (for details, see *Methods*). This system en-
 198 ables precise control of TF copy number (Figure 2D) which is measured either
 199 by wide-field fluorescent microscopy or flow cytometry (Figure 2C, see *Meth-*
 200 *ods*). Using this system we are able to induce these TFs from leaky levels
 201 (several per cell) up to several thousand per cell; the full induction curve
 202 of each strain is shown in Figure 2D. Importantly, although we control the
 203 concentration of the TF, many TFs are capable of existing in distinct binding
 204 conformations that may alter the active TF number, the regulatory param-
 205 eters (α and β), or both. For all but one of the TFs studied here, we expect
 206 that the TFs will always be active in our growth conditions, however BetI is
 207 inactivated by choline (which we do not control for) and as such we expect
 208 to have some fraction of BetI inactive at low TF concentrations [55, 56].

209 In each of these individually tunable *tf-mCherry* strains, a target pro-
 210 moter that drives YFP expression is integrated into the genome at the *galK*
 211 locus, see Figure 2B. The basic promoter incorporates a modified *lac* RNAP
 212 binding sequence where the RNAP occupancy term, P , was measured previ-
 213 ously [54]. Otherwise, the promoter is designed to be free of specific known
 214 TF binding sequences. To study the regulatory role of a specific TF, we intro-
 215 duce a TF-specific binding site (chosen from an array of known binding sites
 216 with strong evidence of that particular TF binding [57, 58]) cloned either
 217 directly downstream of the transcription start site (TSS), annotated as +1
 218 for simplification or centered at 61 bases upstream of the TSS, annotated as
 219 -61. The effect of the TF binding to the promoter is then measured in terms
 220 of YFP fluorescence protein expression as a function of average number of
 221 TFs per cell for a given induction condition (Figure 2C). For this work, our

222 focus is to select a binding site that will bind the TF. The affinity of the site
223 or how the particular choice may influence the regulatory parameters is not
224 something we explore exhaustively here. The specific binding sites chosen
225 for each TF can be found in the supplemental material.

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

227 Figure 3A-F shows the fold-change in YFP expression (promoter activ-
228 ity) as a function of TF copy number for the TFs examined in this study
229 measured using single-cell fluorescent microscopy. In these plots regulation
230 at +1 is shown as red points, regulation at -61 is shown as green points and
231 a control promoter with no TF binding site is shown in blue. The six TFs
232 display diverse regulatory behavior which depends on both the TF identity
233 and TF binding location on the gene. For example, CpxR (figure 3F) is a
234 repressor when bound at +1 but activates at -61, AscG (figure 3D) is a
235 strong repressor at +1, however has almost no regulatory role at -61, and
236 BetI (figure 3E) represses at both +1 and -61 but much more weakly at -61
237 despite binding to the same sequence at both locations. One commonality
238 between the curves is that at +1 every TF in this study acts as a repres-
239 sor. Naively the “strength” of this repression appears rather diverse; the
240 repression from some TFs reduces YFP expression by 10^{-3} in fold-change
241 (BetI), while others never drop below 10^{-1} (AgaR). The solid lines in this
242 figure show a fit to the model in Figure 1A (boxed equation). In this case
243 we are explicitly measuring the number of TFs, N_{TF} , and we fit for the two
244 unknown parameters χ and FC_{max} (see Figure 1A). In four of the six curves
245 (Figure 3A,D-F), the +1 regulation data gives FC_{max} consistent with zero.
246 This result is consistent with the regulatory mode of perfect repression; *i.e.*
247 that the TF completely shuts the gene off when bound. As such, the dif-
248 ference in regulation at +1 between the TFs must be attributed entirely to
249 either binding affinity or differing levels of stabilization (β) between TFs.
250 A typical assumption is that TFs operating at +1 regulate by steric hin-
251 drance ($\beta = 0$), such that when the TF is bound at +1 polymerase binding
252 is occluded. Previous studies [59, 46, 18, 60] support this assumption. The
253 remaining curves (Figure 3B,C) show FC_{max} of order 10^{-1} , however in both
254 cases the binding is weak to the point we do not see the expected saturation
255 of fold-change with TF copy number and, as such, the data from these curves
256 is also consistent with FC_{max} of zero with a slightly increased value of χ to
257 compensate. The collapse of all +1 regulation data to the perfect repression
258 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 fits to a single regulatory contour associated with perfect repression once 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 $\text{FC}_{\text{max}} = 0$), at -61 these TFs operate with a diverse range of regulatory effects; some TFs mirror the function at +1, showing a similar profile to the response function at +1, while others show limited repressive capabilities that saturate at specific fold-change values (FC_{max}), while other TFs activate expression at -61. In order 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 while 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, $\Delta\epsilon_{\text{TF}}$, to depend on where the binding site is located since the binding sequence is not changed [57]. The other possible contribution to χ comes from stabilization and takes the form $(1 + \beta P)$ which means that β needs to be much larger than P for stabilization to impact χ . We expect the data points will fall into two potential outcomes. Points that lay on the black dashed line indicate that stabilization is not playing 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). As can be seen, 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 be playing a role in their respective regulation. The implied magnitude of stabilization is shown by the red lines in the figure. 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 the AgaR, and demonstrates an incoherent regulatory strategy where the TF engages RNAP at two distinct steps with opposing effect.

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 in order to regulate dozens of different genes in *E. coli*. However, the regulatory role of CpxR as a function of binding location is unclear from this data; both 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 base pairs from the TSS to -112 base pairs from the TSS (see Figure 5B). We chose this range because the vast majority of natural CpxR binding sites occur within these limits (Figure 5A). In order 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). We find consistent results with the target gene on plasmid or integrated into the genome (see supplemental material).

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 base pairs) 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 this data. Interestingly, we expected to see an 11 base periodicity in $FC_{\max}$

corresponding to the helicity of DNA [61, 52, 62]; we see this very roughly with maxima in activation around -48, -60, -70 and -80 (± 1 bp), however -64 is a strong outlier which is expected to be close to a minimum but is strongly activating as measured here.

To extract the regulatory parameters α and β precisely from this data, it is helpful to replot the fold-Change data using a “position manifold”. Demonstrated schematically in Figure 6A, the position 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 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 regulation is “pure steric hindrance” ($FC_{\max} = 0$, $\beta = 0$) here. The advantage of this approach is that it enables the 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 Methods for model assumptions). As seen in Figure 6B, data in the position 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 towards 0 and the fold-change at the second position will either increase or decrease depending on the regulatory function at that position; as such activating positions have curves that rise as you move towards zero on the x -axis while repressive positions decrease. The role of β is clear in this formulation. When β is small (compared to $1/P$) 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. A related approach using regulatory manifolds has been applied previously [60].

In Figure 6C we replot the data from Figure 5D as fold-change at a certain 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 1. We find strong stabilization in regulation at positions -50, -60, -54, -48, and -64; visual inspection of Figure 6C shows the curvature we

372 expected to see from strong stabilization. Several positions -58, -56, -70, and
373 -74 have regulation profiles that are approximately straight lines implying
374 that CpxR either destabilizes or does not regulate through stabilization (*i.e.*
375 $(1 + \beta P) \approx 1$).

376 Figure 6D shows a heat map for $\log(\text{FC}_{\max})$ as a function of the regu-
377 lation parameters α and β . The dashed black lines, which denote α and β
378 equal to 1, divide the map into four quadrants, each quadrant with a spe-
379 cific qualitative regulatory scheme. The top-right and bottom-left quadrants
380 represent coherent regulation strategies where α and β both contribute to
381 activate (top-right) or repress (bottom-left) gene expression. On the other
382 hand the top-left and bottom-right quadrant are incoherent in the sense
383 that α and β have opposing regulatory effects: TFs in these quadrant either
384 slow the initiation rate of transcription while also stabilizing polymerase at
385 the promoter (top-left) or increase the initiation rate of transcription while
386 destabilizing the polymerases presence at the promoter (bottom-right). The
387 solid, white contour in this plot shows where these two effects balance and
388 the net fold-change is 1, left of this line represents TFs that repress, right
389 of this line represents TFs that activate. Contours of constant $\text{FC}_{\max}$ are
390 drawn as white dashed lines marking 10 fold increases/decreases in $\text{FC}_{\max}$.
391 On this same plot we also show the inferred probability distribution of the
392 parameters α and β for each position in our data. The black points are lower
393 probability with lighter points representing higher probability values for the
394 α and β parameters. One notable phenomenon is for positions with β less
395 than roughly 10, the inference begins to fail for α and β . This results in
396 inference clouds with "tails" that stretch across quadrants and precludes an
397 assessment into the mode of regulation (see position -74). The alignment
398 between the inference clouds of these positions and the constant $\text{FC}_{\max}$ con-
399 tours, however, shows that while we make very precise estimations of $\text{FC}_{\max}$,
400 the values of α and β are less certain and correlated. This is an unfortunate
401 consequence of the weak stabilization limit in our model.

402 Interestingly, at some positions (-54, -50, -80 and -82), we see the in-
403 coherent behavior discussed above where the TF both stabilizes polymerase
404 at the promoter and also slows the rate of initiation, essentially serving oppos-
405 ing functions in influencing gene expression. For -54 the net effect of these
406 opposing mechanisms is repression, while at -80, -82 and -50 the result is
407 activation. However, the positions with strong activation signatures (-48,
408 -60, -64) as well as some intermediate ones (-56, -58, -70) have stabiliza-
409 tion 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 position manifold space, with the measured extrinsic features (TF copy number and binding affinity) of gene regulation produces model curves using the effective parameterization that fits our data well. The effective parameters $FC_{\max}$ and χ for the CpxR TF data (See Methods section 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 plot in Figure 4A and B.

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 [63, 64, 65]. 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 RNA polymerase at the promoter [66]. Here, we use a synthetic biology approach to measure the isolated regulatory effect of a TF on an otherwise constitutive promoter. This data is 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 for 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 huge 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 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 was primarily due to difference in TF binding affinities rather than in the fundamental regulatory mechanisms of the TFs. In contrast, when these same TFs bind 61 base pairs upstream of the promoter the regulatory function of the TFs varied more substantially. We find some TFs remained as strong repressors (similar to their function at +1) however 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, profiling the regulation of CpxR at upstream binding locations, we find that this TF is capable of regulating multiple steps of the transcriptional process simultaneously and joins a growing body of evidence for TFs engaging in complex regulatory maneuvers at the promoter [67, 68, 69]. 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, where previous work analyzing the kinetics of activator regulation have relied on *in vitro* biochemical assays [70]. 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, 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 both 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-promoter 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 [71, 19, 72, 73], and addressing how these contacts shapes the relative effects of α and β would fill a vital gap in elucidating the molecular determinants that give rise to either coherent and incoherent regulatory modes. Such information would allow for a more complete characterization of TFs, and in conjunction with methods of profiling TFs through genome wide occupancy techniques provides an important 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 [74, 75, 76]. Realizing this goal, however, requires a systematic characterization of the TF position-regulatory axis to determine the spatial landscape of α and β and provides a compelling basis for characterizing a broader set of TFs to generate a "regulatory compendium" where TFs are classified 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 ($\beta \leq 1$) and limits the characterization of TFs that engage through this mode. 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. As such, measuring the isolated regulatory function of TFs may require a range of promoters to fully characterize the wide array of regulatory behaviors possible.

Here we focus entirely on the isolated regulatory role of each TF, however it is clear that one of the next steps is to probe how quantified TFs regulate together. the ability to distinguish and quantify the stabilization and acceleration mode of regulation and characterize TFs based on them is important for developing general theories of regulation that include TFs acting on different kinetic steps of the transcription process [77, 78, 79]; 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 [77]. With each TF characterized, we can develop an empirical baseline or null hypothesis for what a TF should do on a gene: departures from this expectation, due to the emergence of complex regulatory phenomenon brought about by TF-TF interactions [63], allosteric interactions [80, 81] or other effects indicate surprises that warrant testing in these expanded models.

Methods

Thermodynamic model for single TF regulation

In the work here we use the standard form of the thermodynamic model as derived elsewhere [46, 48, 50]. In our framework, we include an additional consideration of the TF altering the rate of transcription as detailed in [52]. The partition function for this system is:

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

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 $\Delta\epsilon_p$ and $\Delta\epsilon_{\text{TF}}$ 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 $\Delta\epsilon_I$. The probability to find polymerase bound as a function of TF number is then,

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

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 $P_{\text{bound}}(N_{\text{TF}})$. For the state in which RNAP is solely bound, we give a rate of expression as r (a coarse-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):

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

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:

$$\text{fold-change} = \frac{\text{YFP}_{\text{expression}}(N_{\text{TF}} \neq 0)}{\text{YFP}_{\text{expression}}(N_{\text{TF}} = 0)} = \frac{1 + \alpha N_{\text{TF}}/N_{\text{NS}}e^{-(\Delta\epsilon_{\text{TF}}+\Delta\epsilon_I)}}{1 + \frac{(N_{\text{TF}}/N_{\text{NS}}e^{-(\Delta\epsilon_{\text{TF}})})(1+N_p/N_{\text{NS}}e^{-(\Delta\epsilon_p+\Delta\epsilon_I)})}{(1+N_p/N_{\text{NS}}e^{-(\Delta\epsilon_p)})}}. \quad (4)$$

We then define the following terms $\beta = e^{-\Delta\epsilon_I}$ and $P = N_p e^{-\Delta\epsilon_p}$. As we have measured the value of P in our synthetic circuit to be 6.6×10^{-2} [54], 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:

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

We now define the final term in our derivation: $\chi = e^{-\Delta\epsilon_{\text{TF}}}/N_{\text{NS}}(1 + P\beta)$ which represents the effective component that modifies the TF copy number with the product $N_{\text{TF}}\chi$ 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-change ($\text{FC}_{\text{max}} = \frac{\alpha\beta}{1+P\beta}$) as presented in Fig1 and the main text.

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

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 [21, 18]. 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* [54]. 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* [54]. Given that we have previously measured the core-promoter strength used in our synthetic circuit to be at $P = 6.6 \times 10^{-2}$ [54], 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 – 1000 fold repression at +1.

Position 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):

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

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.

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

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

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

Based on this reformulation, the key parameters to consider are the acceleration parameters (couched in the $FC_{\max}$ term as described in Fig1) and the stabilization parameters at positions x and y (a total of 4 parameters). In a sense, the “position 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 position manifold for a given regulatory position (as a function of the fold-change at the +1 position - FC^{+1}):

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

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 *ylcN* 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 *ybcN* locus) and 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 $10^4 - 10^5$ 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 0.4 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 [18]. 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 1:10000 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.

665 The aTc dilution solutions were made from a stock solution of 1 mg/mL sus-
 666 pended in ethanol and were made fresh prior to the application of the aTc
 667 for the M9 culture. After M9 dilution, the strains were grown in 96 well
 668 plates to steady state (OD600 of 0.1 – 0.2). Similar to microscopy acquisi-
 669 tion procedure, we had constitutive (CpxR-KO) strains transformed with the
 670 binding position plasmids along with the CpxR-titration strain transformed
 671 with the plasmid having the TF binding sequence removed from the promoter
 672 to account for physiological effects of CpxR-mCherry titration to calculate
 673 the fold-Change. Cells were diluted between 1:16 to 1:32 fold in PBS media
 674 in a 96 well cytometer plate prior to data acquisition and cytometry was
 675 performed on a MacsQuant VYB. At the beginning of each run, an initial
 676 gating strategy involving the FSC and SSC height information was used to
 677 eliminate background events and samples were run to achieve ~ 60000 gated
 678 events for each position strain at a given aTc concentration.

679 Data processing steps for Microscopy and Cytometry Data

680 For the +1 and -61 position sweep regulatory curves of the 6 TFs, we
 681 took the microscopy images and segmented individual cells using a modified
 682 version of the Matlab code Schnitzcells [82]. We use this code to segment the
 683 phase images of each sample to identify single cells. Mean pixel intensities
 684 of YFP and mCherry signals are extracted from the segmented phase mask
 685 for each cell. The autofluorescence is calculated by averaging the mean in-
 686 tensity of the autofluorescence strain in both mCherry and yfp channels and
 687 is subtracted from each measured YFP or mCherry value. Total fluorescence
 688 for each channel is obtained by multiplying the mean pixel-intensity with
 689 the area of the cell. Fold-change in expression for a given binding site is
 690 calculated by the ratio of total fluorescence of strains expressing the TF to
 691 the strains with no TF. For partitioning statistics to estimate the calibration
 692 factor, mother-daughter pairs are first automatically identified and verified
 693 manually to ensure cells made exactly one division. The mean pixel intensity
 694 and area of the mother-daughter pairs are obtained. The background fluo-
 695 rescence is estimated as described in Ali *et al* [7] using the inverse mask of
 696 individual frames. The sum and difference in fluorescence of the two daugh-
 697 ters were then used to find the conversion factor, v , between fluorescence and
 698 number of TFs using the equation $(I_1 - I_2)^2 = v(I_1 + I_2)$, which stems from
 699 the assumption of binomial partitioning of TFs at cell division [82].

700 For the CpxR position cytometry data, we adapted a robust data analysis
 701 procedure [83] 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 height 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 data set for every RFP bin and used these statistics to fit an ellipsoid to the full data set according to the following formulae:

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

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

This step retains events that are within a particular distance from the center of the ellipsoid using an appropriate value for the cut-off (alpha). We based the value of the cut-off on the following rationale: As each event is essentially a vector of FSC and SSC values (the $\log_{10}$ values of the Forward Scatter and Side Scatter), we assume the joint values are normally distributed, which translates to a distance metric that is a chi-squared random variable (the sum of two normally distributed entities is chi-squared with df 2). 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" (*DelBS*) circuit. The YFP signal for these events had large fluctuations and we reasoned to the flow-cytometry approach probably fails in measuring cells with low 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 SI Figures 5 and 6 for details).

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 vs RFP curve for a given TF-binding location across all replicate datasets 1000000 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 94th percent confidence interval.

For the CpxR position sweep data, we used the position 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 FC vs RFP replicates using the position manifold formalism as detailed in the Methods (see Position 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

770 $K = 1 + P\beta$ using a Bayesian approach that relates the parameters un-
 771 derlying our thermodynamic model to the data according to the following
 772 relation:

$$p(\text{FC}_{\max}, K|FC) = \frac{L(FC|\text{FC}_{\max}, K) p(\text{FC}_{\max}, K)}{p(\text{FC})}, \quad (13)$$

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

$$L^K(\text{FC}^k|\text{FC}_{\max}^k, K^k) = \prod_i^{\text{DataPoints}} \text{Normal}(FC_i^k|u = \text{FC}_{\text{prop}}^k(\text{FC}_i^{+1}; \text{FC}_{\max}^k, K^k), sd = \theta^k). \quad (14)$$

780 The superscript k represents an upstream binding location for the CpxR
 781 TF and the subscript i represents the data points for a given position regu-
 782 latory dataset. The proposed fold-change value FC_{prop} from the model takes
 783 the form:

$$\text{FC}_{\text{prop}}(\text{FC}_i^{+1}; \text{FC}_{\max}, K) = \frac{1 + \text{FC}_{\max} K \left(\frac{1 - \text{FC}_i^{+1}}{\text{FC}_i^{+1}} \right)}{1 + K \left(\frac{1 - \text{FC}_i^{+1}}{\text{FC}_i^{+1}} \right)}. \quad (15)$$

784 The crux of the Bayesian approach to model inference is to simulate can-
 785 didate draws of the joint posterior distribution of the $\text{FC}_{\max}$ and K param-
 786 eters by proposing candidate values from the prior distribution, generating
 787 the thermodynamic curve, and evaluating the likelihood function. A transi-
 788 tion in the jointly sampled parameter space from one set of parameter values
 789 to another is based on the premise that parameter sets will be sampled in
 790 proportion to the probability of the posterior distribution (as long as the
 791 sampling chain is drawing from the stationary distribution). This process is
 792 repeated until a given number of draws have been made from the joint pos-
 793 terior distribution. The results of this inference procedure are used to draw
 794 the model curves in Figure 5C and we use the relation between the sampled
 795 parameters ($\text{FC}_{\max}$ and K) and the acceleration and stabilization param-
 796 eters as detailed previously in the methods section. The sampling procedure

797 was implemented with the PyMC3 package that utilizes the NUTS sampler,
798 a particular implementation of the Hamiltonian Monte-Carlo algorithm, to
799 sample the joint posterior distribution [84].

800 We used a uniform distribution as the priors for both the $FC_{\max}$ and K
801 model parameters with appropriate bounds for each parameter. For K , we
802 ensured that the lowest potential value is 1, in line with the assumptions
803 from the derivation of the position manifold formalism. We checked the in-
804 ferences from each position to ensure that the bounds we enforced on both
805 parameters were appropriate and that the sampled values were not tending
806 towards the edge of the sample space. Furthermore, we cast the σ param-
807 eter (the standard deviation) of the position specific likelihood function as
808 a hyper-parameter in our sampling procedure and set the prior distribution
809 as uniform over a defined interval with the lower bound = 0. Overall, our
810 inference approach allowed us to ensure precise inference of α and β and for
811 8 out of the 12 regulatory positions (See Table 1).

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

$$\beta = \frac{K - 1}{P}. \quad (16)$$

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

$$\alpha = \frac{FC_{\max}}{\beta} \left(1 + \left(\frac{P}{1 + P} \right) (\beta - 1) \right). \quad (17)$$

819 Table 1 lists the inferred acceleration and stabilization parameters. We
820 report the median values of the inference chain along with the bounds that
821 encompass the 68%th percent Bayesian credible interval of the parameters.
822 For positions -58, -70, and -74, the inference estimates are not precise due
823 to the phenomenon of the “weak stabilization” limit as discussed in the Main
824 text. Examining the posterior distribution of the $FC_{\max}$ and K parameters,
825 we find the values of the 68% credible interval for these positions (both α
826 and β) to encompass both coherent and incoherent regimes.

Using the position manifold parameters to generate the thermodynamic model in FC vs N_{TF} space

The parameters inferred using the position manifold approach were used to construct the thermodynamic model curves in Figure 5C and Figure 6E. To accomplish this, we used the $K = 1 + P\beta$ and $FC_{\max}$ values inferred for each of the 12 regulatory positions (see Methods Section: 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 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:

$$\prod_k \prod_j \text{Normal} \left(FC_j^k | u = FC_{\text{Thermodynamic}(j)}^{(k)}(\lambda), sd = \sigma^{(i)} \right), \quad (18)$$

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

$$FC_{\text{Thermodynamic}(j)}^k = \frac{1 + FC_{\max}^k \chi^k N_{\text{TF}(j)}^k}{1 + \chi^k N_{\text{TF}(j)}^k}, \quad (19)$$

where $\chi^k = \lambda K^k$. The parameter λ , is the global parameter representing the binding affinity of CpxR TF (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 $\chi N_{\text{TF}} = N_{\text{TF}} e^{-\Delta\epsilon_{\text{TF}}} K$. For the 12 upstream regulatory positions, as the chains of $FC_{\max}$ and K were available from the inference of the position 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 4A, and the K was set to 1 in keeping with the assumption of steric hindrance. The mean value of the Markov chain

856 λ inferred from this model, along with the scaling factor (see SI section -
857 **Converting the mCherry signal to TF copy number**) was used to gen-
858 erate the thermodynamic model curves (mean $\pm 2\sigma$) as seen in Figure 5C and
859 Figure 6E.

860 Cloning the TF specific binding location gene circuits

861 The upstream promoter sequence in our synthetic gene circuit was de-
862 rived from the P_{DL5} plasmid that contains a modified version of the *lacUV5*
863 promoter sequence as used previously [18]. Binding sequences for AcrR,
864 AgaR, ArsR, AscG, BetI, CpxR TFs were cloned at the +1 and the -61
865 locations of that plasmid by ordering the binding sites as a primer dimer
866 with overhangs that amplified the surrounding flanking regions of the P_{DL5}
867 vector where we wished to insert the sequences and PCR amplified the de-
868 sired regions. This amplified sequence contained the YFP coding sequence,
869 transcriptional termination sequence, and the 250 bp upstream promoter se-
870 quence containing the TF-binding site either at +1 or -61 along with the
871 Kanamycin selection cassette. The amplification produced two fragments
872 and Gibson assembly was then used to stitch the two fragments resulting
873 in a sequence that had the desired regulatory sequence, promoter, reporter
874 cassette along with overhangs for the *galK* locus. The resulting reaction was
875 purified, and transformed into the *galK* locus in using the *pKM208* recom-
876 bination *E.coli* strains [85]. The strains were then Sanger sequenced and
877 verified to contain the appropriate regulatory and promoter sequences and
878 used P1 transduction to transfer the *galK* locus into the appropriate TF
879 titration strains (both the TF knockout strain and the TF-mCherry strains).

880 To clone the synthetic target promoters for the CpxR position sweep
881 regulation experiments, we designed an approach to make fast and efficient
882 cloning of any TF binding sequence at defined locations ranging from +1
883 to -112bp relative to the TSS on the unregulated DNA circuit (P_{DL5}). We
884 designed forward and reverse primers to amplify the plasmid at defined lo-
885 cations in the promoter sequence. These primers were used to insert a *ccdB*
886 cassette at the precise location upstream of the gene circuit and had over-
887 hangs for the typeIIIs BbsI restriction site. This allowed for excision of the
888 *ccdB* cassette and cloning of BbsI digested TF binding sites that had com-
889plementary overhangs to excised region. The $P_{DL5} - ccdB$ plasmids were
890 assembled and transformed into *Escherichia coli* DB3.1 strain that harbors
891 key mutations in DNA gyrase that tolerates the *ccdB* toxin [86] for stock
892 cataloguing and sequencing. The plasmids were then incubated with double

893 stranded oligos that had the TF binding sites flanked with the Bbs1 restric-
894 tion sites and the 4 bp complementary sequence to the digested plasmid.
895 This approach allowed the incubation to be a one-step Digestion and Liga-
896 tion reaction to excise the *ccdB* cassette and insert the TF binding sequence
897 at the defined locus. Stock curation, sequencing, and final transformation
898 into the relevant strains were done as for the other TF-position constructs.
899 All primers and TF binding sequences used for the cloning of the synthetic
900 target genes are listed in the Supplement Section.

901 **Engineering the titratable TF-mCherry fusion strains**

902 All strains used in this study are constructed from the parent strain *E.*
903 *coli* MG1655. The TFs investigated in this study include AcrR, AgaR, ArsR,
904 AscG, BetI, and CpxR. Each TF gene is deleted from its wildtype locus and
905 expressed from the *ybcN* locus under the regulation of a Ptet promoter.
906 The autofluorescence strain for each experiment is *E. coli* MG1655 with the
907 corresponding TF knocked out from the wildtype locus.

908 We used the KEIO library [87] as the starting point for the construction
909 of the 6 TFs with the titratable TF-mCherry fusion cassette. We selected the
910 corresponding clone from the KEIO library with the TF-knockout (the coding
911 and upstream regions of the TF gene are replaced by a constitutive promoter
912 expressing Kanamycin), and deletion of TF gene from the wildtype locus was
913 performed by P1 transduction of the corresponding knockout from the KEIO
914 collection to the MG1655 *E. coli* strain. The kanamycin cassette was cured
915 using the *frt* flippase expressed from *pCP20* plasmid. The TF ORF(open
916 reading frame) was fused to mCherry with an AEK linker sequence by PCR
917 amplifying the ORF from the MG1655 strain and cloning into the plasmid
918 pZs3 containing the Ptet promoter with the AEK-mCherry cassette down-
919 stream. The Ptet TF-mCherry fusion construct was then amplified by *ybcN*
920 integration primers for chromosomal insertion. The target gene for inte-
921 gration is constructed by SOEing PCR using the primer pairs listed in the
922 Supplemental Section. Chromosomal integration of the TF and target are
923 performed by lambda red recombineering assisted by plasmid *pKM208* as
924 described in [85]. The strain we used for the integration of the TF-mCherry
925 fusion cassette had the inducer construct (constitutive promoter expressing
926 TetR) already integrated at the *gspI* locus. After integration, we sequenced
927 the regions surrounding the TetR promoter and the TF-Linker-mCherry cas-
928 sette to confirm the regions were free of any mutations. We then catalogued

929 and stored these strains before making the synthetic circuit measurements
930 as presented in the Main Text.

931 References

- 932 [1] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis
933 of proteins. *J. Mol. Biol.*, 3:318–356, 1961.
- 934 [2] M. Ptashne and A. Gann. *Genes and Signals*. Cold Spring Harbor
935 Laboratory Press, New York, 2002.
- 936 [3] N. J. Guido, X. Wang, D. Adalsteinsson, et al. A bottom-up approach
937 to gene regulation. *Nature*, 439(7078):856–60, 2006.
- 938 [4] D. Sprinzak and M. B. Elowitz. Reconstruction of genetic circuits. *Nature*,
939 438(7067):443–8, 2005.
- 940 [5] M. R. Atkinson, M. A. Savageau, J. T. Myers, and A. J. Ninfa. Develop-
941 ment of genetic circuitry exhibiting toggle switch or oscillatory behavior
942 in *Escherichia coli*. *Cell*, 113(5):597–607, 2003.
- 943 [6] A. A. Nielsen, B. S. Der, J. Shin, et al. Genetic circuit design automa-
944 tion. *Science*, 352(6281):aac7341, 2016.
- 945 [7] M. Z. Ali, V. Parisutham, S. Choubey, and R. C. Brewster. Inherent reg-
946 ulatory asymmetry emanating from network architecture in a prevalent
947 autoregulatory motif. *Elife*, 9, 2020.
- 948 [8] G. D. Stormo. DNA binding sites: representation and discovery. *Bioin-*
949 *formatics*, 16(1):16–23, 2000.
- 950 [9] X. Messeguer, R. Escudero, D. Farre, et al. PROMO: detection of known
951 transcription regulatory elements using species-tailored searches. *Bioin-*
952 *formatics*, 18(2):333–334, 2002.
- 953 [10] W. W. Wasserman and A. Sandelin. Applied bioinformatics for the
954 identification of regulatory elements. *Nat. Rev. Genet.*, 5(4):276–287,
955 2004.
- 956 [11] M. T. Weirauch, A. Yang, M. Albu, et al. Determination and inference
957 of eukaryotic transcription factor sequence specificity. *Cell*, 158(6):1431–
958 1443, 2014.

- 959 [12] M. Vidal, M. E. Cusick, and A. L. Barabasi. Interactome networks and
960 human disease. *Cell*, 144(6):986–998, 2011.
- 961 [13] T. Rolland, M. Tasan, B. Charlotiaux, et al. A proteome-scale map of
962 the human interactome network. *Cell*, 159(5):1212–1226, 2014.
- 963 [14] E. Bateman. Autoregulation of eukaryotic transcription factors. *Prog.*
964 *Nucleic Acid Res. Mol. Biol.*, 60:133–168, 1998.
- 965 [15] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in
966 the transcriptional regulation network of *Escherichia coli*. *Nat Genet*,
967 31(1):64–8, 2002.
- 968 [16] S. Klumpp, Z. Zhang, and T. Hwa. Growth rate-dependent global effects
969 on gene expression in bacteria. *Cell*, 139(7):1366–75, 2009.
- 970 [17] A. Schmidt, K. Kochanowski, S. Vedelaar, et al. The quantitative
971 and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.*,
972 34(1):104–110, 2016.
- 973 [18] R. C. Brewster, F. M. Weinert, H. G. Garcia, et al. The transcription
974 factor titration effect dictates level of gene expression. *Cell*, 156(6):1312–
975 1323, 2014.
- 976 [19] T. H. Lee and N. Maheshri. A regulatory role for repeated decoy tran-
977 scription factor binding sites in target gene expression. *Mol. Syst. Biol.*,
978 8:576, 2012.
- 979 [20] S. Oehler, E. R. Eismann, H. Kramer, and B. Muller-Hill. The three
980 operators of the lac operon cooperate in repression. *EMBO J*, 9(4):973–
981 9, 1990.
- 982 [21] H. G. Garcia and R. Phillips. Quantitative dissection of the simple
983 repression input-output function. *Proceedings of the National Academy*
984 *of Sciences of the U.S.A.*, 108(29):12174–82, 2011.
- 985 [22] P. Hammar, M. Wallden, D. Fange, et al. Direct measurement of tran-
986 scription factor dissociation excludes a simple operator occupancy model
987 for gene regulation. *Nat. Genet.*, 46(4):405–408, 2014.
- 988 [23] R. Daber, M. A. Sochor, and M. Lewis. Thermodynamic analysis of
989 mutant lac repressors. *J. Mol. Biol.*, 409(1):76–87, 2011.

- 990 [24] Jose M. G. Vilar and Leonor Saiz. Reliable Prediction of Complex Phe-
991 nototypes from a Modular Design in Free Energy Space: An Extensive
992 Exploration of the lac Operon. *ACS Synthetic Biology*, 2(10):576–586,
993 2013.
- 994 [25] R. B. Lobell and R. F. Schleif. DNA looping and unlooping by AraC
995 protein. *Science*, 250(4980):528–532, 1990.
- 996 [26] Robert Schleif. AraC protein, regulation of the l-arabinose operon in
997 Escherichia coli, and the light switch mechanism of AraC action. *FEMS*
998 *Microbiology Reviews*, 34(5):779–796, 2010.
- 999 [27] S. M. Egan, A. J. Pease, J. Lang, et al. Transcription activation by
1000 a variety of AraC/XylS family activators does not depend on the class
1001 II-specific activation determinant in the N-terminal domain of the RNA
1002 polymerase alpha subunit. *J Bacteriol*, 182(24):7075–7077, 2000.
- 1003 [28] S. M. Egan. Growing repertoire of AraC/XylS activators. *J Bacteriol*,
1004 184(20):5529–5532, 2002.
- 1005 [29] M. Ptashne, A. Jeffrey, A. D. Johnson, et al. How the lambda repressor
1006 and cro work. *Cell*, 19(1):1–11, 1980.
- 1007 [30] M. A. Shea and G. K. Ackers. The OR control system of bacteriophage
1008 lambda. A physical-chemical model for gene regulation. *J Mol Biol*,
1009 181(2):211–30, 1985.
- 1010 [31] Mark Ptashne. *A genetic switch: phage lambda revisited*. Cold Spring
1011 Harbor Laboratory Press, Cold Spring Harbor, N.Y., 3rd edition, 2004.
- 1012 [32] Juan L. Ramos, Manuel Martínez-Bueno, Antonio J. Molina-Henares,
1013 et al. The TetR Family of Transcriptional Repressors. *Microbiology and*
1014 *Molecular Biology Reviews*, 69(2):326–356, 2005.
- 1015 [33] Wanyan Deng, Chunmei Li, and Jianping Xie. The underling mecha-
1016 nism of bacterial TetR/AcrR family transcriptional repressors. *Cellular*
1017 *Signalling*, 25(7):1608 – 1613, 2013.
- 1018 [34] B. C. Stanton, A. A. Nielsen, A. Tamsir, et al. Genomic mining
1019 of prokaryotic repressors for orthogonal logic gates. *Nat Chem Biol*,
1020 10(2):99–105, 2014.

- 1021 [35] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic
1022 toggle switch in *Escherichia coli*. *Nature*, 403(6767):339–42, 2000.
- 1023 [36] J. C. Anderson, C. A. Voigt, and A. P. Arkin. Environmental signal
1024 integration by a modular AND gate. *Mol Syst Biol*, 3:133, 2007.
- 1025 [37] C. Tan, P. Marguet, and L. You. Emergent bistability by a growth-
1026 modulating positive feedback circuit. *Nat Chem Biol*, 5(11):842–848,
1027 2009.
- 1028 [38] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of tran-
1029 scriptional regulators. *Nature*, 403(6767):335–8, 2000.
- 1030 [39] J. Stricker, S. Cookson, M. R. Bennett, et al. A fast, robust and tunable
1031 synthetic gene oscillator. *Nature*, 456(7221):516–519, 2008.
- 1032 [40] R. Amit, H. G. Garcia, R. Phillips, and S. E. Fraser. Building enhancers
1033 from the ground up: a synthetic biology approach. *Cell*, 146(1):105–118,
1034 2011.
- 1035 [41] M. Brunwasser-Meirom, Y. Pollak, S. Goldberg, et al. Using synthetic
1036 bacterial enhancers to reveal a looping-based mechanism for quenching-
1037 like repression. *Nat Commun*, 7:10407, 2016.
- 1038 [42] W. S. Kontur, R. M. Saecker, M. W. Capp, and M. T. Record. Late
1039 steps in the formation of E. coli RNA polymerase-lambda P R promoter
1040 open complexes: characterization of conformational changes by rapid
1041 [perturbant] upshift experiments. *J Mol Biol*, 376(4):1034–1047, 2008.
- 1042 [43] A. Revyakin, C. Liu, R. H. Ebright, and T. R. Strick. Abortive initiation
1043 and productive initiation by RNA polymerase involve DNA scrunching.
1044 *Science*, 314(5802):1139–43, 2006.
- 1045 [44] B. A. Young, T. M. Gruber, and C. A. Gross. Views of transcription
1046 initiation. *Cell*, 109(4):417–420, 2002.
- 1047 [45] K. L. Henderson, L. C. Felth, C. M. Molzahn, et al. Mechanism of
1048 transcription initiation and promoter escape by E. coli RNA polymerase.
1049 *Proc Natl Acad Sci U S A*, 114(15):E3032–E3040, 2017.

- 1050 [46] G. K. Ackers, A. D. Johnson, and M. A. Shea. Quantitative model for
1051 gene regulation by lambda phage repressor. *Proc Natl Acad Sci U S A*,
1052 79(4):1129–33, 1982.
- 1053 [47] T. Kuhlman, Z. Zhang, Jr. Saier, M. H., and T. Hwa. Combinatorial
1054 transcriptional control of the lactose operon of *Escherichia coli*. *Proceed-*
1055 *ings of the National Academy of Sciences of the U.S.A.*, 104(14):6043–8,
1056 2007.
- 1057 [48] L. Bintu, N. E. Buchler, H. G. Garcia, et al. Transcriptional regulation
1058 by the numbers: models. *Curr Opin Genet Dev*, 15(2):116–24, 2005.
- 1059 [49] J. B. Kinney, A. Murugan, C. G. Callan, and E. C. Cox. Using deep se-
1060 quencing to characterize the biophysical mechanism of a transcriptional
1061 regulatory sequence. *Proc. Natl. Acad. Sci. U.S.A.*, 107(20):9158–9163,
1062 2010.
- 1063 [50] N. E. Buchler, U. Gerland, and T. Hwa. On schemes of combinatorial
1064 transcription logic. *Proc Natl Acad Sci U S A*, 100(9):5136–41, 2003.
- 1065 [51] J. M. Vilar and S. Leibler. DNA looping and physical constraints on
1066 transcription regulation. *Journal of Molecular Biology*, 331(5):981–9,
1067 2003.
- 1068 [52] H. G. Garcia, A. Sanchez, J. Q. Boedicker, et al. Operator sequence
1069 alters gene expression independently of transcription factor occupancy
1070 in bacteria. *Cell Rep*, 2(1):150–161, 2012.
- 1071 [53] R. Phillips, N. M. Belliveau, G. Chure, et al. Figure 1 Theory Meets Fig-
1072 ure 2 Experiments in the Study of Gene Expression. *Annu Rev Biophys*,
1073 48:121–163, 2019.
- 1074 [54] R. C. Brewster, D. L. Jones, and R. Phillips. Tuning Promoter Strength
1075 through RNA Polymerase Binding Site Design in *Escherichia coli*. *PLoS*
1076 *Computational Biology*, 8(12):e1002811, 2012.
- 1077 [55] T. Lamark, I. Kaasen, M. W. Eshoo, et al. DNA sequence and analysis
1078 of the bet genes encoding the osmoregulatory choline-glycine betaine
1079 pathway of *Escherichia coli*. *Mol Microbiol*, 5(5):1049–1064, 1991.

- 1080 [56] S. Gama-Castro, H. Salgado, A. Santos-Zavaleta, et al. RegulonDB
1081 version 9.0: high-level integration of gene regulation, coexpression, motif
1082 clustering and beyond. *Nucleic Acids Res*, 44(D1):D133–143, 2016.
- 1083 [57] Heladia Salgado, Martin Peralta-Gil, Socorro Gama-Castro, et al. Reg-
1084 ulonDB v8.0: omics data sets, evolutionary conservation, regulatory
1085 phrases, cross-validated gold standards and more. *Nucleic Acids Res.*,
1086 41(D1):D203–D213, 2013.
- 1087 [58] I. M. Keseler, A. Mackie, A. Santos-Zavaleta, et al. The EcoCyc
1088 database: reflecting new knowledge about Escherichia coli K-12. *Nu-
1089 cleic Acids Res.*, 45(D1):D543–D550, 2017.
- 1090 [59] M. Ptashne and A. Gann. Imposing specificity by localization: mecha-
1091 nism and evolvability. *Curr Biol*, 8(22):R812–22, 1998.
- 1092 [60] T. L. Forcier, A. Ayaz, M. S. Gill, et al. Measuring cis-regulatory ener-
1093 getics in living cells using allelic manifolds. *Elife*, 7, 2018.
- 1094 [61] J. Müller, S. Oehler, and B. Müller-Hill. Repression of *lac* promoter as
1095 a function of distance, phase and quality of an auxiliary *lac* operator. *J
1096 Mol Biol*, 257(1):21–9, 1996.
- 1097 [62] E. Sharon, Y. Kalma, A. Sharp, et al. Inferring gene regulatory logic
1098 from high-throughput measurements of thousands of systematically de-
1099 signed promoters. *Nat. Biotechnol.*, 30(6):521–530, 2012.
- 1100 [63] S. Weingarten-Gabbay and E. Segal. The grammar of transcriptional
1101 regulation. *Hum Genet*, 133(6):701–711, 2014.
- 1102 [64] J. Gertz, E. D. Siggia, and B. A. Cohen. Analysis of combinatorial cis-
1103 regulation in synthetic and genomic promoters. *Nature*, 457(7226):215–
1104 8, 2009.
- 1105 [65] O. V. Bylino, A. N. Ibragimov, and Y. V. Shidlovskii. Evolution of
1106 Regulated Transcription. *Cells*, 9(7), 2020.
- 1107 [66] W. Ross and R. L. Gourse. Analysis of RNA polymerase-promoter com-
1108 plex formation. *Methods*, 47(1):13–24, 2009.

- 1109 [67] S. B. Smith, Y. Cui, and C. Bustamante. Overstretching B-DNA: the
1110 elastic response of individual double-stranded and single-stranded DNA
1111 molecules. *Science*, 271(5250):795–9, 1996.
- 1112 [68] M. Monsalve, B. Calles, M. Mencía, et al. Transcription activation or
1113 repression by phage psi 29 protein p4 depends on the strength of the
1114 RNA polymerase-promoter interactions. *Mol Cell*, 1(1):99–107, 1997.
- 1115 [69] D. Jensen, A. R. Manzano, J. Rammohan, et al. CarD and RbpA
1116 modify the kinetics of initial transcription and slow promoter escape of
1117 the Mycobacterium tuberculosis RNA polymerase. *Nucleic Acids Res*,
1118 47(13):6685–6698, 2019.
- 1119 [70] V. A. Rhodius, D. M. West, C. L. Webster, et al. Transcription activa-
1120 tion at class II CRP-dependent promoters: the role of different activating
1121 regions. *Nucleic Acids Res*, 25(2):326–332, 1997.
- 1122 [71] W. Niu, Y. Kim, G. Tau, et al. Transcription activation at class II
1123 CAP-dependent promoters: two interactions between CAP and RNA
1124 polymerase. *Cell*, 87(6):1123–1134, 1996.
- 1125 [72] Y. Zhou, A. Kolb, S. J. Busby, and Y. P. Wang. Spacing requirements for
1126 Class I transcription activation in bacteria are set by promoter elements.
1127 *Nucleic Acids Res*, 42(14):9209–9216, 2014.
- 1128 [73] N. J. Savery, G. S. Lloyd, S. J. Busby, et al. Determinants of the C-
1129 terminal domain of the Escherichia coli RNA polymerase alpha sub-
1130 unit important for transcription at class I cyclic AMP receptor protein-
1131 dependent promoters. *J Bacteriol*, 184(8):2273–2280, 2002.
- 1132 [74] A. Kinkhabwala and C. C. Guet. Uncovering cis regulatory codes using
1133 synthetic promoter shuffling. *PLoS One*, 3(4):e2030, 2008.
- 1134 [75] M. A. Beer and S. Tavazoie. Predicting gene expression from sequence.
1135 *Cell*, 117(2):185–198, 2004.
- 1136 [76] W. T. Ireland, S. M. Beeler, E. Flores-Bautista, et al. Deciphering the
1137 regulatory genome of Escherichia coli, one hundred promoters at a time.
1138 *Elife*, 9, 2020.

- 1139 [77] C. Scholes, A. H. DePace, and A. Sanchez. Combinatorial Gene Reg-
1140 ulation through Kinetic Control of the Transcription Cycle. *Cell Syst*,
1141 4(1):97–108, 2017.
- 1142 [78] Rosa Martinez-Corral, Minhee Park, Kelly Biette, et al. Transcriptional
1143 kinetic synergy: a complex landscape revealed by integrating modelling
1144 and synthetic biology. *bioRxiv*, 2020.
- 1145 [79] F. Wong and J. Gunawardena. Gene Regulation in and out of Equilib-
1146 rium. *Annu Rev Biophys*, 49:199–226, 2020.
- 1147 [80] Gabriel Rosenblum, Nadav Elad, Haim Rozenberg, et al. Allosteric
1148 through DNA drives phenotype switching. *bioRxiv*, 2020.
- 1149 [81] S. Kim, E. Brostromer, D. Xing, et al. Probing allostery through DNA.
1150 *Science*, 339(6121):816–819, 2013.
- 1151 [82] N. Rosenfeld, J. W. Young, U. Alon, et al. Accurate prediction of gene
1152 feedback circuit behavior from component properties. *Mol Syst Biol*,
1153 3:143, 2007.
- 1154 [83] M. Razo-Mejia, S. L. Barnes, N. M. Belliveau, et al. Tuning Transcrip-
1155 tional Regulation through Signaling: A Predictive Theory of Allosteric
1156 Induction. *Cell Syst*, 6(4):456–469, 2018.
- 1157 [84] M. D. Hoffman and A. Gelman. The No-U-Turn Sampler: Adaptively
1158 Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine*
1159 *Learning Research*, 15:1351–1381, 2014.
- 1160 [85] K. C. Murphy and K. G. Campellone. Lambda Red-mediated recom-
1161 binogenic engineering of enterohemorrhagic and enteropathogenic E.
1162 coli. *BMC Mol. Biol.*, 4:11, 2003.
- 1163 [86] P. Bernard. Positive selection of recombinant DNA by CcdB. *BioTech-*
1164 *niques*, 21(2):320–323, 1996.
- 1165 [87] T. Baba, T. Ara, M. Hasegawa, et al. Construction of Escherichia coli
1166 K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol*
1167 *Syst Biol*, 2:2006 0008, 2006.

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 middle column. The rightmost column shows the transcription rates in these states. (B) Fold-change vs TF copy number (N_{TF}) for 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 ($\Delta\epsilon_{TF}$, 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 knocked out and reintroduced as a TF-mCherry fusion at the *ybcN* 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 systematically controlled. (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. Here we have converted the arbitrary mCherry fluorescence signal into number of TFs using fluctuation counting methods detailed further in the SI.

Figure 3: **Regulatory curves for individual TFs.** 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-FP 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. 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).

Figure 4: **Fold-change for a given regulatory location vs N_{eff} .** The regulatory curve for all TFs when acting at (A) +1 and (B) -61 plotted against $N_{eff} = N_{TF} \exp(-\Delta\epsilon_{TF})$. 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 is 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*. Note that the length of the rectangles represents the span of the binding sequence and the border color represents activation (blue) or repression (red). A majority of the binding sites are centered between the -40 to -80 positions. (B) Schematic strategy for constructing and measuring CpxR acting at a specified binding site(the *ppxA* binding sequence) inserted at 21 upstream positions and 1 downstream position on the promoter. (C) 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 center 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 10.5bp 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). The black points represent the mean and standard error of these replicates. The dashed line running through the data points is the theory prediction based on inference from the model detailed in Figure 6. The shaded regions represent ± 2 standard deviations of the inferred model parameters.

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 position manifold curves are predicted to be straight lines when β is small, but curved for larger values of β (C) Position 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 generated from the inferred acceleration and stabilization parameters and the shaded regions represent ± 2 standard deviations from the mean. (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. (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 position manifold plots in (B).

Binding Position	α	β
-64	24.911 ^{31.849} _{14.911}	29.545 ^{44.169} _{12.984}
-48	3.506 ^{4.008} _{2.898}	116.257 ^{150.131} _{66.453}
-60	2.161 ^{2.33} _{1.974}	219.757 ^{270.481} _{151.499}
-70	6.995 ^{9.748} _{3.101}	6.224 ^{9.316} _{1.776}
-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}
-50	0.675 ^{0.736} _{0.612}	161.766 ^{199.140} _{108.281}
-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}
-54	0.049 ^{0.051} _{0.047}	34.496 ^{43.447} _{23.154}

Table 1: **Inferred acceleration and stabilization parameters.** Median values of the inference chain along with the bounds that encompass the 68%th percent Bayesian credible interval of the parameters α and β . The “weak stabilization limit” limits the precision of inference estimates for positions -58, -70, and -74.