The Role of Binding Sequence, Position, and Promoter Strength on the Regulatory Modes of E. coli Transcription

THE ROLE OF BINDING SEQUENCE, POSITION, AND PROMOTER STRENGTH ON THE REGULATORY MODES OF *E.COLI* TRANSCRIPTION FACTORS

A Dissertation Presented

By

SUNIL GUHARAJAN

Submitted to the Faculty of the

Morningside Graduate School of Biomedical Sciences at UMass Chan Medical School

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 11th , 2023

Systems Biology

THE ROLE OF BINDING SEQUENCE, POSITION, AND PROMOTER STRENGTH ON THE REGULATORY MODES OF *E.COLI* TRANSCRIPTION FACTORS

A Dissertation Presented

By

Sunil Guharajan

This work was undertaken in the Morningside Graduate School of Biomedical Sciences

Program in Bioinformatics and Computational Biology

Under the mentorship of

Robert C. Brewster, Ph.D., Thesis Advisor

Nick Rhind, Ph.D., Member of Committee

Hyun Youk, Ph.D., Member of Committee

Amir Mitchell, Ph.D., Member of Committee

Jane Kondev, Ph.D., External Member of Committee

A.J. Marian Walhout, Ph.D., Chair of Committee

Mary Ellen Lane, Ph.D., Dean of the Morningside Graduate School of Biomedical Sciences

December 11th , 2023

Acknowledgements

This work set before you could not have been done without the immeasurable support of several people.

I am grateful to Robert C. Brewster for accepting me into his lab and giving me excellent training in quantitative biology. My time here has been one of continual improvement, and he has given me ample leeway to hone the skills that I need to contribute to the burgeoning field of synthetic biology. I can truly say that he has been an able instructor in rigorous thinking and scientific conduct, and this time has given me the confidence to excel as an independent scientist.

My tenure as a PhD student would have been far less productive if it was not for the exceptional mentorship of Vinuselvi Parisutham. Her scientific acumen and experimental skills are something I aspire to, and I could not have become a better experimentalist without her guidance and instruction.

The most engaging times in lab were often spent troubleshooting failed experiments, or discussing mathematical models with Zulfikar Ali – a.k.a Zulu. Zulu has always been supportive and encouraging of my foray into mathematical biology, and his patience in pointing out the many missteps I have made while modelling gene expression is nothing short of exemplary.

The PhD experience would not have been complete without a group of smart, funny, and like-minded individuals, and for that I thank the following people for their friendship.

First, my academic career would have been quite different if I had not met Welles Chan in the Spring of 2009 at Hardin-Simmons University. What started out as Friday night jam sessions in his dorm room began a longstanding friendship that has spanned more than a decade. He convinced me to come to UMASS CHAN Medical School, and his continual advice in both scientific and personal areas of my life has enriched me in more ways than I would admit.

While I have made many friends in the crucible of my PhD journey, I would like to thank Jason Fan, Kris Holloway, Stephanie Becker, and Debanjan Goswamy who have made my time here worth it. There is something about a shared experience, and this is particularly true of the ups and downs of our academic journey together. As we go our separate ways to a brighter future, I will look back fondly on those Dairy Queen runs and think of the celebrations we shared from getting through our post qualifying exams.

Lastly, I am deeply indebted to my family, who mean the world to me. My parents instilled in my brothers and I the importance of hard and honest work. The completion of this thesis has been a culmination of their dedication, instruction, and will to see us do what we love. I end this acknowledgment by saying that this milestone in my academic journey would not have been possible without them.

Abstract

The activity of Transcription Factors (TFs) is essential to gene regulation, facilitating the transfer of information encoded in cis-regulatory elements into temporal and spatial control of gene expression. Despite progress in mapping the genomic binding sites of TFs, the regulatory role of TFs once bound is often convoluted. Here, we utilize a synthetic biology approach that allows for precise manipulation of TF concentration *in vivo* and interrogation of the role of promoter features on the regulatory function of *E.coli* TFs. Using thermodynamic models of gene regulation, we decouple TF occupancy from its maximal regulatory output and probe its regulation on two steps of the transcription process: stabilization of RNA polymerase (RNAP) binding and modulation of transcription initiation. Profiling the CpxR activator, we discover universal stabilization across regulated positions, with differences in strong and weak activation set primarily by regulation of the initiation rate. Formulating a high-throughput approach, we probe the regulation of 93 *E.coli* TFs to find that the relative contributions of binding sequence and position on the mode of action are decoupled, with position playing a dominant role for select factors. Building on this information, we assessed the interplay between binding position, TF identity, and basal promoter strength - uncovering a conserved mode of stabilizing regulation across TFs with diverse regulatory outcomes. Taken together, our work delineates the effect of promoter architecture on the quantitative regulatory activity of TFs, with implications for design of synthetic gene circuits and understanding natural promoters

Table of Contents

List of Figures and Tables

List of Copyrighted Materials

Guharajan, S., Chhabra, S., Parisutham, V., & Brewster, R. C. (2021). Quantifying the regulatory role of individual transcription factors in Escherichia coli. *Cell Rep, 37*(6), 109952. doi:10.1016/j.celrep.2021.109952

List of Abbreviations

RNAP – RNA polymerase

E.coli – Escherichia coli

MCMC – Markov Chain Monte Carlo

TSS – Transcription Start Site

TRN – Transcription Regulatory Networks

Chapter I - Introduction

1.1 - Gene Regulation in Prokaryotes

Gene regulation is the fundamental process by which an organism shapes the intracellular concentration of proteins - essential in nearly every facet of cellular biology. Responding to stress, developing specialized cells through differentiation, and ensuring efficient nutrient uptake and metabolism all rely on precise control of pathways. One of the pioneering studies of gene regulation by François Jacob and Jacques Monod (Arthur, François, & Jacques, 1959) was an investigation into a curious decision made by the model bacteria *E. coli* in deciding which sugar to consume from its media. They observed that *E. coli* growing in media containing a mix of two sugars. a preferred (glucose) and less optimal (lactose) one, did not consume the sugars proportionately during growth, but consumed glucose until it was depleted before switching to the less preferred carbon source(Arthur et al., 1959; Hall, 1989; Moses & Prevost, 1966). In the interim before the switch is accomplished a pause in growth occurs as cells adjusts to consume the new carbon source. This phenomenon, dubbed "diauxic growth," led to the discovery that this pause is due to the production a different enzyme to consume the new sugar(Jacques, Jean-Pierre, & François, 1963). The enzyme β-galactosidase, necessary for cleaving lactose for glycolysis, had to have its expression tightly controlled (Moses & Prevost, 1966) as production of the enzyme in an abundance of the primary carbon source would translate to poor management of cellular resources (Pinto, Melo-Miranda, Gordo, & Sousa, 2021; Wilson, Zhan, Swint-Kruse, & Matthews, 2007). This hypothesis, that the

delay seen in the diauxic shift was the result of a change in enzyme identity(François & Jacques, 1961) rather than a repurposing of an existing, active enzyme, was confirmed over the subsequent decades when the genetics of the lac operon were thoroughly elucidated(Brückner & Titgemeyer, 2002; Görke & Stülke, 2008; Hirsh & Schleif, 1973; Stülke & Hillen, 1999; Wanner, Kodaira, & Neidhardt, 1978). These investigations demonstrated that control of the operon involved a balance between a negative regulator of transcription (LacI) and a positive regulator (CRP) and that activity of these regulators was closely tied to the nutrient status of the cell(Stülke & Hillen, 1999; Wanner et al., 1978). This work uncovering the effect of these regulatory proteins, called Transcription Factors (TFs) set a foundation for understanding how cells make simple regulatory decisions.

The discovery that genes could be regulated by TFs spurred attempts to determine the extent to which the expressed genome is regulated in this manner. It is now clear through dissections of promoters from several model organism that TFs play an integral role in control of gene expression, with Monod's hypothesis on the separation of the genome into "structural elements" and portions dedicated to regulating the timing and magnitude of expression of these elements (François & Jacques, 1961) demonstrated across all clades of life (Friedman, 1988; Utley et al., 1998). The mechanisms of this regulation may features that are specific to prokaryotes and eukaryotes, with prokaryotic gene regulation often centered on the operon: a cluster of genes under the control of a common promoter and regulatory elements (Hall, 1989; Okuda et al., 2007) involved in pathways that range from control and organization of cell shape, assembling specific molecular

rotors that generate cell motility, to the synthesis and breakdown of metabolites necessary for cellular homeostasis(Gitai, Dye, & Shapiro, 2004; Ishihama, 2012; Osbourn & Field, 2009; Shih & Rothfield, 2006),. The key feature of the operon involves the grouping of several functionally related genes on a single mRNA (polycistronic transcript) that allows for collective regulation. By constraining several different genes under the control of a single promoter element, the timing of gene expression can be synchronized between related genes, allowing for precise control of gene products necessary for cellular function(Berthoumieux et al., 2013; English, Gayet, & Collins, 2021; Okuda et al., 2007; Rubinstein, Zeevi, Oren, Segal, & Pupko, 2011). As operons form the crux of the prokaryotic transcriptomic response, its distribution throughout the genome is not random, but is concentrated in key regions of the chromosome in a manner that effectively balances cross-talk between the complementary cellular processes of transcription and replication – with evidence of strong selective pressure on the ordering of operons on the leading DNA strand (Lim, Lee, & Hussein, 2011; Omont & Képès, 2004; Rocha & Danchin, 2003). This point highlights the concept of the operon as a versatile medium of genetic change. The preferential clustering of genetic elements and grouping of functionally related genes also facilitates the transfer of genetic material between bacterial strains capable of coding for vital cellular processes, making horizontal gene transfer between bacterial species more efficient and in the process driving evolutionary change in a fundamental way (Hinman & Cary, 2017; Kominek et al., 2019; van Hijum, Medema, & Kuipers, 2009). Indeed, functional interrogation of operons and computational evidence suggests that regulatory elements and structural genes of

common operons are shared between distinct species, providing evidence for a common evolutionary etiology (Bundalovic-Torma, Whitfield, Marmont, Howell, & Parkinson, 2020). However, the ordering of the genes is often configured differently between these species, and along with changes to the regulatory elements modulating the expression of these operons, suggests strong adaptive preferences in organization and regulation contributing to diversity amongst different species (Hazkani-Covo & Graur, 2005; Okuda et al., 2007; Omelchenko, Makarova, Wolf, Rogozin, & Koonin, 2003; Price, Arkin, & Alm, 2006).

Regulation of bacterial operons, once considered to be simple with regards to the number of TFs that co-regulate distinct promoters, has since had this idea after studies probing the occupancy of distinct TFs across the genome. Several studies utilizing occupancy profiling techniques such as Genomic-SELEX have uncovered bacterial promoters regulated by sizeable numbers of individual transcription factors (Ishihama, 2012; Keseler et al., 2021; Salgado et al., 2023). The constellation of factors at these promoters presumably allows for combinatorial utilization of distinct TFs that shapes promoter response in different induction conditions experienced by the cell. Indeed, this combinatorial input from several TFs, much like the interplay of distinct regulators observed in eukaryotes, might be harnessed to accomplish several kinds of regulatory outputs (English et al., 2021; Patange, Girvan, & Larson, 2018). On one hand, the distinct configurations of TF binding sites might be used to set a high threshold of specificity regarding an expression state: the on-off function of the promoter might be tied to these configurations and different cellular states ultimately set the switch by determining the

composition of the TFs expressed in a cell that interact with a given promoter (Golding, Paulsson, Zawilski, & Cox, 2005; McAdams & Arkin, 1997). A second possibility is that the distinct regulators interacting at a given promoter might be required to set a graded response in promoter activity, allowing for a continuum of expression to fine-tune the composition of various proteins necessary to achieve a cellular state (Ackers, Johnson, & Shea, 1982; Dekel & Alon, 2005; Rubinstein et al., 2011).

1.2- Stochasticity in Gene Regulation: Implications for Regulatory Control of Promoters

Control over gene regulation by TFs has implications not just for steady state protein levels, but also for noise in gene expression, where the distribution of expression states mediated through the regulatory architecture of the promoter has consequences on cellular fitness and adaptability (Ferguson et al., 2012; Vlková & Silander, 2022). Cellular processes such as differentiation and metabolic adaptation are affected by stochastic expression of key genes necessary to carry out these functions. representing a challenge for the cells to faithfully execute these highly coordinate processes(Boettiger & Levine, 2009; Chang, Hemberg, Barahona, Ingber, & Huang, 2008; Coomer, Ham, & Stumpf, 2022; MacArthur et al., 2012). Far from functioning in a negative sense, nosiness can also promote cellular niches in an otherwise isogenic population of cells, effectively hedging the population to fluctuation in environmental conditions(Maamar, Raj, & Dubnau, 2007). Indeed the persistence of noise in gene expression likely has a strong environmental component – environments that are often subject to rapid flux create an incentive for cells to tolerate noise in expression, making a trade-off in the precision of

gene expression for an adaptive protein response essential for cell survival(Raj & van Oudenaarden, 2008).

Stochasticity in gene expression can arise from pervasive changes to the cell brought about by shifts in nutrient availability, temperature, and osmotic conditions, all of which shape the number of functionally competent molecular machines such as RNA polymerase (RNAP) and ribosomes, engendering variability in gene expression(Elowitz, Levine, Siggia, & Swain, 2002; Raj & van Oudenaarden, 2008; So et al., 2011). This phenomenon, known as extrinsic stochasticity, is the component of noise in gene expression characterized due to the variability in the concentration of global cellular machinery and the components of the regulatory architecture.

A second contribution to promoter noise comes from the stochastic kinetics of regulators specific to a given promoter and is termed intrinsic stochasticity to contrast its contribution to noise from the global environment of the cell (extrinsic stochasticity). Intrinsic noise in gene expression is primarily dictated by the biochemical interactions of effector molecules that often occur at low copy numbers, and fluctuations in the number of these molecules due to changes in cellular conditions are often considered to be the source of this noise(Cai, Friedman, & Xie, 2006; Dey, Soltani, & Singh, 2020; Rosenfeld, Young, Alon, Swain, & Elowitz, 2005). One of the necessary conditions for stochastic fluctuations in molecules to propagate changes in cell state, is the timescale in which the fluctuation of these molecules operate. If the fluctuations operate on relatively long timescales with respect to the cell cycle, its effects on the variability of the downstream protein is “washed-out” and the steady state distributions of the regulated molecule will

not demonstrate heterogeneity (Kaern, Elston, Blake, & Collins, 2005; Ozbudak, Thattai, Kurtser, Grossman, & van Oudenaarden, 2002; Raser & O'Shea, 2004).

The differential effect of timescales on noise brought about by these fluctuations can be readily demonstrated in the case of eukaryotic promoter dynamics, where fast and slow timescales emerge based on different regulatory interactions at the promoter. Systemic changes to promoter accessibility brought about by nucleosome remodeling shuttle promoters between transcriptionally on and off states over longer time scales, generating an all or nothing bimodal response(Boeger, Griesenbeck, & Kornberg, 2008; Kaern et al., 2005; Tan et al., 2021). In contrast, the binding and unbinding rates of transcription factors often occur on timescales that are a fraction of the cell cycle, creating fast transitions between promoter states, leading to increased heterogeneity of effectors responsible for determining cell state(Kaern et al., 2005; Raser & O'Shea, 2004). This provides the basis for heterogeneity and the emergence of multiple phenotypes within an isogenic population.

Initial investigation into prokaryotic gene expression suggested that both intrinsic and extrinsic noise was largely independent of the regulatory architecture of the promoter (Ferguson et al., 2012; Hensel et al., 2012; Ozbudak et al., 2002; So et al., 2011). This raised an important question: Do cis-acting sequences regulating promoter activity encode information that sets the variability of gene expression? The implications of this question are profound, as it raises the possibility that promoters specify regulatory information at several levels, shaping properties of the distribution of protein states at more than just the level of mean expression and suggests that noise could be a regulated

feature. Indeed, accumulating evidence presents a compelling argument that promoter sequences tune not just the level of expression, but its precision as well (Choi & Kim, 2009; Elowitz et al., 2002; Jones, Brewster, & Phillips, 2014). This raises the hypothesis that evolutionary constraints on promoter architecture do operate at the level of noise and features of the cis-regulatory architecture, such as the operator sequences and interactions of TFs with other regulators, represent control points for evolution through mutation of these sequences. Fundamentally, the control of gene regulation at both the level of mean and variability by TFs is influenced not just by the specific context of the promoter, but also the broader network architecture set by the wiring of the genome(Dey et al., 2020; Raj & van Oudenaarden, 2008).

1.3- A Network View of Gene Regulation

Gene regulation is often conceptualized as a network of interacting factors to delineate how information is processed through these connections (Macneil & Walhout, 2011; Paulsson, 2004; Pedraza & van Oudenaarden, 2005; Shen-Orr, Milo, Mangan, & Alon; Stefan, Zhongge, & Terence, 2009). A "top-down" view of the *E. coli* regulatory network (Ma, Buer, & Zeng, 2004) can reveal the vast interconnectedness of the network, where genes are represented as nodes and the genes that code for TFs connected to their regulated counterparts by lines. This hierarchical view reveals the flow of information in gene regulatory networks and highlights the connectivity of genes and the multi-purposed function of TFs.

The wiring of the Transcriptional Regulatory Networks (TRN) encoded in the genome illustrate the multiple avenues cells may take to exert control over fundamental processes such as cell growth, division, motility, and various metabolic aspects of the cell (Buescher et al., 2012; Fitzgerald, Bonocora, & Wade, 2014; Galagan et al., 2013; Takada, Kijima, Ishiguro, Ishihama, & Shimada, 2023). With the wiring of TRNs now elucidated by the mapping of TFs to their target promoters through expression studies and genome occupancy techniques, emphasis is currently on understanding the fundamental structure of these networks and how it confers functional properties to gene regulation on a systems scale (T. I. Lee et al., 2002). Modelling the connection between regulators and their targets through networks has uncovered the predominance of regulatory motifs – i.e, repeating structures that act as building blocks in the context of the larger tapestry of the genomic TRN – and have emerged as crucial patterns of gene regulation and are conserved from single-cell organisms to the more complex genomic landscapes of metazoans (Babu, Luscombe, Aravind, Gerstein, & Teichmann, 2004; T. I. Lee et al., 2002; Thieffry & Romero, 1999). Importantly, these regulatory motifs can take on several forms, such as the autoregulatory circuit, where a transcription factor regulates target genes including its own expression (Bateman, 1998), and the feed-forward loop that involves a regulator and its downstream promoter simultaneously controlling expression of a target gene (Mangan & Alon, 2003),(Bulcha et al., 2019). Given the predominance of autoregulation in TRNs, investigations probing the regulatory response of this motif uncovered important implications in timing and steady state levels of gene expression (Ali, Parisutham, Choubey, & Brewster, 2020; Rosenfeld, Elowitz, & Alon, 2002),

suggesting that the ubiquity of this motif is related to its functional relevance in gene expression. Indeed, the presence of these motifs mentioned previously at frequencies much higher than what is expected in random networks points to the dependence of the cell on key regulatory functions necessarily constraining the architecture of TRNs (Burda, Krzywicki, Martin, & Zagorski, 2011), delineating design principles prevalent in natural gene networks that could be applied to the engineering of synthetic gene circuits to carry out bespoke activities.

Today, it is known that *E. coli* has nearly 300 TFs which control the expression of some 4000 genes. Nearly 80 years of research has laid out the intricate networks of gene regulation that provide *E. coli* with the ability to efficiently grow in a huge range of nutritional and environmental conditions and persist through stresses (R. Liu et al., 2020; Martínez-Antonio, Janga, & Thieffry, 2008). The extensive annotation pertaining to *E.coli* transcription factors (Keseler et al., 2021; Salgado et al., 2023) coupled with the abundance of gene expression datasets has allowed for determining the degree of modularity within this species gene regulatory network. Indeed, statistical attempts to isolate the fundamental gene network modules responsible for driving expression changes across different conditions have discovered that the *E.coli* architecture is surprisingly modular, with much of the variance in expression associated with specific, individual transcription factors (Sastry et al., 2019). This suggests that systematic characterization of individual transcription factors in *E.coli,* with a focus on elucidating the regulatory principles of promoter control might have merit far beyond individual TF-promoter interactions in delineating global gene regulatory principles.

1.4 - The Kinetic Steps of Transcription

TFs facilitate the process of gene regulation by modulating the accessibility and enzymatic activity of RNA polymerase (RNAP) during transcription. Transcription is a multi-step kinetic process, with several potentially rate limiting steps that provide opportunity for modulation of gene expression. Recruitment of RNAP to its promoter sequence is the primary step: In Eukaryotes, this process requires the assembly of a pre-initiation complex (PIC) with the Mediator Complex and other transcriptional co-activators, leading to a higher effective concentration of RNApolymerase II at promoter regions (Allen & Taatjes, 2015; Poss, Ebmeier, & Taatjes, 2013). Prokaryotic regulation features strong sequence determinants, involved in direct, specific interactions with the RNAP holoenzyme such as the -35, -10 and the UP elements that contact the sigma and α subunits of the holoenzyme respectively(Estrem et al., 1999; Kinney, Murugan, Callan, & Cox, 2010; Murakami, Masuda, Campbell, Muzzin, & Darst, 2002; Siebenlist, Simpson, & Gilbert, 1980; Simpson, 1979). The ability for the sigma subunit of RNAP to associate and dissociate from the core components of the holoenzyme provides specificity and plasticity in the transcriptional response by targeting RNAP to different promoter sequences *(*Feklístov, Sharon, Darst, & Gross, 2014; Paget, 2015; Wade et al., 2006*)*. RNAP at the promoter proceeds through a series of intermediate steps, with the first involving isomerization from the closed to open complex that involves unwinding of DNA(Borukhov & Lee, 2005; X. Y. Li & McClure, 1998). This unwinding precipitates the “transcription bubble” that generates single stranded DNA for a few bases upstream and downstream of the RNAP holoenzyme and facilitates the initial transcription process.

The conformational change into the open complex positions elements of the core enzyme to form a clamp, allowing DNA from the template strand to enter the active site. RNAP does not immediately translocate along DNA, but rather "scrunches" the DNA through the clamp while remaining at its initial bound site and engages in short transcriptional bursts (Kapanidis et al., 2006). These steps form part of a cyclical process: the RNAP holoenzyme starts to engage in initial rounds of abortive transcription, with the number of cycles dependent on the stability of the RNAP-DNA-RNA complex(Goldman, Ebright, & Nickels, 2009; Tang, Roy, Bandwar, Ha, & Patel, 2009). As successive stress is built into the RNAP-DNA-RNA complex through DNA compaction, the contacts anchoring the RNAP holoenzyme to the promoter region are severed, allowing the complex to transition into productive elongation (Revyakin, Liu, Ebright, & Strick, 2006). Several factors play a role in the kinetics of this process, and different sequences alter the speed at which RNAP progresses through these steps (Hsu et al., 2006) - discriminating between weakly and strongly expressed promoters. Importantly, each of these steps can be regulated by TFs, with different factors able to differentially alter the kinetics of recruitment and isomerization of RNAP (Choy et al., 1997; Yankulov, Blau, Purton, Roberts, & Bentley, 1994).

1.5 - The Architecture of Transcriptional Regulation

TFs bind to promoter DNA at specific sequences and interact with RNAP to modulate the transition through the sequence of transcriptional events. Features of the promoter such as TF binding sequence, the position of binding relative to RNAP, and multiplicity of TF sites set the overall occupancy and function of these regulators. The challenge is

understanding how these elements are integrated to impart specific regulatory outputs that facilitate or impede RNAP through the transcriptional process: setting the appropriate level of promoter activity as determined by the state of the cell. Addressing this requirement is crucial for insight into how genes operate in a cellular context, i.e., how genes integrate diverse regulatory cues at the promoter (Harden, Vincent, & DePace, 2023; Martinez-Corral et al., 2023), and to bridge the gap between our ability to "read" sequences and "write" regulatory circuits with specific purpose(Garcia, Brewster, & Phillips, 2016) .

The search to decipher the grammar of promoters (Beer & Tavazoie, 2004; Levo & Segal, 2014; Segal et al., 2006; Seungsoo & Joanna, 2023) has received much attention. Central to this approach is the implicit assumption that the catalogue of realized grammars in natural promoters is sufficient to decode the language of the cis-regulatory code, and that extensive testing of regulatory circuits is key to uncovering this rule set. To this end, multiple lines of interrogation ranging across different cell types and model organisms (Seungsoo & Joanna, 2023; Zhao et al., 2023) have been conducted often by focusing on perturbing short regions around the promoter and assaying the effect on a reporter gene. The increasing affordability of DNA sequencing and synthesis have made possible these "high-throughput" interrogations of promoter architecture that involve testing ever increasing number of promoters that span the combinatorics of the regulatory space (de Boer et al., 2020; Sharon et al., 2012; van Dijk et al., 2017). The designed promoter architectures are modelled on endogenous promoters, systematically perturbing some

aspect of the architecture with the aim of connecting a given regulatory feature to functional changes in gene expression.

Although certain insights into promoter arrangement and broader architecture can be gleaned from high-throughput assays, information on how TFs fundamentally alter gene expression is missed in these approaches. A change in expression due to a perturbed architecture could be attributed to reduced occupancy of a TF or to an alteration in the TFs ability to regulate RNAP at a given checkpoint. This point illustrates a salient conclusion from these investigations: that the relationship between sequence and promoter activity is simply a proxy for the differential activity of TFs in regulating transcription (Martinez-Corral et al., 2023; Scholes, DePace, & Sánchez, 2017). It further suggests that designing promoters with the aim of isolating and evaluating the regulatory function of TFs, as opposed to spanning the gamut of "syntaxes" of the regulatory grammar, would be useful in a "bottoms-up" approach to understanding promoter activity.

1.6 - Structural Features of Transcription Factors

A fundamental feature of transcription factors is their interaction with DNA, and delineating the structural features associated with this activity has the potential to uncover elements of similarity between regulators across the clades of life. Classification of transcriptional regulators into broad families has been carried out through examination of conserved genomic targets, the class of inducer molecules TFs interact with, and application of statistical methods that leverage sequence alignments of various TFs to

identify the presence of structural motifs in the DNA binding domain (Aggarwal, Rodgers, Drottar, Ptashne, & Harrison, 1988; Aravind & Koonin, 1999; Baumgart et al., 2021; Minezaki, Homma, & Nishikawa, 2005. Analyzing the structural features of TFs across bacteria, archaea, and eukaryotes suggest that a fundamental DNA binding sequence, the Helix-Turn-Helix (HTH) motif, is conserved across all domains of life through incorporation into the structural apparatus of transcription factors and members of the basal transcription complex (Rosinski & Atchley, 1999). However, the subsequent divergence in form and function of TFs across these clades has entailed the evolution of several variants of the HTH sequence across bacteria and eukaryotes reflecting the different constraints associated with DNA binding of these regulators (Aravind, Anantharaman, Balaji, Babu, & Iyer, 2005; Rigali, Derouaux, Giannotta, & Dusart, 2002). While the case for sequence divergence is logical when considering the specificity conveyed to TFs in binding to different sequences (and therefore regulating different genes), there seems to be little correlation between regulatory output and classification into an associated TF family based on DNA binding sequence homology (Madan Babu & Teichmann, 2003). Sequence divergence is also pervasive for different TFs across bacterial species, with the LysR family of transcription factors having members associated with a highly conserved N-terminal domain due to a shared HTH-motif, but strong divergence in sequences associated with the C-terminal domain, reflecting structural elements associated with binding to distinct inducers (Schell, 1993). This compartmentalization of functions associated with DNA binding and inducer/regulatory interactions is prevalent across bacterial TFs, with TFs grouped in the same superfamily

often demonstrating variability in the amount of distinct regulatory domains available to those TFs to achieve specific gene expression outcomes (Perez-Rueda et al., 2018) . While DNA binding domains for TFs have been extensively interrogated, the regulatory domains of TFs are difficult to probe in structural considerations due to the presence of intrinsically disordered sequences that generally characterize these domains (Brodsky et al., 2020; J. Liu et al., 2006). While recent investigations have laid to groundwork to systematically profile and classify the regulatory domains of TFs (DelRosso et al., 2023), controlled dissection of critical protein features that mediate activating and repressing interactions by TFs are still lacking for most regulators.

1.7 - The Paradigms of Transcriptional Activation and Repression (Prokaryotes)

Mechanistic studies on the specific regulatory effect of TFs have chiefly been confined to certain "model" TFs. The global regulator CRP (Cyclic AMP receptor protein) has been extensively profiled and serves as a thoroughly dissected TF in *E.coli*, bridging structural considerations and molecular genetics to elucidate a complete regulatory picture (Borukhov & Lee, 2005; Dalma-Weiszhausz & Brenowitz, 1996; C. L. Lawson et al., 2004) . CRP regulated promoters are differentially regulated depending on the distance between the CRP TF and the transcription start site (TSS) (Déthiollaz, Eichenberger, & Geiselmann, 1996; C. L. Lawson et al., 2004). The paradigm of Class I and II activation stems from these investigations, with the discovery that CRP binding at one class of promoter positions the TF to contact specific subunits of the RNAP holoenzyme through acidic regions of bound CRP. These contacts engender specific regulatory output, with the Class I activating contacts working on stabilizing RNAP binding to the

promoter.(Busby & Ebright, 1997, 1999; Schultz, Shields, & Steitz, 1991). For Class II activation, CRP binding occludes contact of RNAP with the -35 sequence but compensates by stabilizing contacts in the α CTD domain of RNAP at the UP element (Busby & Ebright, 1997). In addition to this stabilization, the closer binding to RNAP mediates more extensive contacts, promoting isomerization from closed to open complex and facilitating elongation. Analogous to CRP function, TFs in eukaryotes, including the yeast GAL4 TF work on stabilizing and priming RNAP for elongation at the promoter (Yankulov et al., 1994). Importantly, the acidic domains of the TF that mediate these interactions with RNAP are exceptionally potent, with the SRB2 component of the mediator subunit found to promote strong recruitment of RNAP while fused to the transcriptional repressor LexA (Farrell, Simkovich, Wu, Barberis, & Ptashne, 1996).

Seminal work on the nature of repression has proposed paradigms of gene regulation. Specifically, the idea of "kinetic competition", where the repressibility of the promoter is governed by basal promoter recruitment, has served as a framework for repression(Lanzer & Bujard, 1988). In this model, the susceptibility of a promoter is realized due to mutual inhibition between TF and RNAP: it follows that those promoters with strong basal RNAP binding occludes repressor interactions with the promoter, "crowding" out the repressor and resulting in a weakly regulated promoter(Giordano, Deuschle, Bujard, & McAllister, 1989). Well-studied repressors such as LacI and the lambda phage repressors conform to this model, with ramifications for design of synthetic circuits for tight control of target genes (Lutz & Bujard, 1997).

1.8 - Kinetic and Thermodynamic Models of Gene Regulation –Robust Modeling of TF Function

1.8.1 - Modeling TF-Operator Sequence Search Kinetics

An early question in the study of gene regulation was the search time involved for TFs to find their target DNA sequence. A crucial step in promoter modulation, knowing the fraction of time TFs remained bound to the DNA as opposed to diffusing in the cellular matrix would inform the timescale at which TFs interact with their target sequence (Berg, Winter, & von Hippel, 1981; Halford & Marko, 2004). Using initial estimates for TF and non-specific DNA interactions, the association kinetics of the *lac* repressor were found to be orders of magnitude faster than what was expected for a simple model involving 3-D diffusion(Berg et al., 1981). Theoretical investigations provided several justifications to reconcile the fast search times of this TF that included an initial movement using 3-D diffusion, followed by one-dimensional movement along DNA interspersed with periods of "hopping" between DNA sequences(Berg et al., 1981). Indeed, experimental validation of these movements were later confirmed for both the LacI and TetR transcription factors, with differences in the amount of time portioned between the cellular and DNA diffusion regimes hinting at a material distinction in search kinetics between these factors(Kleinschmidt, Tovar, Hillen, & Porschke, 1988; R. C. Lawson, Jr. & York, 1987). These investigations suggested a possible framework to conceptualize TF movement between specific and non-specific DNA, framing the question of gene regulation in terms of the (on and off) parameters of the TF with regards to binding to sequences associated

with its target promoter. Furthermore, it established the importance of interplay of theory and experimental interrogation in elucidating the regulatory properties of TFs.

1.8.2 - Telegraph Models of Gene Expression: Phenomenological Modelling of Gene Regulation

As kinetic models had been successfully applied to address the question of search times of the TF for its operator sequence, this framework was naturally extended to model the regulation of promoters by TFs. Initial observations of the "bursty" nature of gene expression led to the adoption of the Telegraph model to parametrize the process of transcription (Elowitz et al., 2002; Raj, Peskin, Tranchina, Vargas, & Tyagi, 2006). In this model, gene expression is characterized by the kinetic process where the promoter is modeled as being in a state of readiness or inactivity with the transition between these two states set by molecular factors that determine the probability the promoter resides in a primed state. In the primed state, the promoter generates mRNA in a series of bursty events given by the burst parameter which equates to multiple rounds of transcription initiation and elongation. In this framework the effect of the TF can be modeled as determining the rate of transition between the active and inactive promoter, with activation of gene expression generally skewing the distribution of promoter states to the ready or primed state and repressors to the quiescent (off) state. Telegraph models have been widely used to model both prokaryotic and eukaryotic transcriptional regulation (Bartman, Hsu, Hsiung, Raj, & Blobel, 2016; Golding et al., 2005; Molina et al., 2013; So et al., 2011) due to the phenomenon of bursty transcription dynamics observed at the promoter and has allowed for quantitative determination of the kinetic effects of TFs on

the promoter states. Appraising TF function in terms of its modulation of this transition between promoter on and off states has been used to compare different promoters and delineate similarities and differences in the regulatory landscapes of genomic targets of TFs (Haberle & Stark, 2018; Larsson et al., 2019).

1.8.3 - The Chemical Master Equation: Discrete State Kinetics Allows for Modelling TF Activity on RNAP Recruitment and Post-Recruitment Steps

One drawback of the Telegraph model is the inability to model the direct regulatory function of the TF on the enzymatic processes of RNAP. Detailed investigation of TF function *in vitro* has revealed several modes in which TFs can regulate the kinetic steps of RNAP that encompass the recruitment step and the post-recruitment activity. Furthermore, evidence from the phage psi 29 repressor has shown that TFs can often act with a high degree of specificity at the promoter, allowing RNAP binding to occur at the unregulated kinetic rate constant, but slowing the time taken for RNAP to initiate transcription (Monsalve, Calles, Mencía, Salas, & Rojo, 1997). The challenge of modelling the TFs ability to regulate multiple steps of the kinetic process of transcription is a limitation when using Telegraph models, where these steps are coarse-grained into the on and off states of the promoter.

An alternative approach is modelling TF-RNAP promoter occupancy explicitly via a kinetic model that accounts for the distribution of states, where the relevant parameters include the rates of the detailed inter-molecular events such as the interaction of TF and RNAP with DNA, the rate or mRNA synthesis, and the change in rates relevant to the

kinetic processes of RNAP brought about by TF function (Sanchez, Choubey, & Kondev, 2013). The kinetic model of a simple regulatory architecture, where a single TF regulates the activity of RNAP from one operator sequence is shown in Figure 1.1.

In panel A, the regulatory effect of the TF is modeled as acting on two steps of transcription – (de)stabilization of RNAP at the promoter and (de)acceleration of the initiation rate of bound RNAP. These parameters are independent and span the full regulatory space (encompassing both positive and negative regulation of the respective kinetic steps) to realize net outcomes of activation and repression of gene expression. In panel B, we list out the possible states of occupancy at the promoter for both TF and RNAP, as well as the kinetic rate constants that determine the transition rates between the discrete states of promoter occupancy. Note that the impact of the TF on RNAP is embedded in the rate constants: the effect of β is modeled as changing the off rate (k_p) of RNAP (either stabilizing or destabilizing RNAP at the promoter) while the effect of α works on altering the transcription initiation rate.

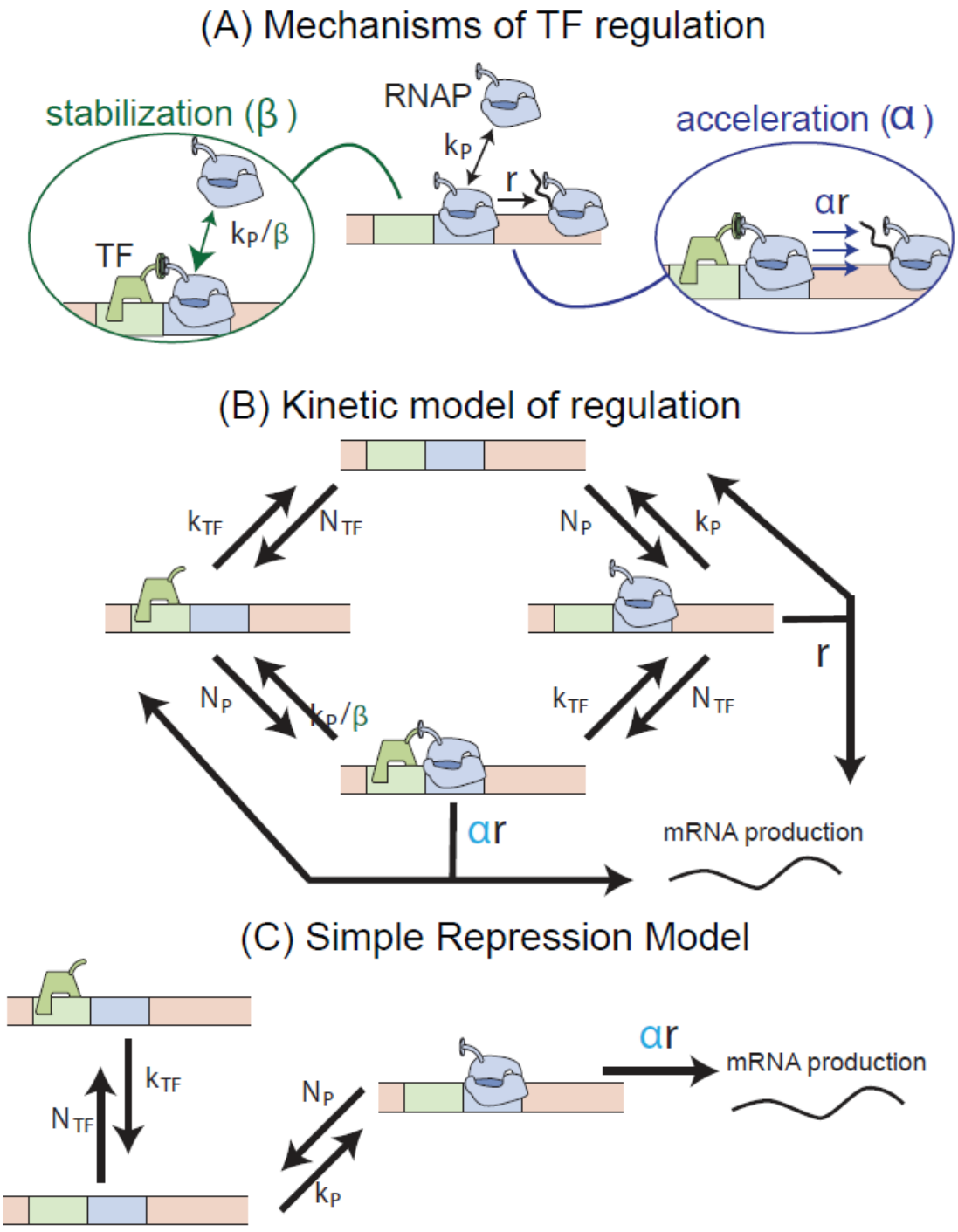

(A) Mechanisms of TF regulation
stabilization (β)
RNAP
acceleration (α)
kP
r
αr
kP/β
TF
(B) Kinetic model of regulation
kTF
NTF
NP
kP
r
NP
kP/β
kTF
NTF
αr
mRNA production
(C) Simple Repression Model
kTF
NTF
NP
kP
αr
mRNA production

Figure 1.1: Explicit Modelling of TF-RNAP Regulation Under a Simple Regulatory Architecture. (A) Regulatory interactions between TFs and RNAP as specified by a 2-step model of transcription initiation. We model the TFs ability to effectively promote RNAP binding as stabilizing its interaction with the promoter sequence – hence the parameter β functions by modifying the off-rate. **(B)** The full kinetic model for promoters under a simple regulatory architecture (single TF with one operator site). **(C)** A simplified model of the full kinetic model for a repressor and RNAP mutually inhibiting binding to regulatory DNA.

The goal of the kinetic model is to compute the expected number of mRNA molecules given the transition rates and the TF regulatory effect (encoded in α and β), and the Fold-Change in mRNA between the regulated and unregulated state takes the following form at steady state:

Equation 1.1:

$$FC = 1 + \frac{\alpha\beta(1+P) - 1 - P\beta}{P\beta}\langle Pr_{11}\rangle$$

Where the quantity $\langle Pr_{11}\rangle$ is as follows:

Equation 1.2:

$$\langle Pr_{11}\rangle = \frac{RP\beta(1+P+V_2+V_1(1+R)}{P(1+R\beta)(1+P+V_2)+RV_1(1+R)(1+\alpha\beta V_2+P\beta)+(1+V_1+\alpha\beta V_2)(1+R)(1+P+V_2)}$$

With the following definitions for the parameters: $R = \frac{N_{TF}}{k_{TF}}, P = \frac{N_p}{k_p}, V_1 = \frac{k_{TF}}{k_P}, V_2 = \frac{r}{kp}$

The rate constants depicted in Figure 1.1 are encapsulated by the macroscopic parameters R,P,, V_1, V_2. The parameters R and P represent the effective concentrations of the TF and RNAP respectively while the V_1 parameter represents the ratio between the off-rates for the TF and RNAP. The final parameter, V_2, represents the ratio between the transcription initation rate and the off-rate of RNAP unbinding from the DNA. Note the large number of terms in the expression $\langle Pr_{11} \rangle$ that obscures an intuitive understanding of the fundamental features of the regulatory response (the Fold-Change in the expected number of mRNA molecules produced) as a function of TF regulation. While the full kinetic model allows for simplification of the response function in special cases of TF activity (see Panel C), a general treatment of gene regulation requires measuring or fitting several parameters, introducing many degrees of freedom in the parameter regimes. This creates redundancy in parameter composition, complicating the interpretation of the regulatory effect of the TF on gene expression.

1.8.4 – Thermodynamic Models of Gene Regulation: Equilibrium Assumptions for an Intuitive Framework

A robust framework for understating TF regulation requires an interplay of theory and measurement to capture the fundamental elements of gene regulation and build frameworks of sufficient complexity to discriminate between competing hypotheses. The large number of parameters inherent in kinetic models often preclude an intuitive correspondence between the fundamental regulatory impact of the TF and the net regulatory output. Thermodynamic models of gene regulation are idealized cases of the kinetic model and have been utilized to explain the transcriptional programs of the

Drosophila embryo to the genetic logic of the lambda phage promoters that govern the switch from lysogeny to the lytic life cycle (Ackers et al., 1982; Bintu et al., 2005; Fakhouri et al., 2010; He, Samee, Blatti, & Sinha, 2010). The appeal of these models is intuitive: the properties of gene regulation can be distilled into biochemically meaningful parameters such as dissociation rate constants, cooperativity between different regulatory molecules and chiefly, the effect of the TF on RNAP (Bintu et al., 2005; Brewster et al., 2014). These models along with experiments designed to test specific aspects of promoter architecture, can investigate the contributions of these features to properties of the TF's regulatory response.

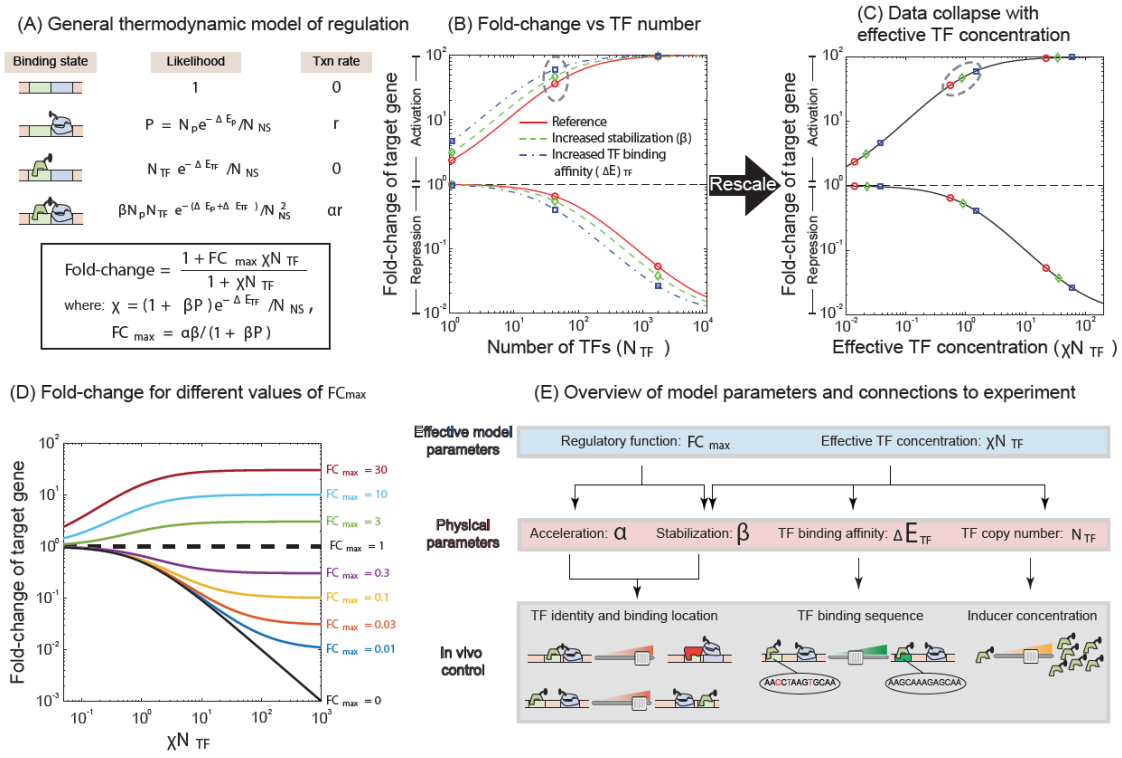


Figure 1.2: Thermodynamic Model of Gene Regulation Under a Simple Regulatory Architecture (A) The states of the thermodynamic model with the respective weight. The enumeration of each state is normalized to the unbound state, giving a likelihood of 1 for that weight. The parameters ΔE_p and ΔE_{TF} determine the binding affinity of RNAP and TF respectively. The role of the TF on gene expression is captured in two parameters β - the stabilization coefficient and α - the coefficient that modifies the rate of initiation.**(B,C)** Plot showing the role of stabilization on gene expression. When scaled by TF occupancy the curves collapse on the effective TF concentration.**(D)** The degree of stabilization, acceleration, and promoter strength set the maximal regulatory response of the TF. **(E)** The thermodynamic model probes the role of specific in vivo features to gene regulation giving the ability to isolate what properties of the regulatory function are dictated by these features.

In our work, we use a thermodynamic model of regulation as shown in Figure 1.2A: the possibilities of the system are enumerated with respect to the promoter of interest and the likelihood corresponds to a function of the biophysical parameters that make the state probable. The regulatory properties of the TF can be captured in the two effective model parameters (Panels A,E) of χ and FC_{max} that set the TF occupancy and maximal regulatory response.

In Panels B and C, we demonstrate the role of TF occupancy on the regulatory response. Scaling the measured TF concentration with occupancy (the product of the TF binding affinity and its degree of stabilization) collapses the regulatory response: TFs with high effective occupancy move further along the curve toward the maximal regulatory output, while TFs with low effective occupancy require higher TF copy number to realize their maximal response. For the FC_{max} parameter, both stabilization and acceleration play a role (A,D) with the TFs change in the initiation rate functioning as a scaling factor for the Fold-Change in expression. In Figure E, we demonstrate the assumptions that relate how these effective parameters are related to the *in vivo* regulatory features. Importantly the role of these features, such as binding sequence, and position of regulation can be tested rigorously to see what aspects of the thermodynamic model they influence.

The potential for regulating two steps of the transcription process independently allows for modelling the regulatory regimes that encompass the full spectrum of regulation as mentioned in Section 1.7. TFs that operate in a regime that stabilize RNAP binding yet slow the rate of initiation would function in a manner distinct to that of kinetic competition as mentioned in the previous section: the stronger promoters would be more

repressible, with a decline in susceptibility of regulation as the basal RNAP binding weakens. This raises a broader question as to the kinds of stabilization and acceleration regimes that are present among TFs. However, the absence of detailed, mechanistic driven regulatory information for most TFs, even in well-characterized model systems such as *E.coli*, underscores the potential scope of regulatory ignorance and precludes the field from developing a first-principles approach to interpreting the complexity of endogenous promoters(Garcia et al., 2016; van Hijum et al., 2009).

1.9 - Synthetic Control of Promoter Activity

Over the past two decades, synthetic transcriptional factors have been increasingly used to accomplish specific gene expression. This has thrust the role of design principles of gene expression to the forefront of synthetic biology with attempts to design gene regulatory circuits based on functionally defined and modular building blocks (English et al., 2021). Rather than focus on a mechanistic dissection of a suite of model TFs limited to prokaryotic gene regulation, rational design of TFs with specific properties have dominated eukaryotic gene regulation studies, with the aim of designing synthetic gene circuits for specific purpose (Chen et al., 2022; Pandelakis, Delgado, & Ebrahimkhani, 2020). These synthetic TFs are often based on modular DNA binding domains such as the Zinc Finger binding domain (Beerli & Barbas, 2002). By determining the protein-DNA interface for these domains, the discovery that these protein sequences bind with specificity to triplicate nucleotides spurred design of zinc finger “arrays’ capable of targeting regions of the genome precisely(Gogos, Hsu, Bolton, & Kafatos, 1992; Moore, Klug, & Choo, 2001). This has served as a foundation to design gene circuits with

regulatory features, where the specificity to DNA binding conferred through zinc finger arrays is merged with regulatory domains of well characterized protein sequences that activate gene expression (Ptashne, 1988).. Indeed, the tunability of specific elements of the transcription response, such as the degree of strong/weak binding, cooperativity with neighboring molecules, and the interchangeability of the regulatory domains of these features has been a pull towards using these synthetic TFs to accomplish transcriptional goals with modular construction (Bragdon et al., 2023; Swank, Laohakunakorn, & Maerkl, 2019).

Applications of synthetic transcription factors to drive gene circuits that accomplish specific cellular functions have benefitted from the great degree of modularity associated with the individual components. The use of the zinc-finger DNA binding domains fused to well-characterized activation domains allow for specific targeting of transcriptional activators to defined gene sequences. This programmable toolkit for genetic control has been used in the context of remediating diseased states, engaging in cellular surveillance, and reprogramming cells by inducing the expression of key factors (Heiderscheit, Eguchi, Spurgat, & Ansari, 2018; H. S. Li et al., 2022; Newby et al., 2017; Piatek & Mahfouz, 2017). Indeed, the application of models of gene regulation to these effectors have demonstrated the effectiveness of the thermodynamic model of gene regulation in framing the regulatory behavior of synthetic transcription factors and suggests that these models have predictive capabilities in both prokaryotes and eukaryotic gene expression programs (Bashor et al., 2019).

However, the goal of understanding the principles of natural gene regulation can oftentimes be difficult to glean from the synthetic approach. In particular, the synthetic approach leverages well characterized DNA binding and regulatory domains to facilitate modular circuit design (Khalil et al., 2012), which sometimes runs counter to endogenous TF regulation featuring a multitude of different regulators with often competing functions at the promoter. As these approaches are tailored toward eukaryotic gene regulation where activation is the norm (Ptashne, 1988), the focus on positive regulation obscures the vast regulatory landscape of repression of gene expression. Furthermore, the focus of synthetic transcription factors on well-characterized parts, while essential for maintaining robustness and predictability of the designed regulatory circuits, often cannot reveal the full scale of possible determinants associated with endogenous gene regulation, where the diversity of promoter architectures outpaces the limited repertoire of functional synthetic parts.

1.10 - Deconvolving the Role of Binding Sequence and Position in TF Regulation

Understanding the role of different aspects of promoter architecture has been essential in deciphering core elements that set the overall regulatory profile of a gene. Early investigations probed key features such as binding sequence composition, orientation of and location of the binding site on the activity of TFs, with several studies highlighting the sensitivity of TF regulation to these features. These studies elucidated the sequence determinants that affected the kinetics of TF binding to the promoter (Lehming et al., 1987), establishing that prokaryotic gene regulation involves information dense regions that set the occupancy of TFs at the promoter (Barker, Fickert, Oehler, & Müller-hill,

1998; Niland, Hühne, & Müller-Hill, 1996). In tandem with sequence, the position axis has been extensively interrogated (Besse, von Wilcken-Bergmann, & Müller-Hill, 1986) revealing strong spatial constraints that shape distinct properties of the TF regulatory function. The influence of TF binding position on net regulatory effect is nuanced: while prokaryotic regulators often bind to regions close to the promoter to achieve specific protein-protein contacts with RNAP for both activation and repression, the presence of DNA looping allows for stable repressor complexes to develop from TFs bound at a distance (Becker, Peters, Maher, & Lionberger, 2013; Boedicker, Garcia, Johnson, & Phillips, 2013). This effect allows for synergistic repression that cannot be realized by individual binding of TFs in proximity due to steric hindrance between these factor, with implications for precision and magnitude of gene regulation (Vilar & Leibler, 2003).

However, there has been little systematic attempt to link aspects of promoter architecture to the TFs kinetic role - its ability to influence one or multiple aspects of the transcriptional process - at promoters. This leaves the role of fundamental characteristics such as TF binding sequence, binding location relative to the TSS, and the degree of basal promoter strength on the kinetics of RNAP regulation unclear. The role of TF binding sequence in prokaryotes has canonically been assigned to mediate TF recruitment to DNA: in the language of thermodynamics, the sequence sets the binding affinity of the TF. However, techniques that illuminate the critical step of TF-DNA interactions such as SELEX have qualified aspects of TF recruitment to regulatory DNA: highlighting the existence of sequence determinants that mediate cooperative or antagonistic interactions

between TFs (Inukai, Kock, & Bulyk, 2017; Jolma et al., 2015; Morgunova & Taipale, 2017).

Indeed, the potential of DNA binding sequence to facilitate not only the affinity of the TF for the promoter but also its contacts with neighboring molecules, suggests a possibility that binding sequence could shape TF-RNAP contacts – potentially influencing the activity of the TF on the kinetic steps of transcription. Furthermore, direct evidence on the contribution of TF operator sequence to gene expression in LacI has confirmed its role is inconsistent with that of solely setting TF occupancy (Garcia et al., 2012) – indicating that not only where, but what TFs bind to has relevant properties for kinetic regulation of RNAP. However, a statistical approach that formally investigates the contributions TF binding sequence and position to the regulatory modes of the TF (the ability to stabilize and accelerate RNAP) has not to the best of our knowledge been systematically carried out across *E.coli* TFs.

The influence of basal promoter strength on TF regulation is equally important and is one component of promoter architecture that influences the regulatory output of TFs. The model of kinetic competition, where the repressor and RNAP mutually compete for binding at the promoter, is one of several ways repression can occur. Considering the two -step model proposed in Figure 1.1, repression can arise through the interplay between the modes of regulation (stabilization and acceleration). The avenue of destabilization ($\beta < 1$) is the intuitive way of conceiving repression, however, TFs that regulate in the regime of stabilization RNAP ($\beta > 1$), but greatly impair the kinetics of promoter escape ($\alpha < 1$) will result in net repression if certain conditions are met . This

result has been shown experimentally for several TFs, with *in vitro* work clarifying the mechanism of a "kinetic trap" (Roy, Semsey, Liu, Gussin, & Adhya, 2004; Schröder & Wagner, 2000) by TF stabilization of RNAP-DNA kinetic intermediates leading to delayed progression in promoter escape. Indeed, some TFs regulate RNAP progression with remarkable specificity, influencing only the stability of the kinetic intermediates with no effect on RNAP recruitment (Swapna, Kumari, & Nagaraja, 2015). The features that determine promoter strength, as outlined in Section 1.4, play a role here, with basal promoter strength setting the degree to which stabilization of these kinetic intermediates translates into an effective kinetic trap (Djordjevic & Bundschuh, 2008; Jensen & Galburt, 2021). The *Mycobacterium* TFs CarD and RbpA have been shown to alter the kinetics of the RNAP-DNA intermediates, slowing the complex and preventing fast transition to elongation (Jensen, Manzano, Rammohan, Stallings, & Galburt, 2019),

The challenge of deconvolving the relative contributions of sequence, location, and promoter strength in natural promoters, as opposed to *in vitro* reconstitutions of minimal promoters, is apparent. Measurements of promoter activity from natural genes are often complicated by the presence of multiple regulating factors and the coupling of promoters in gene regulatory networks (feed-back loop, etc.). Furthermore, the ability to precisely manipulate the *in vivo* concentration of TFs is essential to deconstruct how specific elements of the regulatory architecture shape promoter activity (Naqvi et al., 2023). Yet the number of TFs for which specific information on inducers is available is often a subset of the total regulatory repertoire available to a cell, limiting the characterization of TFs in an organism.

1.11 - Focus of this Thesis –Probing the Interplay Between TF Binding Sequence, Location, and Promoter Type on Gene Expression

In this thesis, a synthetic biology approach is used to probe the regulatory function of a subset of *E.coli* TFs. Specifically, the challenge of controlling and measuring TFs *in vivo* is overcome by utilizing a library of TFs fused to the mCherry fluorescent protein. This library of *E.coli* strains has the endogenous TFs deleted from their native locus, with the TF coding sequence integrated into a defined locus under the control of the pTet promoter (Gossen & Bujard, 1992). The TFs are expressed by adding aTC (anhydrous tetracycline) to the growth medium, and different aTC concentrations can be used to tune the copy number of the TF under investigation. This approach effectively decouples expression of the TF from its endogenous context, allowing us to explore the function of these regulators without subjecting the cell to changes in physiology necessary to induce expression from the TFs native promoter.

Utilizing this library and designing synthetic gene circuits to test the influence of TF binding sequence, location, and the type of regulated promoter on expression, the quantitative relationship between promoter determinants and its regulatory output is measured for these TFs. The measured response function is then interpreted using a thermodynamic mode of gene regulation that enables us to decouple TF occupancy from its maximal regulatory effect. As described in Section 1.8, we model a TFs regulatory activity as acting on two steps of the transcription kinetic cycle: the recruitment of RNAP (RNA polymerase) to the promoter sequence (stabilization) and the rate of transcription initiation by bound RNAP (acceleration).

To uncover spatial rules regarding the kinetic modulation, we investigated the role of the position axis in setting the regulatory modes of the CpxR activator. Building on this approach, 93 *E. coli* TFs were interrogated using the synthetic constructs to deconvolve the effect of TF binding sequence and position on the TFs regulatory function. Finally, using information on the regulatory modes uncovered for both activating and repressing TFs, we engineered promoters with different strengths to test predictions on the effect of stabilizing TFs across diverse promoter types to discover that stabilization of RNAP is a conserved mode of regulation employed by TFs that activate and repress. Overall, this work characterizes the regulatory modes of several *E.coli* TFs, uncovering the role of specific elements of promoter architecture in mediating the function of these regulators.

Chapter II – Results

Delineating the Role of TF Binding Position on the Regulatory Modes of the CpxR TF

Substantial Portions of this section, including several figures and one table have been adapted from the following research publication: Guharajan, S., Chhabra, S., Parisutham, V., & Brewster, R. C. (2021). Quantifying the regulatory role of individual transcription factors in Escherichia coli. *Cell Rep, 37*(6), 109952. doi:10.1016/j.celrep.2021.109952

2.1 - Introduction

The genomics era has enabled multiple techniques capable of determining where a TF will bind and with what specificity (Stormo, 2000; Messeguer et al., 2002; Wasserman and Sandelin, 2004; Weirauch et al., 2014). While this information is crucial for building occupancy-based models of gene regulation, there is still an important missing component to gene regulation; the quantitative regulatory role of a TF, when bound, is often unclear. Determination of TF function are often accomplished through overexpression of the endogenous TF or by deleting the regulator of interest and observing the expression changes from a target promoter. This serves a purpose in predicting the specific role of that TF on a given gene but offers less predictive power when examining regulation of other genes by that same TF or to different binding sites. These measurements of gene regulation are often entangled in indirect regulatory effects

such as TF-TF interactions (Vidal et al., 2011; Rolland et al., 2014), feedback (Bateman, 1998; Shen-Orr et al., 2002), and physiological (i.e., growth rate) (Klumpp et al., 2009; Schmidt et al., 2016) and off-target competitive effects of decoy binding sites or other genes in the network (Brewster et al., 2014; Lee and Maheshri, 2012). Because of this, a single TF that binds at the same relative location on two different natural promoters can appear to have opposite regulatory roles. The entanglement between indirect and direct regulation likely contributes to this ambiguity and prevents the field from developing a basic intuition of TF regulatory function.

To systematically interrogate the position axis of TF binding and uncover the effect of binding location on the TFs ability to regulate gene expression through either regulatory mode (stabilization of RNAP and acceleration of transcription initiation), we designed synthetic circuits that allow for facile cloning of TF binding sequences at several locations relative to the Transcription Start Site (TSS). The circuits were designed using a low copy number plasmid (pSC101) that harbors about 8-12 copies per cell (Brewster et al., 2014) and use a well-characterized mutant of the *LacUV5* promoter regulating the YFP gene with defined ribosomal binding sequence and termination elements (Figure 2.1A). This promoter was chosen because it had sufficient dynamic range to confidently measure strong repression (100-fold decreases from constitutive bulk expression levels are measurable above noise) and activation. Overlapping primers were designed and synthesized to amplify various portions of the synthetic circuit to clone the *ccdB* cassette at specific locations. This cassette is part of the toxic and anti-toxin proteins found in *E.coli*, and the presence of this cassette in strains lacking the anti-toxin or specific

mutations in DNA gyrase kills the cell (Afif, Allali, Couturier, & Van Melderen, 2001). These synthetic position circuits were cloned into the DB3.1 strain of *E.coli* that harbors the mutation, allowing for propagation of the constructs. The *ccdB* cassette insertion presented overhangs for the Type IIs restriction enzyme *BbsI* and digestion of TF binding sequences with the flanking overhangs allowed for scarless ligation of binding site oligos. The constructs were then transformed into standard *E.coli* strains to generate the synthetic circuits with TF binding sequences at specified locations.

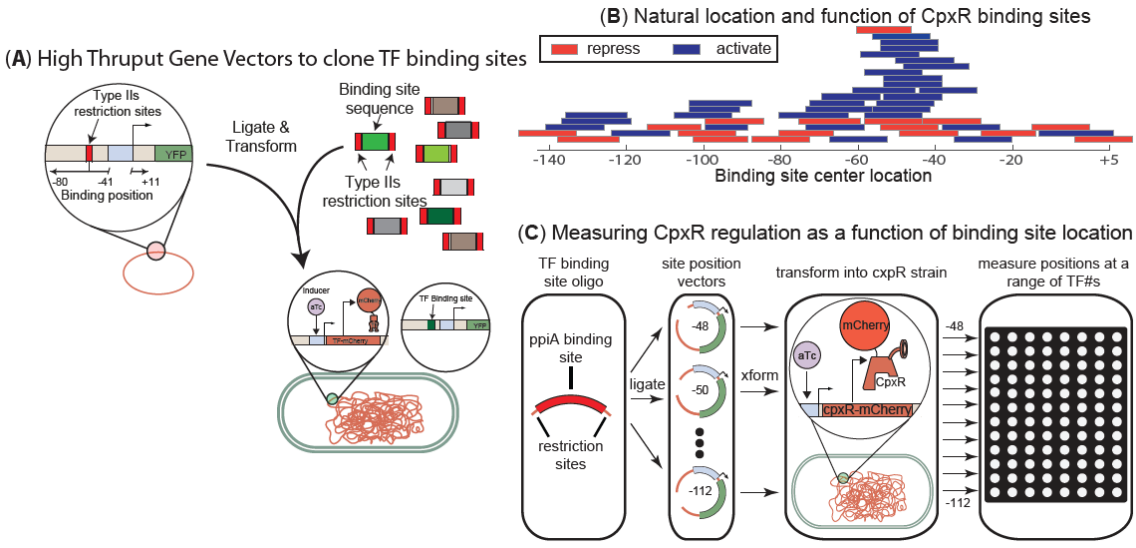


Figure 2.1: Synthetic Biology Approach for Facile Cloning and Measurement of CpxR Regulatory Activity at Defined Binding Locations (A) High Throughput cloning vectors were designed to interrogate the regulatory effect of any given TF at multiple positions upstream and downstream of the promoter. The bases circuit was a modified LacUV5 promoter sequence used previously and position primers were designed to clone in Type IIs restriction sites at defined locations relative to the Sigma 70 promoter. Cloning of the TF binding sequences is accomplished by digesting the plasmids and ligating oligos containing the sequence with complementary overhangs. **(B)** 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. **(C)** Schematic strategy for constructing and measuring CpxR acting at the ppiA promoter sequence inserted at 21 upstream positions and 1 downstream position in our synthetic gene circuit.

We profiled the RegulonDB database for information on candidate TFs to select for interrogation of the positional landscape of regulatory function. The transcriptional regulator CpxR, came to our attention: CpxR is a member of the two-component system of TFs, and is involved in cellular response to membrane stress, regulating dozens of genes involved in maintaining cellular integrity (Raivio & Silhavy, 1997). As demonstrated in Figure 2.1B, CpxR naturally binds to a wide range of promoter locations to regulate dozens of different genes in E. coli. However, the regulatory role of CpxR as a function of binding location is unclear from these data; repressive and activating interactions are attributed to many of the locations upstream of the promoter. We sought to clarify the isolated regulatory function of this TF using our synthetic circuit, with the goal of decoupling the modes of regulation as a function of binding position.

2.2 - Results

2.2.1 Profiling the Regulatory Output of CpxR Across 22 Binding Locations

To measure the isolated regulatory behavior of CpxR we cloned the synthetic gene circuits with the CpxR binding sequence regulating the *ppiA* promoter at multiple locations centered at -48 bp from the TSS to -112 bp from the TSS (Figure 2.1C). We chose this range because the vast majority of natural CpxR binding sites occur within these limits (Figure 2.1B). We then transformed these position circuits into a CpxR inducible strain with the ability to precisely manipulate the concentrations of the TF *in vivo* (Parisutham, Chhabra, Ali, & Brewster, 2022). Crucially, the TF had the *mCherry* cassette fused to the C-terminus via a linker sequence, allowing us to measure TF abundance. This gave us the ability to measure the quantitative input/output function for CpxR through measurements of TF concentration and promoter output via flow cytometry.

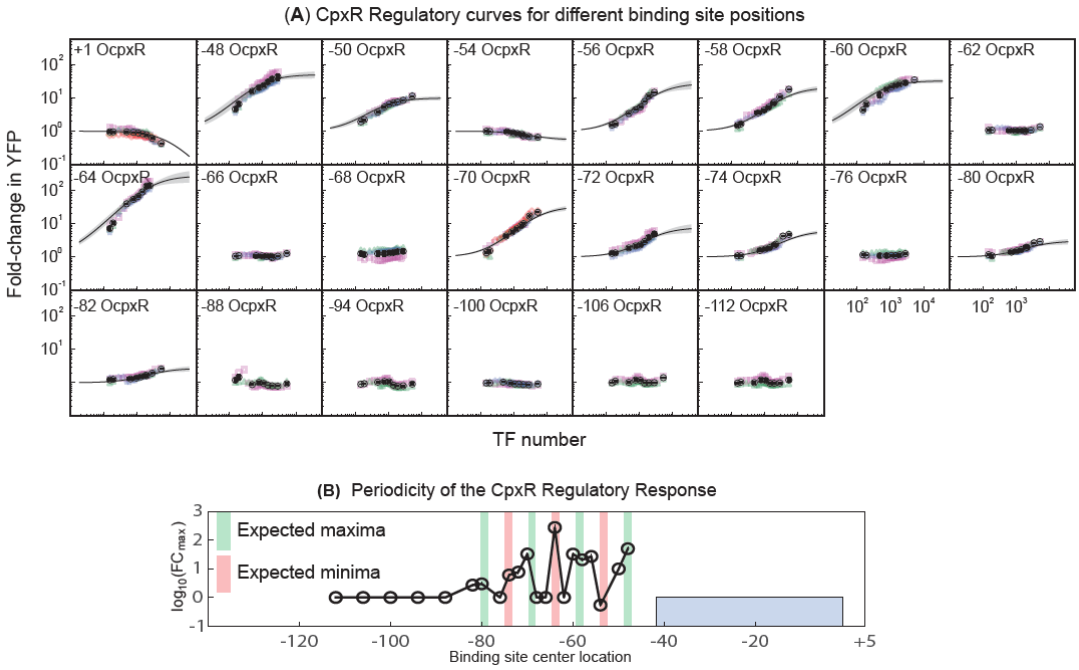


Figure 2.2: The Activation Landscape of CpxR is Strongly Sensitive to Position (A) The regulatory profile of CpxR as a function of TF copy number for the 22 binding locations. Each panel shows the TF copy number on the X-axis and the fold-change in YFP on the Y-axis of individual replicates (colored points). For all positions that show regulation (plots with the model curves) data point is the average of 3 replicates. The black points represent the mean and standard error of these replicates. For convenience, this is not shown for every TF number in the plot. The dashed line running though the data points is the theory prediction based on inference of the model parameters detailed in Methods 2.3.2.2. The shaded regions represent 2 standard deviations of the thermodynamic model conditioned on the inferred model parameters. **(B)** The mean of the inferred maximal fold-change (FC_{max}) for each of the 21 upstream binding locations as a function of the binding location at the promoter. The 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.

The data for fold-change as a function of CpxR copy number for the binding location sweep is shown in Figure 2.2A. 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 2.2B, we show fit values of the thermodynamic parameter FC_{max} (See Section 1.8.4 for details on the Thermodynamic Model of Gene Regulation) 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 (Besse et al., 1986; Garcia et al., 2012); we see this very roughly with maxima in activation around -48, -60, -70 and -80 bp, however -64 is a strong outlier which is expected to be close to a minimum but is strongly activating as measured here.

2.2.2 - CpxR Binding Location Determines the Degree of Stabilization in the Regulatory Response

To decouple the regulatory modes of CpxR, we developed a manifold approach that utilizes the Fold-Change in regulation at distinct binding positions. Previous studies of the CRP regulator used a manifold approach to decouple the regulatory mode of the TF at its canonical binding positions. To leverage the position sweep data for the CpxR TF, we coupled the response functions at two positions via TF concentration: this allowed for

removal of the TF binding affinity as a parameter in our models. In this case, the decoupling of regulatory modes in the manifold approach arises from the dependence of the curvature in our data on the degree of stabilization of CpxR. This approach is demonstrated schematically in Figure 2.3A. The concentration manifold plots the fold-change at one position against the fold-change of another position, each data point in this space corresponds to a measurement of both TF binding positions at the same effective TF concentration. In this case, we chose to plot all position data against the corresponding fold-change at +1; we chose +1 because we assume regulation is ``pure steric hindrance'' (FC_{max}=0, β=0) here (See Section 1.8.4 on the Thermodynamic Model of Gene Regulation). 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 2.3.2.1).

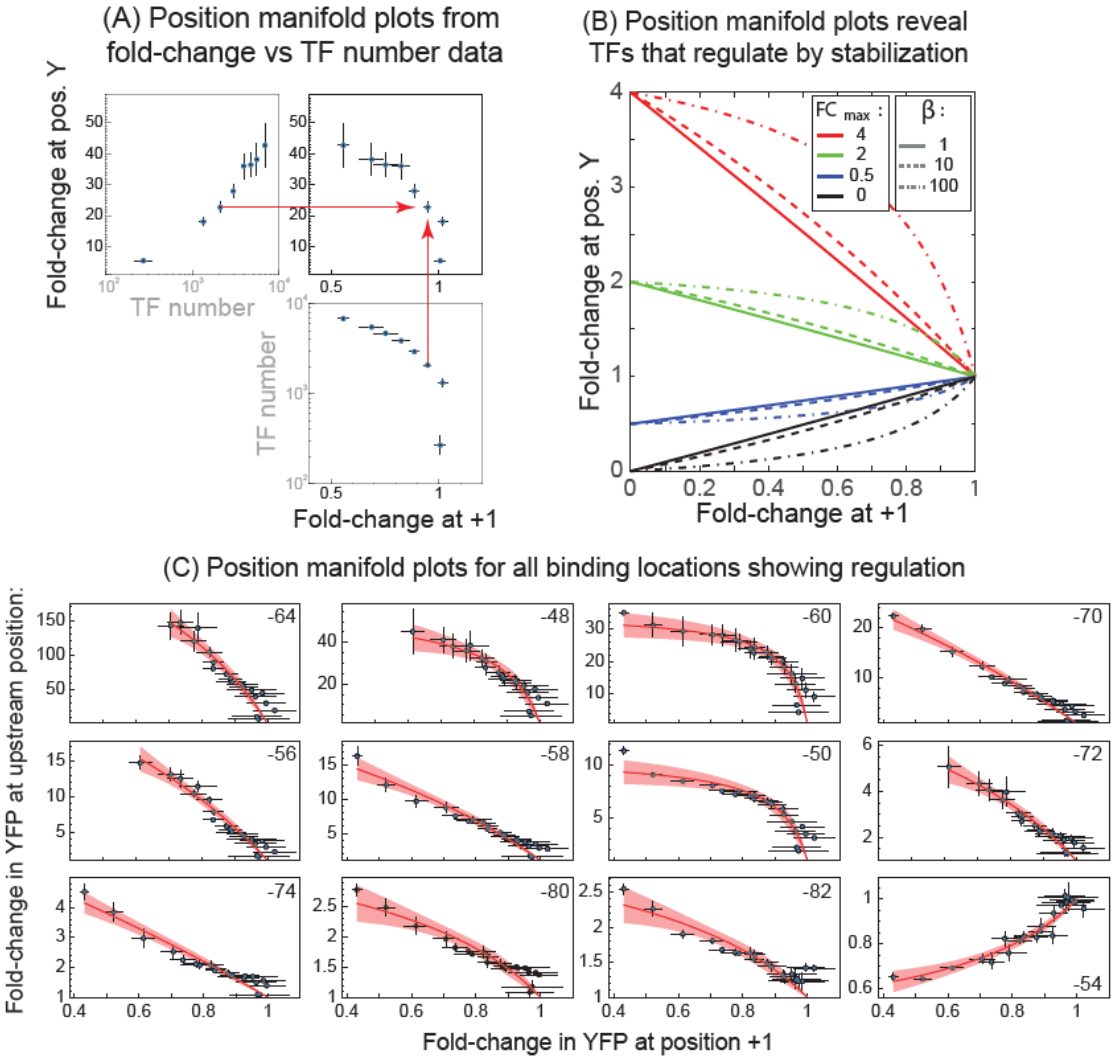

(A) Position manifold plots from fold-change vs TF number data
Fold-change at pos. Y
TF number
Fold-change at +1
(B) Position manifold plots reveal TFs that regulate by stabilization
FC max :
4
2
0.5
0
β :
1
10
100
Fold-change at pos. Y
Fold-change at +1
(C) Position manifold plots for all binding locations showing regulation
Fold-change in YFP at upstream position:
-64
-48
-60
-70
-56
-58
-50
-72
-74
-80
-82
-54
Fold-change in YFP at position +1

Figure 2.3: Concentration Manifold Approach Uncovers Stabilization of RNAP by CpxR at Multiple Binding Locations (A) Tracing out the position regulatory manifold using TF abundance from two different binding locations. **(B)** The concentration manifold curves are predicted to be straight lines when beta is small, but curved for larger values of beta. **(C)** Concentration Manifold plots for the 12 upstream regulatory positions as a function of the Fold-Change at +1. The solid red line represents the mean of the model regulatory profile (thermodynamic model) generated from the inferred acceleration and stabilization parameters and the shaded regions represent 2 standard deviations from the mean. Data points represent the mean and standard error of the Fold-Change as seen in the plot of Figure 2.2A for each respective position.

As seen in Figure 2.3B, data in the concentration manifold are expected to be lines emanating from (1,1) which is the defined fold-change of both locations when the number of TFs is zero. As TF concentration is increased, the fold-change at +1 decreases 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) we predict profiles are straight lines connecting the point (1,1) with the point corresponding to (*FCmax* of position 1, *FCmax* of position 2). However, larger β will cause the curve to rise or fall more rapidly than linear reaching FC_{max} at higher values of the corresponding +1 fold-change.

In Figure 2.3C we replot the data from Figure 2.2A as fold-change at an upstream regulatory position against fold-change at +1; each data point represents a measurement

of fold-change at these two different binding locations for a given TF copy number within the cell. The plots are arranged from the highest FC_{max} (strong activation at -64) to the lowest FC_{max} (weak repression at -54). The solid line represents our model curve using the inferred values of α and β. Based on the inferred stabilization values, we find that the activation profiles across the position sweep is driven by varying degrees of stabilization and acceleration. We find strong stabilization in regulation at positions -50, -60, -54, -48, and -64; visual inspection of Figure 2.3C shows the curvature we expected to see from strong stabilization. Several positions -58, -56, -70, and -74 have regulation profiles that are approximately straight lines implying that CpxR either destabilizes or does not regulate through stabilization.

2.2.3 - CpxR Activation Encompasses Both Coherent and Incoherent Modes of RNAP Regulation

The inferred values for α and β of all measured positions are shown in Table 2.1. Figure 2.4A shows a heat map for FC_{max} as a function of the regulation parameters α and β with each quadrant having a specific qualitative regulatory scheme. The top-right and bottom-left quadrants represent coherent regulation strategies where α and β both contribute to activate (top-right) or repress (bottom-left) gene expression. On the other hand the top-left and bottom-right quadrant are incoherent in the sense that the modes have opposing regulatory effects: TFs in these quadrant either slow the initiation rate of transcription while also stabilizing polymerase at the promoter (top-left) or increase the initiation rate

of transcription while destabilizing the polymerases' presence at the promoter (bottom-right).

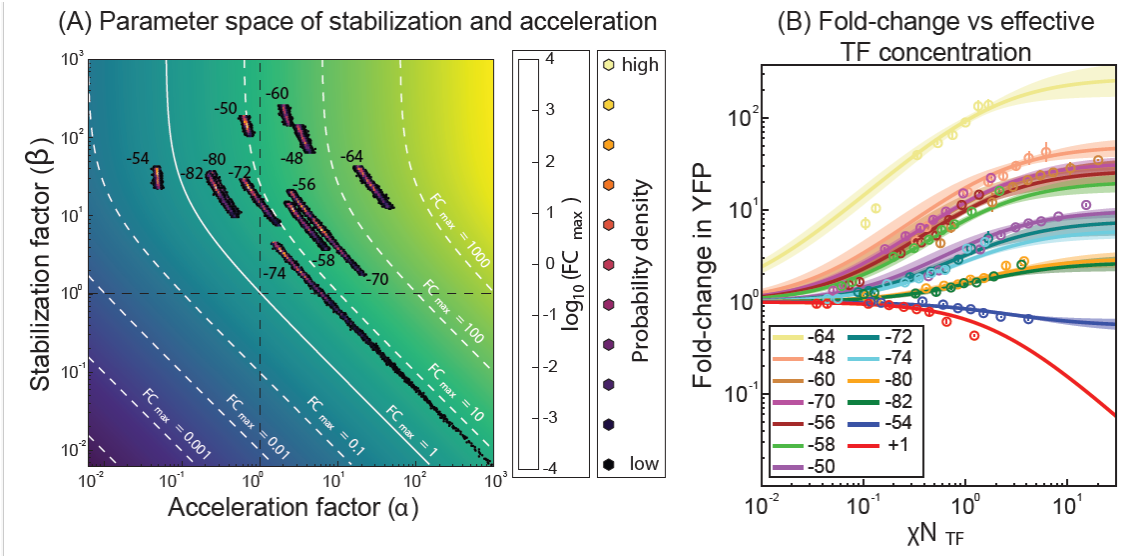


Figure 2.4: CpxR Binding Location Determines the Mode of TF-RNAP Regulation. (A) Phase plot of FC_{max} as a function of alpha and beta parameters. The horizontal black line marks beta=1 (no stabilization or destabilization) and the vertical black dashed line marks alpha=1 (no acceleration or deceleration). The white lines represent contours of constant FC_{max}. The colored points represent the parameter inference of alpha and beta for each of the regulatory positions. **(B)** Plot of fold-change against the effective TF concentration $\chi * N_{TF}$ for the 13 regulatory positions (12 upstream and 1 downstream) using parameters derived from the concentration manifold plots in Figure 2.3.

On this same plot we also show the inferred probability distribution of the parameters for each position in our data. The black points are lower probability with lighter points representing higher probability values for α and β. One notable phenomenon is for positions with β less than roughly 10, the inference begins to fail. This results in

inference clouds with ``tails' that stretch across quadrants and precludes an assessment into the mode of regulation (see position -74). The alignment between the inference clouds of these positions and the constant FC_{max} contours, however, shows that while we make very precise estimations of FC_{max}, the values of α and β are less certain and correlated. This is an unfortunate consequence of the weak stabilization limit in our model resulting in parameter combinations that encompass both coherent and incoherent regulatory regimes which could explain the data well.

Assessing the degree of stabilization and acceleration across the 12 positions with regulation, we see that CpxR regulates predominantly through the stabilization mechanism: this feature is robust across positions with all but 3 of the locations having definitive stabilization values greater than 2-fold (Table 2.1). The strongest stabilization position did not correspond to the peak activation position as seen in Figure 2.3B: in fact, the position with the strongest curvature in the manifold response curve were aligned closer to the sigma 70 promoter, **which falls in line with the Class I activation scenario** (See Section 1.7 for details). However, many positions with either low activation or repression are also found to be stabilizing. For these cases the TF must decelerate RNAP to realize the Fold-Change values leading to the two modes working antagonistically. This feature is apparent for the regulatory positions discussed previously (-54, -50, -80 and -82) (Figure 2.4A). The remaining 8 positions have stabilization and acceleration working in tandem, and we see this predominantly in the peak activation position (-64) where the inferred acceleration is 10-fold.

Combining the inferred regulatory parameters, 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 are used to plot the fold-change as a function of χ for the 13 regulatory positions (Figure 2.4B). Crucially, we demonstrate that the two key variables χ and FC_{max} are effective in capturing the fold-change position sweep profiles.

Binding position	α (alpha)	β (beta)
−48	$3.506^{4.008}_{2.898}$	$116.257^{150.131}_{66.453}$
−50	$0.675^{0.736}_{0.612}$	$161.766^{199.140}_{108.281}$
−54	$0.049^{0.051}_{0.047}$	$34.496^{43.447}_{23.154}$
−56	$3.38^{4.496}_{1.745}$	$14.723^{21.647}_{5.304}$
−58	$3.107^{4.127}_{1.635}$	$10.519^{15.386}_{3.796}$
−60	$2.161^{2.33}_{1.974}$	$219.757^{270.481}_{151.499}$
−64	$24.911^{31.849}_{14.911}$	$29.545^{44.169}_{12.984}$
−70	$6.995^{9.748}_{3.101}$	$6.224^{9.316}_{1.776}$
−72	$0.813^{1.053}_{0.451}$	$20.905^{30.744}_{7.678}$
−74	$2.329^{3.944}_{0.467}$	$2.982^{4.685}_{4.789\times10^{-6}}$
−80	$0.289^{0.355}_{0.197}$	$27.025^{37.948}_{10.944}$
−82	$0.271^{0.334}_{0.182}$	$24.811^{35.501}_{9.835}$

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

Table 2.1: Inferred Acceleration and Stabilization Values for Regulated Positions in the CpxR Data Set.

2.3 - Methods

2.3.1 Culture and Measurement Conditions

E.coli strain MG1655 was the base strain used for all synthetic regulatory circuit strains constructed and measured for the CpxR analysis. 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 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. The aTC dilution solutions were made from a stock solution of 1 mg/mL suspended in ethanol and were made fresh prior to the application of the aTC for the M9 culture. After M9 dilution, the strains were grown in 96 well plates to steady state (OD600 of 0.1 – 0.2). We had the constitutive (CpxR-KO) strains transformed with the binding position plasmids to measure constitutive expression of the gene circuit. To normalize growth rate effects on gene expression, we transformed our position plasmids with the TF binding sequence deleted into the CpxR-titration strain. Cells were diluted between 1:2 to 1:3 fold in M9 minimal media in a 96 well cytometer plate prior to data acquisition and cytometry was performed on a MacsQuant VYB. At the beginning of each run, an initial gating strategy involving the Forward Scatter and Side Scatter area

information was used to eliminate background events and samples were run to achieve 60000 gated events for each position strain at a given aTC concentration.

2.3.2 Data Quantification and Statistics

2.3.2.1 Deriving the Concentration Manifold Equation

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):

Equation 2.1

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

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

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

Equation 2.2

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

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

Equation 2.3

Based on this reformulation, the key parameters to consider are the acceleration parameters (couched in the FC_{max} term as described in Figure 1.2) and the stabilization parameters at positions x and y (a total of 4 parameters). In a sense, the "concentration manifold" allows us to remove what we consider to be the extrinsic feature of TF regulation (the binding affinity and TF copy number) from the intrinsic features of the TF regulatory response (the regulatory activity as of the TF on RNAP through stabilization and acceleration). We can further simplify the model by taking into account a judicious binding location for the position x. Taking the +1 binding location, where the assumption of steric hindrance in our thermodynamic model for all the TFs surveyed in this work is

justified, we set β =0 (which makes K=1) and FC_{max}=0. This leads us to the final form of the concentration manifold for a given regulatory position (as a function of the fold-change at the +1 position – $FC^{(+1)}$):

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

Equation 2.4

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).

2.3.2.2 MCMC Sampling of the Posterior Distribution of Stabilization and Acceleration Parameters for Regulated Positions

We used the concentration manifold formalism to delineate the values of the acceleration and stabilization parameters. For positions that showed discernible regulation (12 out of the 22 upstream positions profiled), we assume that the binding affinity is constant between the regulatory positions as the only changing variable is the binding location (the binding sequence is the same) and recast the binned data from the fold-change vs RFP replicates using the concentration manifold formalism as detailed in Methods 2.3.2.1**.** To sample the probability space of the acceleration and stabilization parameters for a given binding location, we started by inferring the joint posterior distribution FC_{max} and K=1+Pβ using a Bayesian approach that relates the parameters underlying our thermodynamic model to the data according to the following relation:

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

Equation 2.5

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

$$L^K\left(FC^k|FC^k_{max}, K^k\right) = \prod_{i}^{DataPoints} Normal\left(FC^k_i|u = FC^k_{prop}\left(FC_i^{+1}; FC^k_{max}, K^k\right), sd = \theta^k\right).$$

Equation 2.6

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

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

Equation 2.7

The crux of the Bayesian approach to model inference is to simulate candidate draws of the joint posterior distribution of the $FC_{max}\ and\ K$ parameters by proposing candidate values from the prior distribution, generating the thermodynamic curve, and evaluating the likelihood function. A transition in the jointly sampled parameter space from one set of parameter values to another is based on the premise that parameter sets will be sampled in proportion to the probability of the posterior distribution (as long as the sampling chain is drawing from the stationary distribution). This process is repeated until a given number of draws have been made from the joint posterior distribution. The results of this inference procedure are shown in Figure 2.5. We used the sampled parameter values for the Markov Chain and plotted the expected model curves in Figures 2.2A and 2.4B. The sampling procedure was implemented with the PyMC3 package that utilizes the NUTS sampler, a particular implementation of the Hamiltonian Monte-Carlo algorithm, to sample the joint posterior distribution.

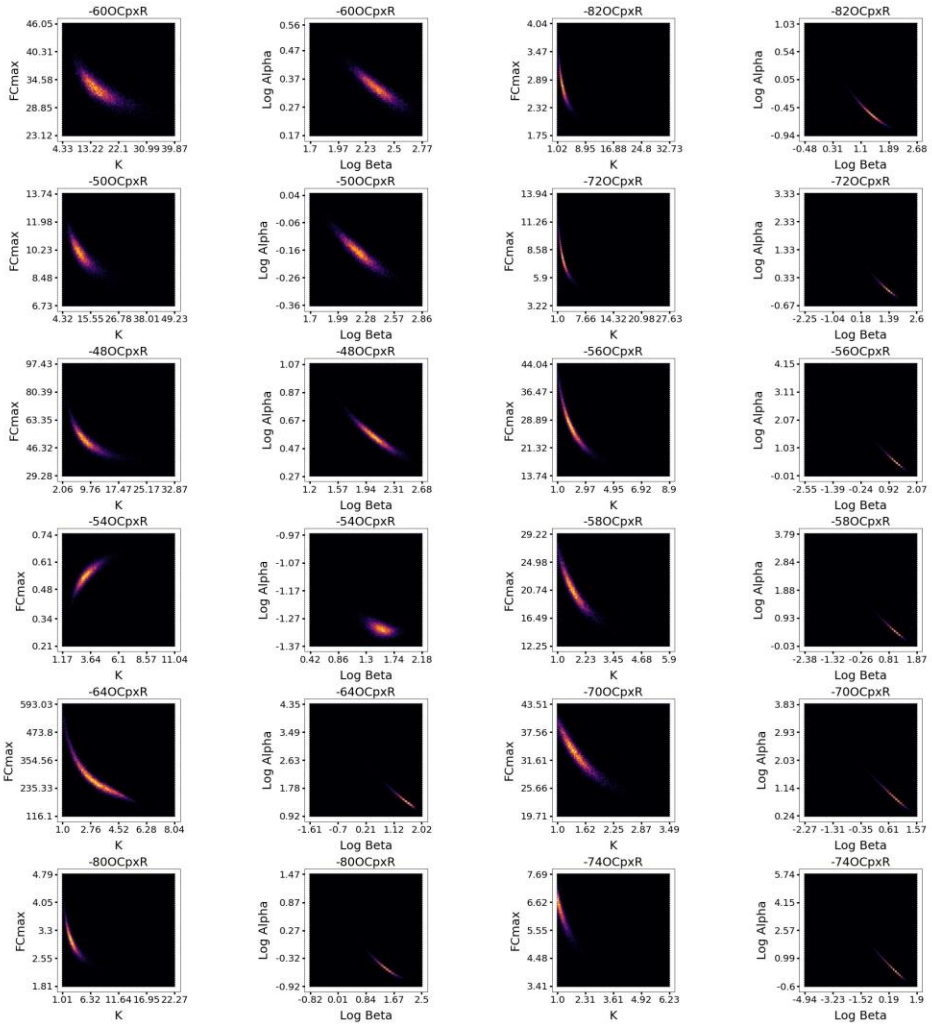


Figure 2.5: Results From MCMC Inference of the Fold-Change Data. The Left Panel represents the results of the inference for the parameters χ *and* FC_{max} while the right panel represents the conversion to α *and* β given the relation in Equation 1.8

We used a uniform distribution as the priors for both the FC_{max} *and* K model parameters with appropriate bounds for each parameter. For K, we ensured that the lowest potential value is 1, in line with the assumptions from the derivation of the concentration manifold formalism. We checked

the inferences from each position to ensure that the bounds we enforced on both parameters were appropriate and that the sampled values were not tending towards the edge of the sample space. Furthermore, we cast the θ parameter (the standard deviation) of the position specific likelihood function as a hyper-parameter in our sampling procedure and set the prior distribution as uniform over a defined interval with the lower bound = 0. Overall, our inference approach allowed us to ensure precise inference of $\alpha \; and \; \beta$ and for 11 out of the 12 regulatory positions (See Table 2.1).

To get the stabilization values from this inference procedure, we simply used the following relation between the K and β and the fact that the polymerase occupancy has a measured value of .0655 in our synthetic promoter,

$$\beta = \frac{K - 1}{P}.$$

Equation 2.8

Given the sta3ilization value for a draw in the chain, we then find the corresponding acceleration value using the jointly sampled FC_{max} value and the relation:

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

Equation 2.9

Table 2.1 lists the inferred acceleration and stabilization parameters. We report the median values of the inference chain along with the bounds that encompass the 68^{th}percentile Bayesian credible interval of the parameters. For position -74, the inference estimates are not precise due to the phenomenon of the ``weak stabilization'' limit as discussed in the Results Section. Examining the posterior distribution of the $FC_{max}\ and\ K$parameters, we find the values of the 68^{th}percentile credible interval for this positions (both $\alpha\ and\ \beta$) to encompass both coherent and incoherent regimes.

2.3.2.3 Using the Concentration Manifold Parameters to Generate the Thermodynamic Model in Fold-Change vs N_{TF} Space.

The parameters inferred using the concentration manifold approach were used to construct the thermodynamic model curves in Figure 2.2A and Figure 2.4B. To accomplish this, we used the $K = 1 + P\beta$ and FC_{max}values inferred for each of the 12 regulatory positions (see Methods 2.3.2.1 and 2.3.2.2). To transform the Markov chain of K values to the χ for the main thermodynamic model 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 position – the +1 position). The binding affinity was treated as a global parameter in our Bayesian inference scheme and was inferred according to the following model:

$$\prod_{k}^{Position} \prod_{j}^{Data\ Points} Normal\left(FC_{j}^{k} \middle| u = FC_{Thermodynamic(j)}^{(k)}(\lambda), sd = \sigma^{(k)}\right)\Bigg)$$

Equation 2.10

with the mean of the likelihood function specified by the thermodynamic model outlined in Section 1.8.4

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

Equation 2.11

The parameter λ, is the global parameter representing the scaled binding affinity of the CpxR TF ($\chi = N_{NS} e^{-\Delta\varepsilon_{TF}}$) and is assumed to be constant across the regulatory positions assessed in this work. FC^{k} is the general thermodynamic model specified in the previous section with our objective to infer λ. For the 12 upstream regulatory positions, as the chains of FC_{max} and K were available from the inference of the concentration manifold dataset, we inferred the global χ 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 and the K was set to 1 in keeping with the assumption of steric hindrance. The mean value of the Markov chain λ inferred from this model, and was used to generate the thermodynamic model curves (mean $\pm 2\theta$) as seen in Figure 2.2A and Figure 2.4B.

2.3.2.4 Processing CpxR Flow Cytometry Analysis Data

For the CpxR position cytometry data, we adapted a robust data analysis procedure (Razomejia2018) 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 off unequal size in RFP space with the number of cells in each bin constant). We then took these binned events and gated them using the log_{10} values of the Forward Scatter and Side Scatter area profiles for each event (referred to as FSC and SSC respectively) to improve the likelihood that the final retained events were single cell measurements. To construct this gate, we computed the mean and covariance matrix for each data set for every RFP bin and used these statistics to fit an ellipsoid to the full data set according to the following formula:

Equation 2.12

$$\begin{bmatrix} \boldsymbol{FSC} \\ \boldsymbol{SSC} \end{bmatrix}^{\boldsymbol{T}} \boldsymbol{\Sigma}^{-\frac{\mathbf{1}}{\mathbf{2}}} [\boldsymbol{FSC} \quad \boldsymbol{SSC}] \leq \boldsymbol{\alpha}$$

Equation 2.13

$$with\ \boldsymbol{\Sigma} = \begin{bmatrix} Var(FSC) & Cov(SSC, FSC) \\ Cov(FSC, SSC) & Var(SSC) \end{bmatrix}$$

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 log_{10} values for the FSC (Forward Scatter) and SSC (Side Scatter), we assume the joint values are normally distributed, which translates to a distance metric that is a chi-squared random variable (the sum of two normally distributed entities is chi-squared with 2 degrees of freedom). We selected α as the 5th percentile of values from the cumulative distribution, and events within the cutoff were taken to be single cell measurements used to compute the fold-change values presented in the results section. The resulting gated events for each of the 16 RFP intervals were pooled from all position strains and replicates, and we excluded events with RFP measurements below a certain threshold determined by visually assessing the the fold-change profile of the control no binding site circuit. The YFP signal for these events had large fluctuations and we reasoned that 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.

Chapter III – Results

Probing the Contributions of Binding Sequence and Position to the Regulatory Modes of Several *E coli* TFs

3.1 - Introduction

TF binding sequence and position are important elements that set the quantitative regulatory function of the TF at different promoters. While the role of TF binding position is well attested in setting the fundamental properties of the thermodynamic curve (Boedicker et al., 2013; Vilar & Leibler, 2003) (see Section 1.8.4 for detailed analysis of the thermodynamic model) analysis of TF binding sequence has often been in the context of motif discovery: statistical methods have been designed to isolate the important nucleotides at TF binding sequences that form the sequence determinants of TF-DNA interaction (Baleja & Sykes, 1991; Benson, Sugiono, & Youderian, 1988; Stormo, 2000) and leverage information on sequence and structure of TF DNA binding domains to predict genomic binding sites (Messeguer et al., 2002; Wasserman & Sandelin, 2004; Weirauch et al., 2014). Understanding the role of sequence in TF binding, in terms of setting TF occupancy and the distribution of binding sites across the genome, has certain implications for TF regulatory function, as TFs with a high number of target sequences would regulate their target promoters distinctly than TFs with only a few binding sites (T. H. Lee & Maheshri, 2012). These approaches have led to the idea of the canonical role of sequence in gene expression: the impact of sequence on the regulatory function of a TF is

often construed as acting on the binding affinity of the TF to its operator site (Barker et al., 1998). There has been considerable investigation into the role of operator sequence in regulating the search kinetics of TFs that has unmasked important kinetic steps of TF activity under the direct control of sequence(Berg, 1981; Halford & Marko, 2004; Niland et al., 1996). With this information, modelling the function of binding sequence in the thermodynamic model is often viewed in the context of tuning overall TF occupancy through changes in the free-energy associated with TF-DNA interaction (Ackers et al., 1982). However, quantitative investigations on the role of LacI operator sites in gene expression have shown that certain sequences have a material impact on regulation outside of TF occupancy (Garcia et al., 2012). Furthermore, analysis of DNA binding motifs from TFs bound individually and TFs interacting cooperatively open the possibility that TF-DNA interactions facilitate conformational change (Inukai et al., 2017; Morgunova & Taipale, 2017) – allowing for TFs to interact with neighboring molecules. This fact suggests a potential role of sequence in setting the regulatory modes of TFs, however, there has been little systematic attempt to isolate its contribution to the TFs ability to stabilize and accelerate RNAP.

To assess the spectrum of regulatory responses from Transcription Factors in *E.coli*, we used a library of strains where the concentration of any TF in the cell is titratable with the small molecule inducer, aTC (anhydrous tetracycline) (Parisutham et al., 2022). This system allows for precise control of TF copy number within the cell for each of the TFs we investigate. In this library the endogenous TF was knocked-out of its native locus, and the TF gene was cloned into an expression construct under the regulation of the Tet

promoter. This construct was integrated into the *ybcn* locus in an *E.coli* strain expressing the TetR repressor. Each of these TFs are fused to the mCherry open reading frame through a linker sequence, which allows us to measure the input level of TF in flow cytometry at the single-cell level (Guharajan, Chhabra, Parisutham, & Brewster, 2021; Parisutham et al., 2022). We selected 93 TFs based on comprehensive regulatory information that included the promoters they bind to, the TF binding sequences they interact with and the qualitative regulatory function of the TF (activation or repression) from the RegulonDB. These TFs represent 26 out of the 40 TF families grouped according to sequence conservation of the respective DNA binding domains and regulate genes involved in diverse cellular processes such as cellular respiration, membrane-stress response, nutrient switching, and a host of other cellular processes (Keseler et al., 2021; Salgado et al., 2023). We selected an average of 2 binding sequences for each of these TFs based on features that included the evidence for the interaction, the degree of similarity to the consensus sequence, among other factors (See Methods 3.3.1 for details on binding site curation). The titratable library allowed us to decouple the induction of a TF of interest from its native genetic context, giving a throughput advantage by expressing all 93 TFs under the same induce and in a single growth medium (M9 minimal media supplemented with glucose).

We selected binding locations to profile the regulatory activity of these TFs by evaluating the distribution of binding locations of *E.coli* TFs relative to the Transcription Start Site (TSS).

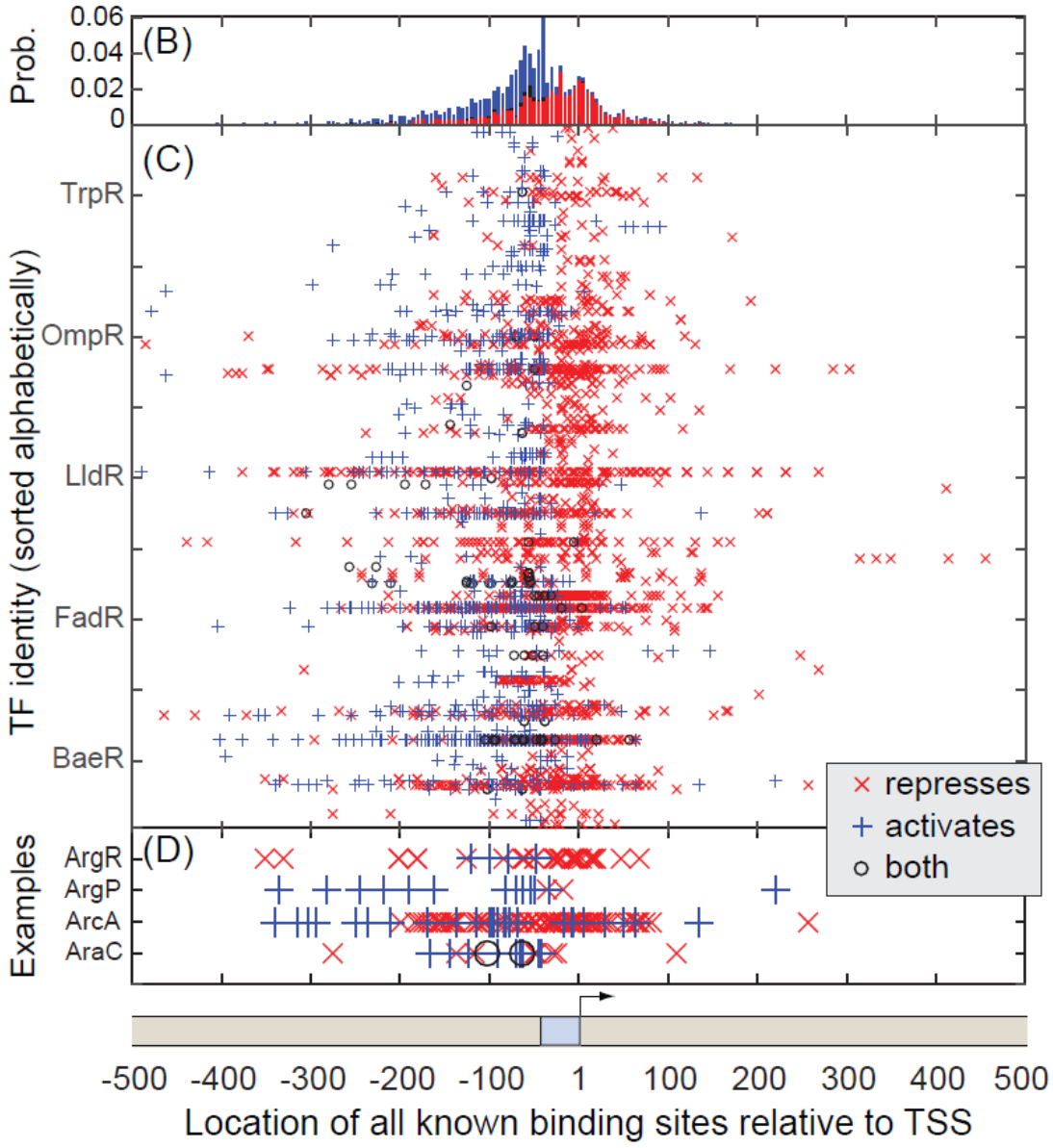

Prob.
0.06
0.04
0.02
0
(B)
(C)
TF identity (sorted alphabetically)
TrpR
OmpR
LldR
FadR
BaeR
(D)
Examples
ArgR
ArgP
ArcA
AraC
represses
activates
both
-500 -400 -300 -200 -100 1 100 200 300 400 500
Location of all known binding sites relative to TSS

Figure 3.1: Survey of Qualitative Regulatory Function Across Binding Positions for All Annotated *E.coli* TFs. TFs are arranged alphabetically from bottom to top on the Y-axis, and the regulatory function is as indicated. Note the histogram at the top demonstrating a clear position dependence on TF function. As most activator and repressor interactions are contained within +20 to -100bps from the promoter, we decided to use -103 as our terminal position in evaluating the regulatory function of the chosen 93 TFs.

In Figure 3.1, we show the binding position and qualitative regulatory function for all annotated TFs. Note that the while activating and repressing interactions are scattered throughout binding position, a pattern emerges where TF binding positions situated proximal and downstream to the promoter are repressive, and the majority of activating interactions take pace an upstream window between -40 and -80 nucleotides. With this information in hand, we selected 3 binding positions to evaluate the isolated regulatory effect of the TF: the +1 position (immediately downstream of the promoter), the -57 position (a site commonly associated with activation) and the more distal -103 position to determine how the response function of the TF varies across a wide spatial range.

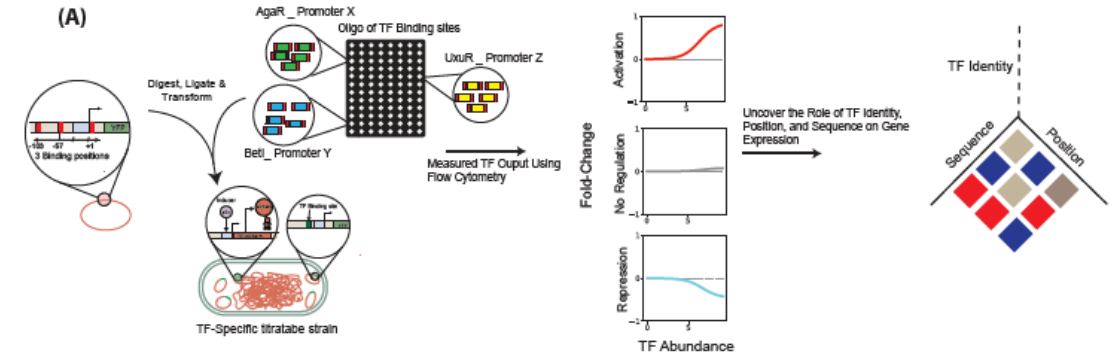
3.2 - Results

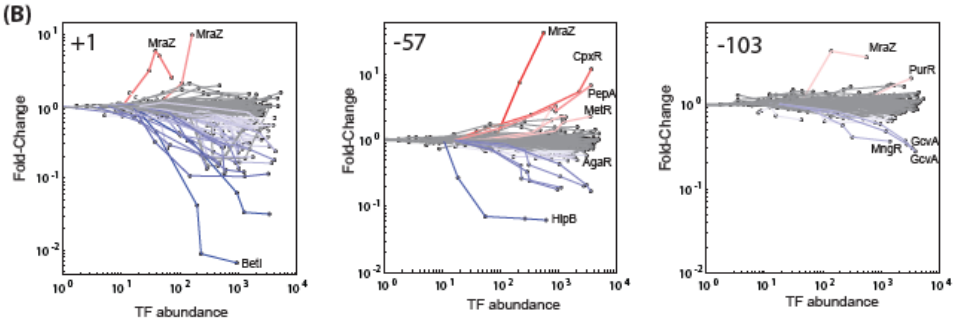
3.2.1 Assessing the Isolated Regulatory Function of 93 *E.coli* TFs

To comprehensively assess the regulatory function for a broad number of *E.coli* TFs, we selected a total of 93TFs that were included in our library and profiled the RegulonDB database (Salgado et al., 2023) to extract binding sequences with strong information content for each of these TFs. We curated a list of TF binding sequences across all

annotated promoters and retained sequences in which there was either direct evidence of TF binding or strong imputation of binding. We then filtered each of the binding sites to ensure that the regulated promoters were sigma70 promoters and selected on average 2 binding sequences per TF.

In total, 186 TF binding sequences (Table 6.1) were selected that covered 93 TFs. These sequences were spliced into the landing pad gene constructs using the single digestion and ligation procedure described in Chapter 2 into the 3 positions relative to the sigma70 promoter sequence as discussed previously (Figure 3.2A). By controlling both the position and sequence of a given TF in a designed regulatory circuit, we sought to investigate the role of these features on the quantitative response function of the TF. A broad overview of the regulatory function of these 93 TFs at each position is shown for the three binding positions (Figure 3.2C). Here, we group the regulatory response according to TF families: the rows represent the 3 binding positions ordered from proximal to distal and the color map represents the average for each TF at the 2 highest induction conditions. The area of the squares is partitioned based on the number of binding sequences assayed for that TF, as such each section of a square represents measurements of regulation for a given TF at a given position for different binding sequence. Most TFs had two binding sequences except for TFs CRP and Fis (3 binding sequences) and a few other TFs with only one binding sequence. The full TF-titration response curves for all the data are shown in Figure 3.2B.


(A)
AgaR _ Promoter X
Oligo of TF Binding sites
UxuR _ Promoter Z
BetI _ Promoter Y
Digest, Ligate & Transform
Measured TF Ouput Using Flow Cytometry
TF-Specific titratabe strain
Activation
No Regulation
Repression
Fold-Change
TF Abundance
Uncover the Role of TF Identity, Position, and Sequence on Gene Expression
TF Identity
Sequence
Position


(B)
+1
-57
-103
MraZ
CpxR
PepA
MetR
AgaR
HipB
PurR
MngR
GcvA
BetI
Fold-Change
TF abundance

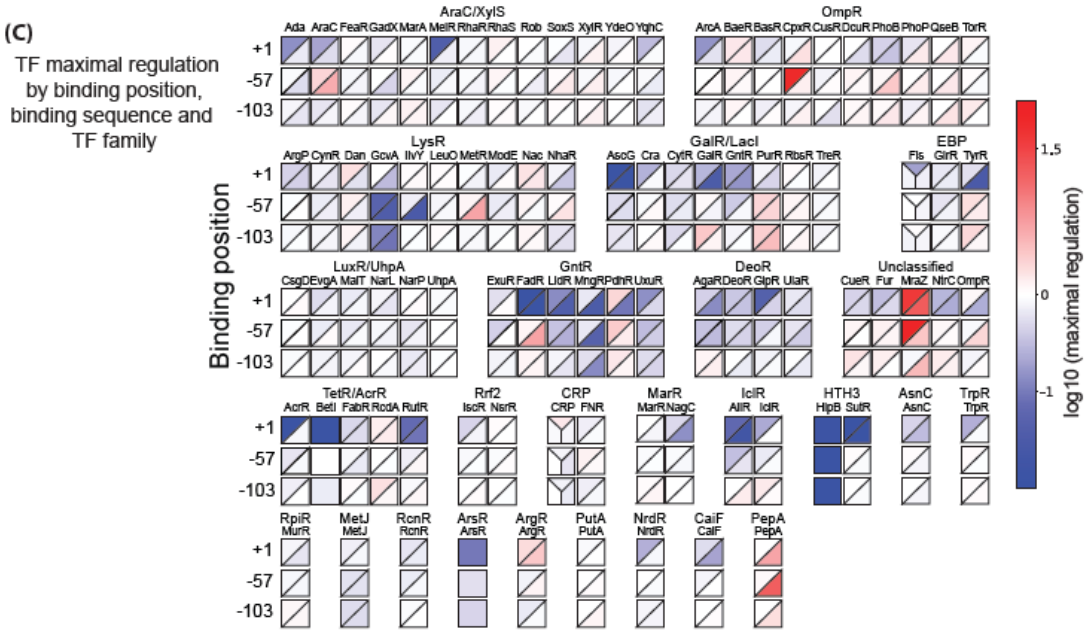

(C)
TF maximal regulation by binding position, binding sequence and TF family
AraC/XylS
Ada AraC FeaR GadX MarA MelR RhaR RhaS Rob SoxS XylR YdeO YqhC
OmpR
ArcA BaeR BasR CpxR CusR DcuR PhoB PhoP QseB TorR
LysR
ArgP CynR Dan GcvA IlvY LeuO MetR ModE Nac NhaR
GalR/LacI
AsoG Cra CytR GalR GntR PurR RbsR TreR
EBP
Fls GlrR TyrR
LuxR/UhpA
CsgD EvgA MalT NarL NarP UhpA
GntR
ExuR FadR LldR MngR PdhR UxuR
DeoR
AgaR DeoR GlpR UlaR
Unclassified
CueR Fur MraZ NtrC OmpR
TetR/AcrR
AcrR BetI FabR RodA RutR
Rrf2
IscR NsrR
CRP
CRP FNR
MarR
MarR NagC
IclR
AllR IclR
HTH3
HipB SutR
AsnC
AsnC
TrpR
TrpR
RpiR MurR
MetJ MetJ
RcnR RcnR
ArsR ArsR
ArgR ArgR
PutA PutA
NrdR NrdR
CaiF CaiF
PepA PepA
Binding position
+1
-57
-103
1.5
0
-1
log10 (maximal regulation)

Figure 3.2. Isolating Quantitative Regulatory Profiles for 93 *E.coli* TFs Using a Synthetic Biology Approach

(A) Thruput Cloning Procedure and measurement scheme to assay the regulatory response of *E.coli* TFs: 186 TF Binding Sequences were selected that encompassed 93 E.coli TFs with anaverage of 2 binding sites per TF. The binding sequences were cloned into the Thruput position cloning vectors used in Chapter 1 at 3 binding positions relative to the Sigma70 promoter (+1. -57, and -103). The circuits were transformed into the *E.coli* TF titration strains, cultured in M9 media (glucose supplemented) with aTC, and measured for YFP expression via flow cytometry. Overall, we profiled the Fold-Changes of approximately 558 *E.coli* strains to map the quantitative output of TFs as a function of TF identity, binding position, and sequence. **(B) The full induction response for the186 binding sequences evaluated at the 3 binding positions**. Solid lines are colored to reflect the heat map in (C) reflecting the degree of activation or repression. **(C) Position and Sequence Dependent Regulatory Profiles for 185 TF Binding Sequences:** The profiles are grouped according to TF Families, with each square representing a TF with the area of the square partitioned based on the number of sequences assayed for the TF.

In Figure 3.2C, we present an atlas of quantitative regulatory information for 93TFs regulating 2 binding sequences at 3 distinct positions relative to the promoter. We see that the response function of these TFs allows us to interpret data gathered from molecular genetics experiments on TF function. For example, the UxuR and ExuR TFs are known paralogues that regulate hexuronate metabolism and bind interchangeably to their cognate sequences as the TFs share a conserved DNA binding motif (Rodionov, Mironov, Rakhmaninova, & Gelfand, 2000; Suvorova et al., 2011). Here, we demonstrate that while both TFs directly repress transcription, the quantitative activity of the UxuR is noticeably stronger than its counterpart at the downstream position, suggesting either an

increased binding affinity of this paralogue to the sequence or potentially a stronger maximal regulatory effect. Furthermore, comparison within the GntR family reveals that the TF FadR has both activating and repressing interactions across binding position much like the AraC TF from the AraC/XylS family, and that MngR has a remarkably consistent regulatory profile across a range of binding positions.

For the downstream position (+1, top row in each family), we find most of the TFs act as repression. This is expected as binding in the downstream region immediately adjacent to the TSS is a hallmark of gene repression where the TF sterically hinders RNA polymerase binding and initiation of transcription. However, this rule does have an interesting exception with the TFs MraZ (Unclassified Group)- and PepA, suggesting that the proximal downstream position does allow for some TFs to interact with RNAP in a way that facilitates the transcriptional process. However, it is important to note that the mechanism of activation for TFs binding at overlapping positions downstream to the sigma70 promoter is unclear. An important consideration would be to determine if there is a stabilizing component to the regulatory response for both the MraZ and PepA TFs at position +1. The strongest repressor, at this position was BetI, repressing about 100-fold, which was followed by the TFs HipB and GcvA. Looking at TF family groupings, we find that the TF families *GntR, GalR, TetR/AcrR*, and *HTH-3* all strongly repress at the proximal downstream position.

Regulation at -57 reveals more occurrences of activation, with the strongest activators, MraZ, PepA, and CpxR all increasing gene expression more than 10-fold and the weaker activators, FadR, AraC, MetR, and Nac, – demonstrating more than 2-fold activation. The

degree of activation for CpxR was in line with our expectations as -57 corresponds to the -64 position in the CpxR position sweep data (Figure 3.2C). At -103, we see the TF MraZ and PurR activating expression. Both -57 and -103 see moderate number of TFs repressing, with the strongest repressors GcvA, MngR, and HipB decreasing expression by more than 10-fold. However, the number of activating TFs seen in our data is less than expected, as several of the 93 TFs have annotated activating functions that were not recapitulated in these measurements. This result could be due to the fundamental regulatory property of the TF, i.e the annotated function does not capture the isolated regulation of the TF, or the fact that our culture conditions did not create the conditions necessary for induction of TF activity. Overall, our measurements revealed about 50 TFs with robust differences in Fold-Change, which we defined as more than 2-fold activation or repression, in at least one binding position and with one sequence. The fact that roughly half of the TFs had muted Fold-Change profiles might suggest that the fusion of the TF via the linker domain to the mCherry cassette compromised its regulatory function and highlights a need to potentially profile un-fused versions of the TF to verity this. A second possibility is that the positions selected do not adequately cover the regulatory activity of the remainder TFs, and that further interrogation along the position axis might demonstrate robust regulation from these factors.

3.2.2 - Evaluating the Relative Contributions of TF Binding Sequence and Position to the Regulatory Output of 93 TFs.

Armed with the quantitative regulatory response for a subset of *E.coli* TFs, we were motivated to determine the precise contributions of TF binding sequence and TF position

to the regulatory outcome. Specifically, we aimed to determine the relative contributions of these 2 elements of promoter architecture to the maximal regulatory response of the thermodynamic model. As the FC_{max} parameter encompasses both the TFs ability to stabilize and accelerate RNAP, determining if position or sequence are necessary to set this parameter would inform the contributions of each feature to the TF regulatory mode. A simple idea is that the TF binding sequence plays a fundamental role in setting the binding affinity, while position sets aspects of TF occupancy and crucially, the maximal regulatory output by modulating the dual regulatory modes.

Here we test these assumptions for FC_{max} by contextualizing the thermodynamic model in a Bayesian inference scheme and extract the degree of importance of TF binding sequence and position to this parameter (See Methods 3.3.2 for details).

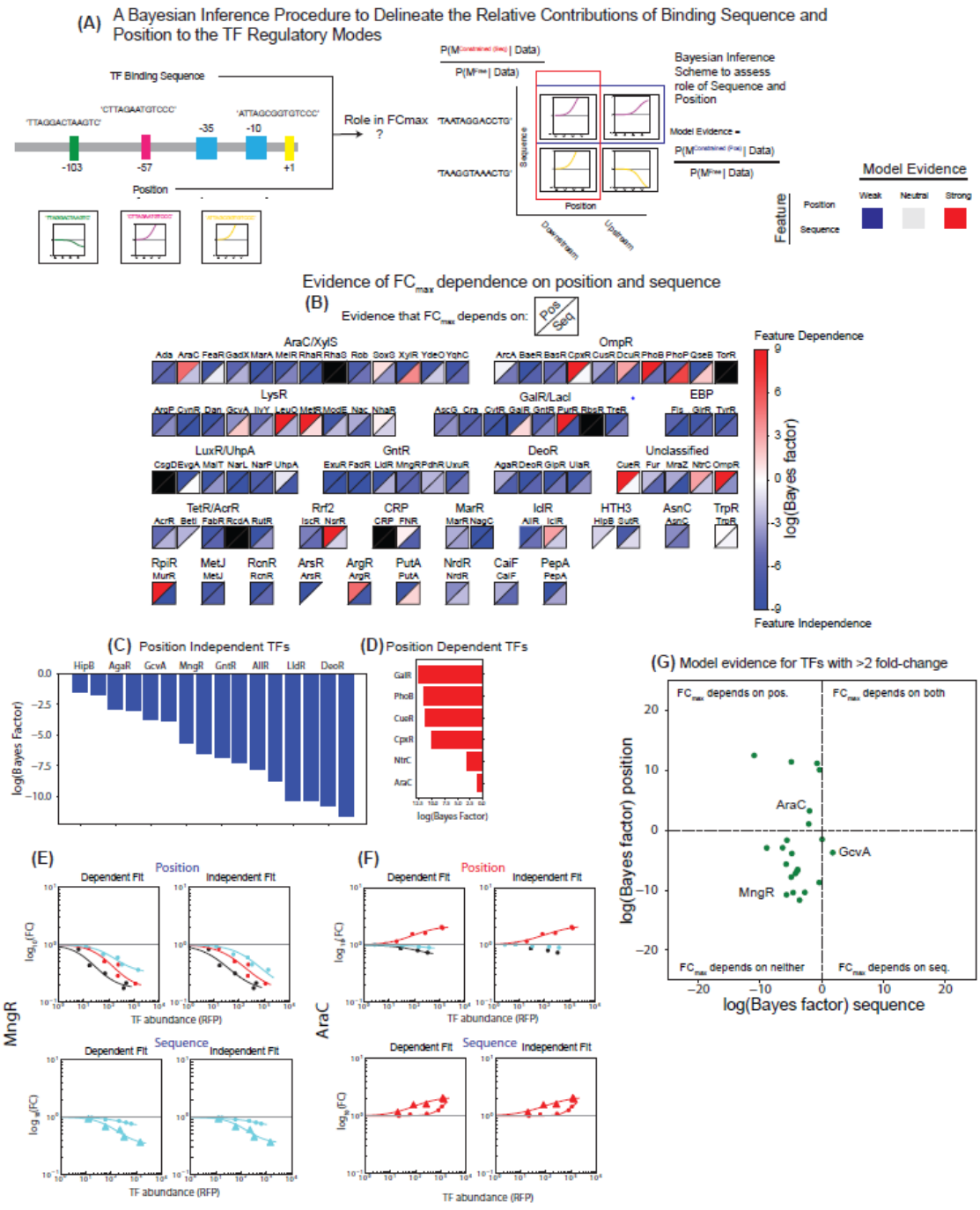

(A) A Bayesian Inference Procedure to Delineate the Relative Contributions of Binding Sequence and Position to the TF Regulatory Modes
TF Binding Sequence
'TTAGGACTAAGTC'
'CTTAGAATGTCCC'
'ATTAGCGGTGTCCC'
-35
-10
-103
-57
+1
Position
Role in FCmax ?
P(M | Data)
P(M | Data)
'TAATAGGACCTG'
'TAAGGTAAACTG'
Sequence
Position
Downstream
Upstream
Bayesian Inference Scheme to assess role of Sequence and Position
Model Evidence
Model Evidence
Weak
Neutral
Strong
Feature
Position
Sequence
Evidence of FCmax dependence on position and sequence
(B)
Evidence that FCmax depends on:
Pos
Seq
AraC/XylS
OmpR
LysR
GalR/LacI
EBP
LuxR/UhpA
GntR
DeoR
Unclassified
TetR/AcrR
Rrf2
CRP
MarR
IclR
HTH3
AsnC
TrpR
RpiR
MetJ
RcnR
ArsR
ArgR
PutA
NrdR
CaiF
PepA
Feature Dependence
Feature Independence
log(Bayes factor)
(C) Position Independent TFs
log(Bayes Factor)
HipB
AgaR
GcvA
MngR
GntR
AllR
LldR
DeoR
(D) Position Dependent TFs
log(Bayes Factor)
GalR
PhoB
CueR
CpxR
NtrC
AraC
(G) Model evidence for TFs with >2 fold-change
FCmax depends on pos.
FCmax depends on both
FCmax depends on neither
FCmax depends on seq.
log(Bayes factor) position
log(Bayes factor) sequence
AraC
GcvA
MngR
(E)
Position
Dependent Fit
Independent Fit
MngR
TF abundance (RFP)
Sequence
Dependent Fit
Independent Fit
TF abundance (RFP)
(F)
Position
Dependent Fit
Independent Fit
AraC
TF abundance (RFP)
Sequence
Dependent Fit
Independent Fit
TF abundance (RFP)

Figure 3.3: The Position and Sequence Axes are Decoupled in Setting the TFs Maximal Regulator Output

(A) The role of TF Binding Sequence and TF location in gene regulation. Different sequences interface with binding location to set the regulatory response - yet the precise contributions of each to the regulatory modes (represented as FC_{max}) are not know for most TFs. **(B) The results of the Bayesian Inference Scheme for the Position and Sequence Constrained Models**. The Heat Map indicates the Bayes Factor -the ratio of the model evidence between the independent and constrained models. The squares are partitioned with the position model evaluation in the top-half and the sequence model in the bottom-half. The larger the value, the greater the relative importance of the feature (TF binding position or sequence) (**C,D**) Model evidence of representative TFs with robust regulatory responses (greater than 2-fold activation or repression). TFs in blue do not require binding position to set the FC_{max} parameter, i.e, the regulation across +1, -57, and -103 positions is effectively captured by a single parameter. The converse is true for the TFs in red **(E,F) Examples of robust regulators with different profiles of position and sequence constraints.** Blue Outline indicates a promoter feature (position or sequence) is constrained while the Red Outline indicates independence. Data points are as follows: cyan is the -103 position, red is the -57, and black is +1. Different sequences are represented by circles and triangles. For MngR, the TF is captured by the model in which position and sequence can be constrained. For AraC, the different positions require their own FC_{max} parameter but the sequences are constrained. **(G) Plot of the Bayes Factor for Position and Sequence**. Each point is a robustly regulating TF. Note that several robust regulators are constrained by both sequence and position. The only TF with sequence acting as an important feature in setting the regulatory mode is GcvA.

In Figure 3.3A, we outline this procedure. The crux of this approach is to compute a quantity known as the Model Evidence associated with a given feature (TF binding position or sequence). The space of possible models in the thermodynamic framework that incorporate sequence and position are vast, with these features potentially involved in setting both TF occupancy and the maximal regulatory response (FC_{max}). Here we

constrain the models to evaluate our hypothesis of TF regulation – that sequence and position play a role in contributing to the FC_{max} parameter. As the maximal regulatory output encapsulates both the TFs ability to (de)stabilize RNAP and (de)accelerate transcription initiation (Figure 3.3A, Section 1.8.4), sequence and position must play a role in shaping the parameter if these features do indeed affect the TF regulatory modes. Therefore, our inference procedure compares the evidence in favor of a model where position (or sequence) is constrained in setting FC_{max} - i.e a single FC_{max} value can sufficiently describe the TF regulatory response across different positions or sequences (feature independence) - to a model where each regulatory curve associated with a given position or sequence requires a unique FC_{max} (feature dependence). In short, we evaluate the ratio of model evidence for the scenario in which FC_{max} is independent of the specified feature, to the scenario in which FC_{max} is dependent on that feature (See Panel 10A). If the feature is dispensable in setting the regulatory mode, the inference would assign the feature independence model with a higher probability than the feature dependence model (Figure 3.3A) and vice versa. The ratio of the model evidence between feature independence and dependence is the Bayes Factor, representing a rigorous statistical evaluation of the role of sequence or position in setting the regulatory mode for a given TF.

The results of the inference procedure are presented in Figure 3.3B. Here we plot the model evidence for both the position and sequence features for a given TF. The color scheme indicates the model evidence in favor of the unconstrained (feature dependence) model (red) on one end of the spectrum and the constrained (feature independence)

model (blue) on the other. ***We note that for much of the regulatory response across TFs, both the position and sequence models can often be constrained as seen in the predominance of low model evidence in favor of the FC_{max} as an independent parameter.*** This suggests that the key factor in setting the regulatory modes of these TF is TF identity, and not the extrinsic features of gene regulation such as TF binding position and sequence. Looking at several TFs that robustly regulate at multiple binding positions and across both sequences, we plot the model evidence in Figure 3.3C demonstrating that regulation at these multiple positions is specified by a singular regulatory effect – the regulatory response across 3 distinct positions (+1, -57, -103) can effectively be captured by a single maximal fold-change (FC_{max}). However, TF binding position matters for select TFs: In Figure 3.3D we show the large model evidence for 6 TFs, suggesting each binding position specifies a unique regulatory function. This stands in contrast to TF binding sequence, where the model evidence is weak across all TFs except one (GcvA – see Figure 3.3G).

In Figure 3.3E and 3.3F, we show 2 examples of TFs classified according to different model outcomes. For MngR, there is discernable regulation at each of the 3 binding positions and both binding sequences, yet the position (left) and sequence (right) curves are effectively set by a single FC_{max} for each feature. The TF AraC highlights the importance of position for that TF, where the response is 2-fold activation for -57 but repression at +1, suggesting position must play a role in setting the maximal response. However, across binding sequence, we see no reason for a different FC_{max} to explain the data at the -57 position. Both sequences are effectively captured by the same parameter,

indicating that the AraC's ability to stabilize and accelerate RNAP is independent of the sequence it interacts with.

Plotting the model evidence for TFs along the axes of position and sequence, we clarify the contribution of these features (Figure 3.3G). Here we only included 22 TFs from the analysis, as these TFs regulated robustly in at least 2 positions and at both binding sequences. We see the variability of the data distributes along the position axis and **conclude that the axes of sequence and position are decoupled in setting the maximal response for these 6 TFs – revealing an underlying gene regulatory principle for these factors.** Taken together, our inference approach leads us to the conclusion that most of the TFs considered (16 out of the 22) do not require sequence and position to set the TFs regulatory modes, and only effectively contribute to TF occupancy at the promoter. However, we do find that position is uniquely important for select TFs, with only one TF (GcvA) requiring sequence as a key feature in setting the regulatory modes of the TF.

3.3 - Methods

3.3.1 186 TF Binding Sequence Selection from RegulonDB

The TF binding sequences used in interrogating the role of position and sequence on TF regulatory modes were curated from RegulonDB (Salgado et al., 2023). We performed an initial inventory of TFs in our synthetic library and determined which of these TFs overlapped with detailed information on TF binding sequence in this database. We selected 93 TFs with information on binding position of TFs throughout the genome,

their qualitative regulatory function (activation or repression) and the degree of evidence for the binding and regulation. We then filtered the sequences to promoters that were bound by sigma 70 and selected the two (or three) binding sequences. These sequences were then cloned into the throughput gene circuits described in the previous results section to be measured in their respective TF titration strains.

3.3.2 – Bayesian Inference Scheme to Evaluate the Model Evidence of Sequence and Position Contributions to TF Regulatory Output

To decouple the role of binding position and sequence on TF regulatory response, we applied a Bayesian inference approach to assess the degree each of these promoter features explain the parameters in the thermodynamic model. This model of simple TF regulation (See Figure 1.2 for details) parametrizes TFs regulatory effect in terms of the regulators occupancy (χ) and its maximal output (FC_{max}). The fundamental regulatory modes of the TF, its ability to stabilize RNAP binding (β) and accelerate transcription initiation (α), are embedded in these χ and FC_{max} terms (See Section 1.8.4 for explanation on the Thermodynamic Model). Crucially the stabilization term features in both χ (setting the effective TF occupancy) and FC_{max}, while the acceleration parameter (α) is only in the FC_{max} term. The position axis is well attested in the setting the TF regulatory modes (Besse et al., 1986; Guharajan et al., 2021). Furthermore, it is generally established that prokaryotic TF binding sequences are information dense (Halford & Marko, 2004; Kleinschmidt et al., 1988; R. C. Lawson, Jr. & York, 1987), and these TFs react specifically with given sequences to set the binding affinity in the thermodynamic

model. However, the role of TF binding sequence in setting the maximal response (FC_{max}) is unclear: conformational change brought about by the interaction of TF to different binding sequences could impact the TFs ability to stabilize and accelerate gene expression.

This problem can be framed as inference on the model evidence for sequence (or position) in setting the FC_{max} parameter as it contains both the regulatory modes of the TF. Specifically, we ask if the measured curves for a given TF – consisting of $p\ x\ s$ curves where p is the number of positions and s is the number of TF binding sequences – require unique FC_{max} values to explain the response for different sequences/position or a single FC_{max} response to explain the data across a given feature. We call this the degree of independence of the feature, and can be evaluated formally using Bayesian inference by evaluation of the model evidence in favor of the free/independent model (where FC_{max} is not constrained across the feature) to the constrained model (where a single FC_{max} value is used to set the TF response across the different curves)

Equation 3.1 - $P(M|D) = \frac{P(D|M)_{*P(M)}}{P(D)} = \frac{P(M)}{P(D)}\left[\int_{\theta lb}^{\theta ub} \mathcal{L}(D|\theta) * p(\theta|M)\, d\theta\right]$

Where M represents the model and D represents the data. The distribution of $\mathcal{L}(D|\theta)$ represents the likelihood of the data, $p(\theta|M)$ is the prior parameter distribution of the data, and P(M) and P(D) are the prior distributions of the model and data respectively. The prior distribution over the parameter space was modeled as follows:

Equation 3.2 – $P(\theta|M) = P(\chi, FC_{max}|M) = P(\chi|M)P(FC_{max}|M); with\ P(\chi|M) \sim Uniform(1e^{-9}, 0)\ and\ P(FC_{max}|M) \sim Uniform(0,100)$

The key step is the evaluation of the integral over the entire parameter space θ that includes parameters that are local to a specific curve (χ and FC_{max} in the free model) or parameters that are global, i.e shared between curves (FC_{max} in the constrained model). For evaluation of the sequence model, this is:

Equation 3.3 – $Const_{seq} = \theta_{const\_seq}^{\rightarrow} = [\chi_{1,1}\chi_{1,2}, \dots \chi_{s,p}, FC_{max1}\ FC_{max2}, \dots FC_{maxp}]$

Where the total number of occupancy parameters are s x p (the total number of curves) and the number of FC_{max} parameters is the number of positions.

For the Position Model, as we have 3 binding positions (+1, -57, -103) giving us the following parameters

Equation 3.4 – $Const_{pos} = \theta_{const\_pos}^{\rightarrow} = [\chi_{1,1}\chi_{1,2}, \dots \chi_{s,p}, FC_{max1}\ FC_{max2}, \dots FC_{max\_s}]$

Where the total number of occupancy parameters is the same as the sequence model and the number of FC_{max} parameters is the number of sequences.

The Free (Independent) Model has the the following number of parameters:

Equation 3.5 – $Free = \theta_{Free}^{\rightarrow} = [\chi_{1,1}\chi_{1,2}, \dots \chi_{s,p}, FC_{max_{1,1}}, FC_{max_{1,2}}, \dots, FC_{max_{s,p}}]$

Where the number of FC_{max} parameters is p x s (the total number of c

The Likelihood function can be represented as a product of normal distributions, with mean~0 and a set standard deviation

Equation 3.6 – $\mathcal{L}(D|\theta) = \prod_{i=1}^{p}\prod_{j=1}^{s}\prod_{k=1}^{d} N_{i,j,k}(\mu = 0, \vartheta)$

Which follows the likelihood the product of the residuals over all the curves for a given TF with a total of p(positions) x s(sequences) x d(data points) as the residuals are independently and normally distributed. Evaluation of the model evidence requires an integration over an N-dimensional parameter space which is infeasible as the number of parameters is non-trivial. To circumvent this, we adopted an Approximate Bayesian Computation approach that replaced the integral with a sum over the parameter draws randomly sampled over the space, approximating the likelihood with the following form:

Equation 3.7 – $\mathcal{L}(D|\theta) \sim \mathcal{L}(D|\theta, d = \varepsilon) \sim Delta\_Function(|r| < \varepsilon = 1\,; |r| > \varepsilon = 0\,)$

Where the likelihood is now conditioned over distance representing an arbitrary value cut-off value for the residual sum of squres (RSS). Modelling the likelihood as a Dirac Delta function ensures that only parameter draws that provide a reasonable approximation to the data (as set by ε) are accepted in the tally. The model evidence is then a weighted sum of $p(\theta|M)$ distribution with the weights as 0 and 1 (rejection or acceptance) based on the sample draw and the RSS associated with it. The weighted sum of the priors is where the model discrimination manifests. The free model has more parameters than the constrained one and therefore fits the data better resulting in more increments to P(D|M). However, as there are more parameters, $p(\theta|M)$ is naturally smaller than the constrained model (all priors are bounded between 0 and 1) – conforming to Occam's Razor (out of a set of competing models, the more parsimonious

one that explains the data is accepted). We then compare the evidence between the two models as follows:

Equation 3.8 –

$$log\left(\frac{P(M_{free}|D)}{P(M_{const}^{pos}|D)}\right) \sim Bayes\_factor^{(pos)} \; ; \; log\left(\frac{P(M_{free}|D)}{P(M_{const}^{seq}|D)}\right) \sim Bayes\_factor^{(seq)}$$

If the Bayes Factor is larger than 1, that would indicate the free model outperformed the constrained one and implies that the promoter feature under consideration (sequence or position) is essential in setting FC_{max} and consequently, the regulatory modes of the TF. For computation of the Bayes Factor we simulated draws for the free and constrained modes of the χ *and* FC_{max} parameters from the uniform prior distribution by randomly sampling two numbers and taking the average to be used as the parameter value for computation of the residual. If the residual fell below the cut-off value, the value of $p(\theta|M)$ was incremented to the initialized P(D|M) variable. We performed the simulation for 10000 runs and then computed the Bayes factor at the end for each of the 93 TFs.

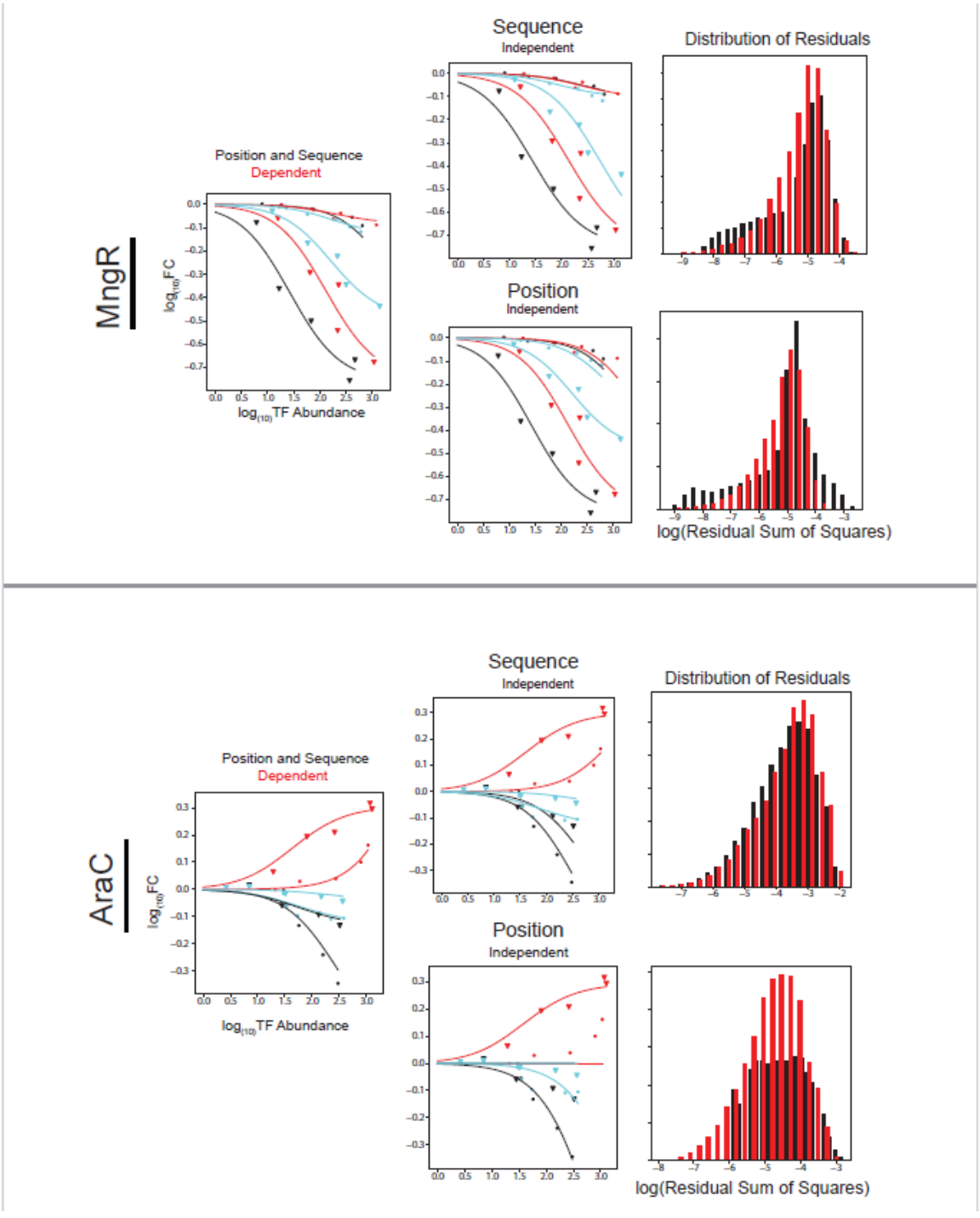

MngR
Position and Sequence
Dependent
log(10)FC
log(10)TF Abundance
Sequence
Independent
Position
Independent
Distribution of Residuals
log(Residual Sum of Squares)
AraC
Position and Sequence
Dependent
log(10)FC
log(10)TF Abundance
Sequence
Independent
Position
Independent
Distribution of Residuals
log(Residual Sum of Squares)

Figure 3.4 Breakdown of Sampling Procedure and Distribution of Residuals for Distinct TFs. Calculation of the model evidence requires sampling over the parameter space and computing the distribution of residuals across the regulatory curves. In **(A),** we demonstrate the results of the sampling for evaluation of the contribution of the features to the regulatory modes for the MngR TF. The first plot from the left represents the dependence model: all regulatory curves (across position and sequence) are specified by their unique FC_{max}.parameter followed by the position and sequence constrained models. The model lines are the result of sampled parameters that generated residuals that fell below the cutoff. For MngR, the constrained models capture the data well – on par with the feature dependence model. This is reflected in the distribution of the residuals in the last sub-panel (red is dependence, black is constrained). Colors in the scatter plot represent the 3 binding positions (+1 - black, -57 - red, and -103 -cyan) and the different markers represent distinct binding sequences. **(B)** Evaluation of the AraC TF. Note that while the sequence constrained model captures the data equally as well as the feature dependence model, the position constrained model cannot explain the regulatory curves – demonstrating the importance of position as a free parameter and its role in shaping the regulatory modes of the TF.

As demonstrated in Figure 3.4, we breakdown the sampling procedure for both the MngR (Panel A) and AraC (Panel B) TFs. We plot the regulatory curves across position and sequence on the same plot, with colors representing different positions and the markers representing different binding sequences. For evaluation of the position independence model, the algorithm takes a given sequence, and assumes all the curves across different positions have the same FC_{max}. It then samples over the parameters unique to each of the regulatory curves ($\chi -$ TF occupancy) and samples a single FC_{max} value to represent each of the curves. For the feature dependent model, each regulatory curve gets its unique FC_{max} value. The thermodynamic model is then computed for all the position specific curves related to the chosen sequence with the sampled parameters, and the output is then

subtracted from the data to compute the residuals for both models. The sum of squares of the residuals (RSS) is then compared to an arbitrary cut-off as mentioned previously. Figure 3.4A plots the residuals, sampling both the feature dependent and independent models for MngR where the distribution of residuals between the constrained and feature dependence modesl are close- implying that increments to the posterior function for each model would be of similar frequency. However, at every increment, the dependent model increments a lower value (due to the larger number of parameters to explain the data) than the constrained model, and accounts for the large difference in model evidence in favor of independent model across both features as described in the Results section. This is in contrast with the AraC TF (Figure 3.4B), where we observe the position feature accesses smaller residual values in the distribution for the feature dependence model. This is reflected in the model curves unable to capture the data well for this feature.

3.3.3 – Cloning of TF Binding Sequences (TFBS) Strains and Culture Conditions for Growth and Measurement of Sequence-Position Promoter Constructs.

To measure the regulation of 93 TFs at 3 binding positions (+1, -57, and -103) and 2 binding sequences, we used the library of titratable transcription factors developed by (Parisutham et al., 2022). Specifically, these TFs had their endogenous coding regions deleted from the genome and the TF coding sequence replaced in the *ybcn* locus fused to an mCherry cassette under the control of the pTet promoter. The TetR protein was integrated in the *gsp1* locus under regulation of a constitutive promoter. The concentration of each of the 93 TFs could be manipulated *in vivo* by addition of aTC

(anhydrous tetracycline) to the growth medium – titrating the number of active TetR repressors in cells.

To clone the 186 TFBS at the three binding positions, we took the TFBS oligos with the Type IIs restriction overhangs and annealed them to their reverse complements in a predefined arrangement in a 96-well PCR plate. The overhangs were designed for digestion by the BbsI restriction enzyme with the following sequence '5'-GAAGAC-3" and a 4 bp sequence at the 3'end to facilitate cloning into the digested position HT plasmid (see Methods 3.3.1, Table 6.1)**.** The arrangement of the oligos in the plate corresponded to the layout of the TF-titration strains: these strains were organized into two 96-well growth plates separating TFs with comparable (or higher) growth rates than those with lower growth rates with strains duplicated in the plate layout as there were 2 binding sequences per TF (a total of 186 wells across 2 96 well plates were used). Cloning of the 186 TFBS at a given position was done in a single digestion and ligation reaction by incubation with BbsI and T7 ligase with the addition of isolated position construct plasmids. Temperature oscillations between 37° and 16° at 5-minute intervals were carried out for optimal digestion and ligation and proceeded for 30 cycles. The reaction mixture was then transformed into chemically competent cells, cultured in SOC media for 1 hour, and plated on agar plates. PCR sequencing was done to confirm the incorporation of the TFBS oligo with select wells from the 96 well plates sent for Sanger sequencing to confirm incorporation of the oligo at the defined position. Final transformation of the cloned TF binding sequence and position plasmids into the respective TF-titration strain was accomplished by generating competent cells for each of

the 93 TFs from the library and cloning the isolated plasmids into the strains. Strains were then prepped for measurement as described in Methods 2.3.1.

Chapter IV: Results

Promoter Titration Uncovers the Mechanism of Regulation for a Diverse Pool of *E.coli* Regulators

4.1 - Introduction

In prokaryotic transcription, one of the chief regulators of expression is the promoter sequence involved in recruiting RNAP to the region. This recruitment is mediated through contact between the subunits of the RNAP holoenzyme such as the sigma subunit and the alpha-C terminal domain tail by sequences such as the Up element along with -35 and -10 consensus sequences (Borukhov & Lee, 2005; X. Y. Li & McClure, 1998; Murakami et al., 2002). Sequence affects both DNA binding and unwinding of initially transcribed DNA, appreciably altering the initial transcriptional kinetics and the time for RNAP to clear the promoter(Goldman et al., 2009; Hsu et al., 2006; Revyakin et al., 2006). The interplay between TF regulation and the kinetic steps of RNAP binding and promoter clearance, therefore, is important in understanding the context by which TFs exert control over the transcriptional process and regulate endogenous promoters.

Early considerations of basal promoter strength were primarily viewed in the context of repression, particularly with the model of kinetic competition proposed in (Lutz & Bujard, 1997). Subsequent work on prokaryotic TFs explored the mechanistic interactions of regulators on the stability of initial RNAP-DNA complexes, with the

observations that TFs that regulate these intermediates of the transcription process can have vastly different regulatory outcomes depending on the initial stability of these complexes (Djordjevic & Bundschuh, 2008; Jensen et al., 2019). Investigations of the phage psi 29 repressor protein p4 demonstrated highly specific regulation at keys steps of the transcription process across different promoters. In promoters associated with early expression of the phage genome, the p4 protein represses transcription, while activating transcription at the late promoter ((Monsalve et al., 1997),(Rojo, Mencía, Monsalve, & Salas, 1998)). Binding of the p4 protein occurs cooperatively with RNAP, stabilizing the RNAP-DNA open complex and prevents RNAP from progressing to productive elongation (Monsalve, Mencia, Rojo, & Salas, 1996). In effect, the mechanism of this repression occurs in the post-recruitment step with a significant stabilization component ($\beta > 1$ in the Thermodynamic Model) and bears resemblance to other TFs in *E.coli* that employ a similar mode of repression (Choy et al., 1997),(Schröder & Wagner, 2000). The function of these repressors stands in direct contrast to repressor action in the model of kinetic competition and suggests that the net regulatory outcome (activation or repression) is contingent on the relative strength of the promoter. Promoters with sequence determinants that promote RNAP-DNA open complex stability would not be activated by these TFs but rather repressed due to its role in inhibiting promoter escape (Galburt, 2018).

The predominance of these modes of regulation offers testable hypotheses on how regulators with distinct stabilization and acceleration profiles operate on different promoter types. An important question we could address is if the method we developed to

infer the degree of stabilization from *in vivo* data (the concentration manifold approach) in Section 2.2.2 does indeed inform us about fundamental aspects of TF regulation and allow us to predict the regulation of characterized TFs on distinct promoter types. Specifically, we sought to test predictions on how TFs with different propensities to stabilize regulate promoters of different strengths. For activators and repressors that function through stabilization of RNAP-DNA intermediates, the thermodynamic model generates specific regulatory outcomes for promoters of different strengths. In essence we could determine if what we learn from the manifold (inference of stabilization) could indeed be predictive as opposed to simply classifying TF function.

An important conclusion from the investigation of sequence and position on TF regulatory modes in Chapter 2, was that for select TFs, the position feature seemed to play a dominant role over sequence in setting the TF regulatory modes. Building on this information, a thorough dissection of the role of position on the regulatory modes for certain TFs from the 93 profiled regulators would allow us to uncover positions of (de)stabilization for distinct TFs. This point allows us to make informed predictions on how these regulators should work on different promoters to chart the role of position, TF identity, and promoter strength on the regulatory modes of TFs.

4.2 - Results

4.2.1 Quantifying Proximal Promoter Regulation of 23 *E.coli* TFs

We selected 23 of the 93 TFs for systematic interrogation of the position axis akin to our work on CpxR. The proportion of TFs belonging to each classified Family are shown in

Figure 4.1B, with the TFs chosen based on the dynamic response of the regulation seen in the systematic interrogation of TF position and sequence. Many of these regulators are listed as affecting both activation and repression on distinct *E.coli* promoters (dual regulators) (Keseler et al., 2021), with interesting implications for the regulatory modes employed by these TFs. We focused on proximal regulation by these factors, as profiling the 93 TFs in the prior section showed only a handful of TFs that regulate at the distant binding location of -103. We selected the binding sequences that provided robust Fold-Changes from the 93 TF profiling and cloned a single binding sequence for each TF at 11 proximal positions. These positions range from -59bps relative to the transcription start site (the 3'end of the binding sequence) to +5bps downstream (5'end of the binding sequence) (Figure 4.1A).

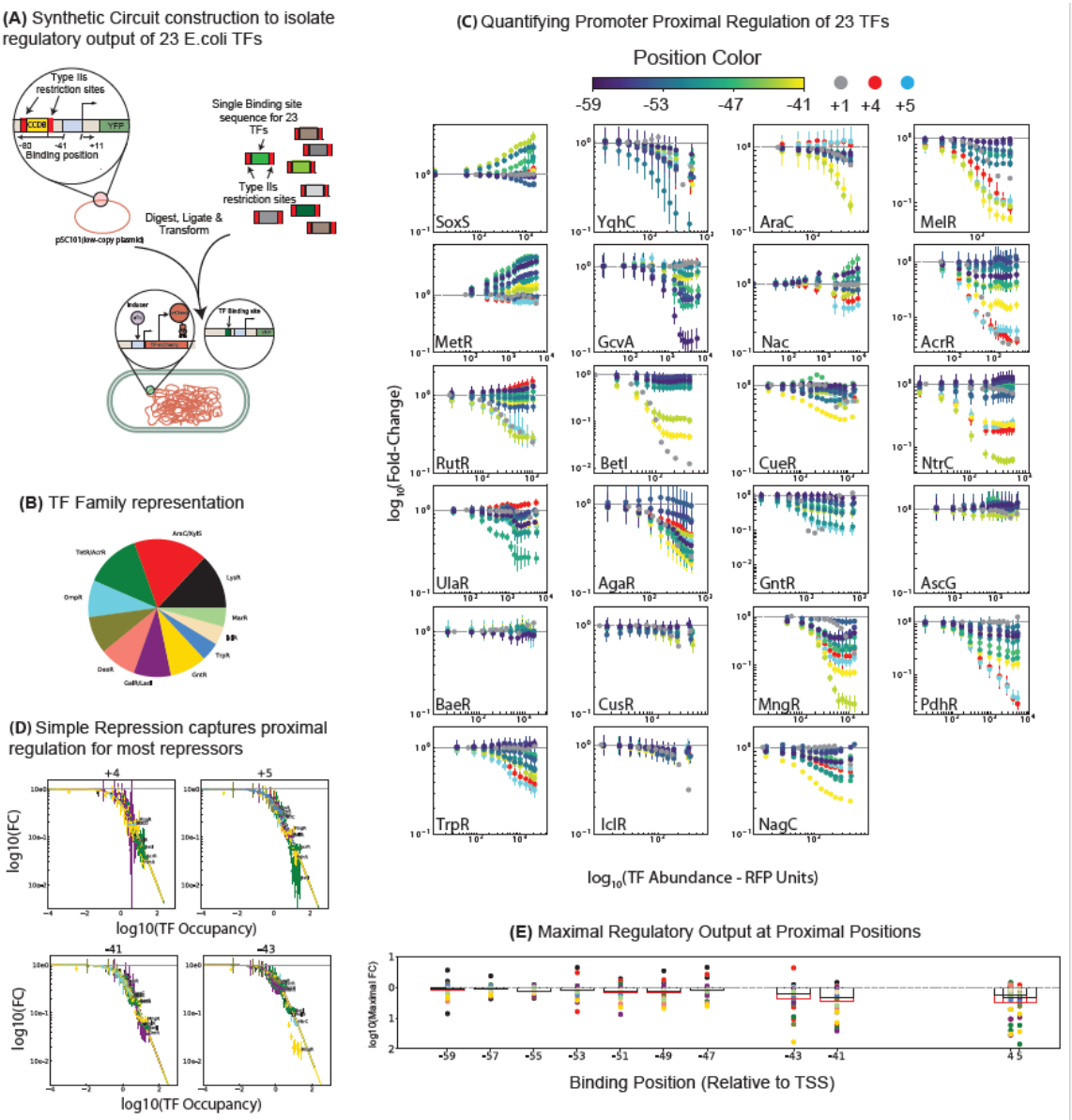

(A) Synthetic Circuit construction to isolate regulatory output of 23 E.coli TFs
Type IIs restriction sites
CCDB
YFP
-80
-41
+11
Binding position
Single Binding site sequence for 23 TFs
Type IIs restriction sites
Digest, Ligate & Transform
p5C101(low-copy plasmid)
Inducer
TF Binding site
(B) TF Family representation
(C) Quantifying Promoter Proximal Regulation of 23 TFs
Position Color
-59
-53
-47
-41
+1
+4
+5
log10(Fold-Change)
SoxS
YqhC
AraC
MelR
MetR
GcvA
Nac
AcrR
RutR
BetI
CueR
NtrC
UlaR
AgaR
GntR
AscG
BaeR
CusR
MngR
PdhR
TrpR
IclR
NagC
log10(TF Abundance - RFP Units)
(D) Simple Repression captures proximal regulation for most repressors
+4
+5
log10(FC)
log10(TF Occupancy)
-41
-43
log10(FC)
log10(TF Occupancy)
(E) Maximal Regulatory Output at Proximal Positions
log10(Maximal FC)
-59
-57
-55
-53
-51
-49
-47
-43
-41
4 5
Binding Position (Relative to TSS)

Figure 4.1: Characterizing the Regulatory Arsenal of *E.coli* Across Position Reveals Rules for Proximal Promoter Regulation (A) Schematic detailing the synthetic circuits used to measure several TFs at 11 binding positions (ranging from -59 bps from the TSS to +5). **(B)** The breakdown of TFs into the classified families. **(C)** The regulatory data for 23 E.coli TFs at multiple binding position measured by flow cytometry. Each data point represents the median fold-change of the binned data for that position normalized to the fold change in response to the changing growth rate of the cell (measured by the change in YFP expression from TFs regulating a circuit with no binding site). The error-bars represent error propagation of the uncertainty of this normalized fold-change. The color scheme for position is split with the downstream positions separated from the upstream ones. **(D)** TFs that repressed gene expression were fitted to the thermodynamic model to get the TF occupancy. Scaling the TF abundance measurements with this parameter allowed us to gauge how many repressors regulated by simple repression. As seen, the data clearly shows that most of the repressors collapse on the model curve.**(E)** The maximal regulatory effect (mean of the last three points from the regulatory curve) as a function of position demonstrates the landscape of activation and repression. TFs are colored according to their Family as depicted in (**B**). Note the prevalence of strong repression at positions adjacent to the promoter (-35 and -10 sequences not shown).

In Figure 4.1C, we show the result of these measurements: most TFs demonstrate repression across all proximal positions except for 3 activators: SoxS, Nac, and MetR. This activation ranges from 2-Fold (Nac at position -49) to roughly 5-fold activation (MetR at positions -47 and -53). Repression, however, tends to have stronger dynamic range, with the TF MngR from the *GntR* family reducing expression by almost 100-fold. The strongest repression is found at positions closest to RNAP binding, encompassing TFs from the *GntR, AraC/XylS,* and *TetR/AcrR* families. However, there is a clear dependence on the orientation of the binding position upstream or downstream relative to

the promoter, with certain TFs such as GcvA, BetI and MngR regulating strongest at the proximal upstream position, with little to moderate repression in the proximal downstream positions. The TFs Nac, PdhR, and AcrR, function in an opposite manner and have the strongest repression in the proximal downstream position (Figure 4.1C).

We then asked if the repressors at these proximal positions regulate in a way consistent with simple repression – a model of gene regulation demonstrated by well-studied regulators such as the LacI TF (Brewster et al., 2014). In this regime, the occupancy of the TF solely determines the regulatory response, with promoter regulation set by the TFs binding affinity to the site and the degree of stabilization imparted to RNAP by the regulator. We fit the thermodynamic model of gene regulation to the repression data at the 4 positions closest to the promoter (+4, +5, -41, -43) and plot the data scaled by TF occupancy in Figure 4.1D with the simple repression curve as the solid yellow line. We see good agreement with the data uniformly across these positions, however, the -43 region has more TFs that fail to conform to the model. The TF MngR clearly tapers at this position along with the TFs MelR and MngR at -41. This indicates that these repressors clearly require a different model to explain their function at these binding positions with the potential of different regimes of stabilization featuring in the regulation of these TFs.

In Figure 4.1E, we show the landscape of position regulation, where a clear trend can be seen for the repression data. Positions immediate to the -35 and -10 sequence demonstrate the predominance of strong repression with multiple TFs and a gradual tapering of the response for most TFs further away. The activating responses tend to be seen after the initial proximal positions of -41 and -43, with the exception for SoxS at -43 (5-fold

activation), with peak activations for the LysR family (MetR and Nac) further from the promoter including positions -49, -53, and the terminal position -59. Taken together our quantitative regulatory data spanning proximal binding locations of various TFs points to a variety of regulatory behaviors, with implications for differential utilization of stabilization and acceleration by these TFs.

4.2.2 Concentration Manifold Analysis Uncovers Stabilization Across Diverse Regulators

To determine the degree of stabilization present for each of these TFs we used the manifold approach applied to CpxR in Chapter 2. However, a key assumption used to isolate stabilization from the position curves that the TF operates with complete destabilization ($\beta \sim 0$) at a downstream reference position, was not valid for several TFs as the regulatory response at the downstream positions (+4 and +5) were muted (see the response function for the TFs GcvA, CueR, AraC, GntR – Figure 4.1C, Figure 4.1E). We therefore resorted to a qualitative approach to assess the predominance of stabilization at different binding positions. By plotting the concentration manifold between two positions (Figure 4.2A), we detect an increase in the stabilizing coefficient of one position relative to the other if the curve deviates from a straight line which implies either no stabilization or comparative stabilization between positions. We term this coefficient Kappa (K). It is formally the ratio of TF occupancy due to its stabilizing property with RNAP between two positions. An example of the workflow is shown in (Figure 4.2B), where we plot the pairwise comparison across 4 binding positions for a given regulator to extract positions with significant curvature in data.

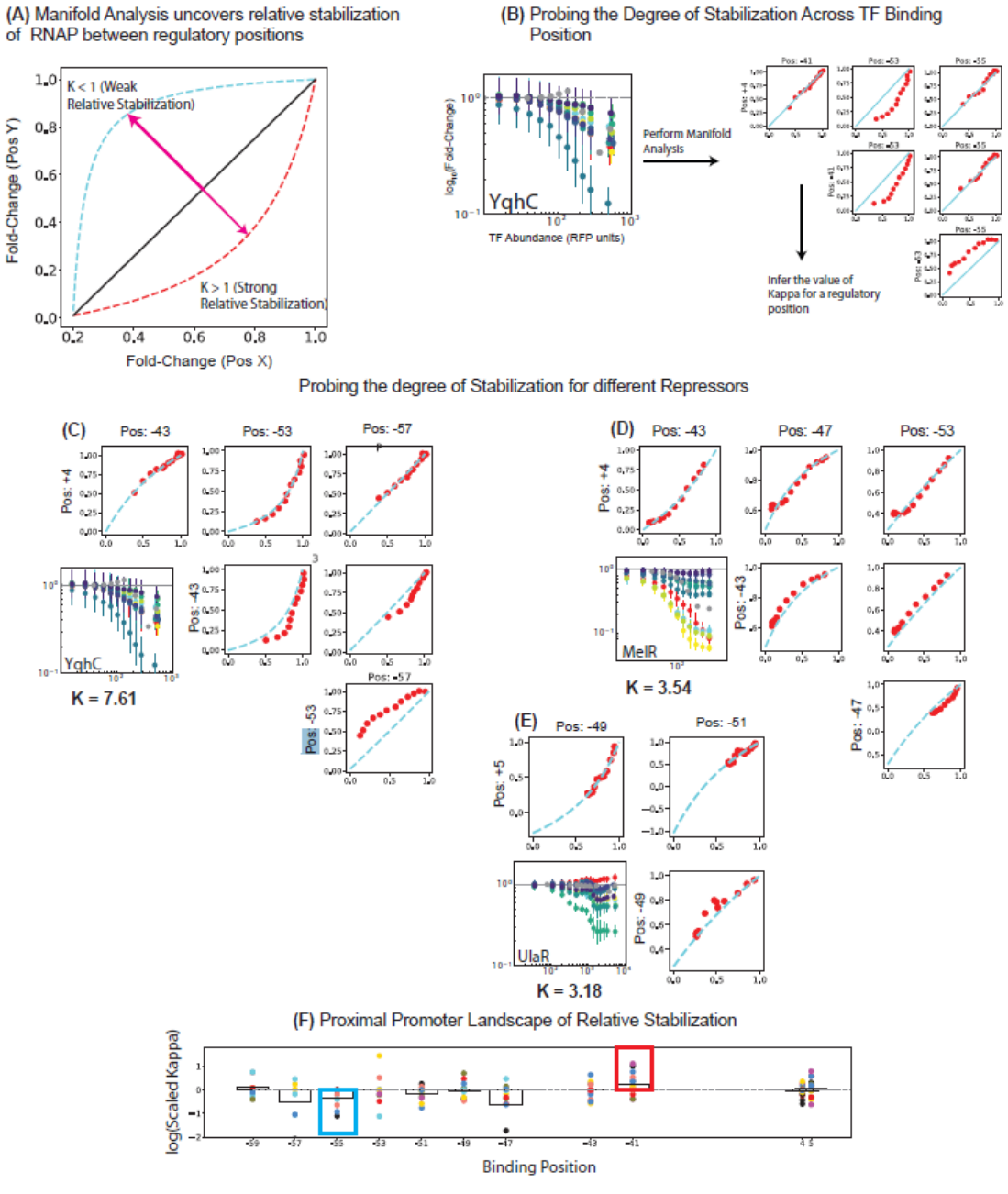

(A) Manifold Analysis uncovers relative stabilization of RNAP between regulatory positions
K < 1 (Weak Relative Stabilization)
K > 1 (Strong Relative Stabilization)
Fold-Change (Pos Y)
Fold-Change (Pos X)
(B) Probing the Degree of Stabilization Across TF Binding Position
YqhC
TF Abundance (RFP units)
Perform Manifold Analysis
Infer the value of Kappa for a regulatory position
Probing the degree of Stabilization for different Repressors
(C)
Pos: -43
Pos: -53
Pos: -57
YqhC
K = 7.61
(D)
Pos: -43
Pos: -47
Pos: -53
MelR
K = 3.54
(E)
Pos: -49
Pos: -51
UlaR
K = 3.18
(F) Proximal Promoter Landscape of Relative Stabilization
log(Scaled Kappa)
Binding Position

Figure 4.2: Probing the Degree of Relative Stabilization Across Binding Position for Several TFs. (A) Plotting the Fold-Change in one regulatory position as a function of another provides insight into relative stabilization (K) between binding positions. **(B)** Workflow to extract the degree of stabilization between TFs. We selected positons that showed robust regulation (more than 2-fold for activation and at least a third of repression) for pairwise analysis, inferring the N-1 Kappa values from the N positions that met this criterion. The pairwise manifold plots have the dashed cyan line to denote the null model (Kappa = 1). **(C,D,E)** Manifold analysis for 3 repressors are shown. Note the degree of curvature seen in the data at certain regulatory positions. For the YqhC TF, position -53 demonstrates the strongest K value. For MelR and UlaR, the positions are -43 and -49 respectively. Positions with stronger stabilization relative to others are highlighted in the red box for the YqhC TF. The dashed cyan line demonstrates the model with the best fitted Kappa Values. **(F)** To facilitate comparison of relative stabilization across position, we scaled the Kappa value for each TF with its median across position to give insight into hotspots of proximal positions that have high relative stabilization. The colors represent TFs according to their classified TF family. Note several TFs wth strong relative stabilization at +5. We note the large number of TFs at -41 with strong curvature providing evidence that TF regulation from this position utilizes a strong stabilizing

In Figure 4.2C, we demonstrate the inference of Kappa (K) for 3 repressing TFs. For the YqhC TF, position -53 had the strongest curvature when assayed pairwise against all the positions with appreciable regulation. We see the curvature clearly when fold-change at position -53 (Y-axis) is plotted against the fold-change at positions +4 and -43. The data demonstrates downward curvature implying regulation at that position reaches its maximum faster than its counterpart (i.e higher degree of stabilization). The inferred Kappa value was approximately 4 times the reference position (+4 for this TF). We see the same pattern for the TFs UlaR and MelR, with Kappa > 1 at positions -49 and -43 respectively.

As stabilization is featured prominently in at least one of the regulated positions for these 3 repressors, we sought to uncover the positional landscape of relative stabilization across the 23 TFs we profiled. In Figure 4.2D, we plot the Kappa values for each TF scaled to its respective median across positions that it regulates. This scaling allows us to visualize positions of strong curvature in the manifold response by uncovering the spatial landscape of curvature in our dataset. At position -53, we see the largest of this response for a single TF, which corresponds to YqhC. At -55, however, we see all the values fall below the zero line, indicating that relative stabilization is uniformly lower across all TFs. A striking feature from the manifold analysis is that the +5 regulatory response seems to have higher relative stabilization than the median position for several TFs, suggesting that downstream repression is not exclusively mediated through destabilization. Considering the upstream proximal positions (-41 and -43), we see the scaled Kappa tends toward values higher than the median and is consistent with the paradigm of Class I activation where the TF makes stabilizing contacts with RNAP. The interesting aspect of our data is that most of these positions repress (Figure 4.2D) in good agreement with the simple repression model. ***This feature of stabilization across different TFs hints that despite opposing regulatory outcomes, proximal regulation by model activators and the profiled repressors involves a stabilizing component of RNAP – pointing to a conserved mechanism of TF modulation at these positions***. However, while it is important to note that we have uncovered evidence of stabilization by several repressors, the manifold approach we designed only allows us to get the relative differences – with the assumption that the binding affinity of the TF does not change between positions.

4.2.3. Promoter Titration Confirms the Role of (de)Stabilization Among Activators and Repressors of Gene Expression

Given the evidence of stabilization by distinct TFs, we probed the role of promoter strength on the regulation of these putative stabilizers. As we have quantitatively dissected the influence of the position axis in setting gene regulatory outcomes, promoter strength is equally important and its intersection with the TFs ability to stabilize have clear regulatory implications in the thermodynamic model. For strong promoters, the degree of stabilization plays no role in setting the maximal response: this is set solely by the regulators ability to modulate the initiation rate (acceleration). This interplay between the regulatory mode and promoter strength is diagrammed in Figure 4.3B. In the phase plot of stabilization (β) and acceleration (α), the effect of promoter strength can be divided into 6 regulatory regimes. For TFs that regulate by positively modulating stabilization and acceleration, the response is activation despite changing promoter strength (insert plot). For TFs that negatively alter both modes, the response is repression, with the degree set by the promoter strength. In the incoherent regimes, where the TF regulates one mode positively and the other negatively, the expression output is nuanced. In 2 of these regimes the TFs can cross from repression to activation depending on promoter strength. This is important, as certain repressors from the manifold analysis

(See Figure 4.3) have implied regulation by stabilization, setting these factors in the incoherent regime with the potential to realize dynamic quantitative outputs based on promoter strength.

To test the role of promoter strength on the TFs we profiled, we used the synthetic gene circuits designed for high throughput cloning in the previous Chapters and cloned different -35 and -10 sequences (Table 6.2) into these constructs for a total of 7 promoters with good dynamic range, spanning almost 4 orders of magnitude in expression (Figure 4.3A). An important assumption is the fact that these mutations in the -35 and -10 solely change the RNAP-DNA binding kinetics, and not the promoter escape rate, which is viable given the extensive characterization of sigma70 promoters (Wade et al., 2006). As we determined that the CpxR TF regulated by stabilization in Chapter 1, we tested the regulation of CpxR at position -64 (a stabilizing position) at these different promoters. The results are consistent with the phenomenon of stabilization: the stronger promoters have muted activating responses, but the weaker promoters having strong activation (Figure 4.3 C,E) – the weakest having an approximately 80-fold increase in gene expression. Plotting the fitted FC_{max} from each promoter response as a function of promoter strength, we see the curve decrease from smaller to larger promoter strength values as depicted in region 1 and 2 of Figure 4.3B. This same phenomenon is seen for SoxS regulation at -53, where strong activation is seen at the weakest promoters with the degree of activation following the rank order of promoter strength for all but one promoter (the purple coded promoter – Figure 4.3 D,F). These findings highlight that the inference of stabilization from the manifold is predictive: our application of an inferred

stabilizing activator (CpxR at -64) to novel engineered promoters resulted in a response consistent with the thermodynamic model.

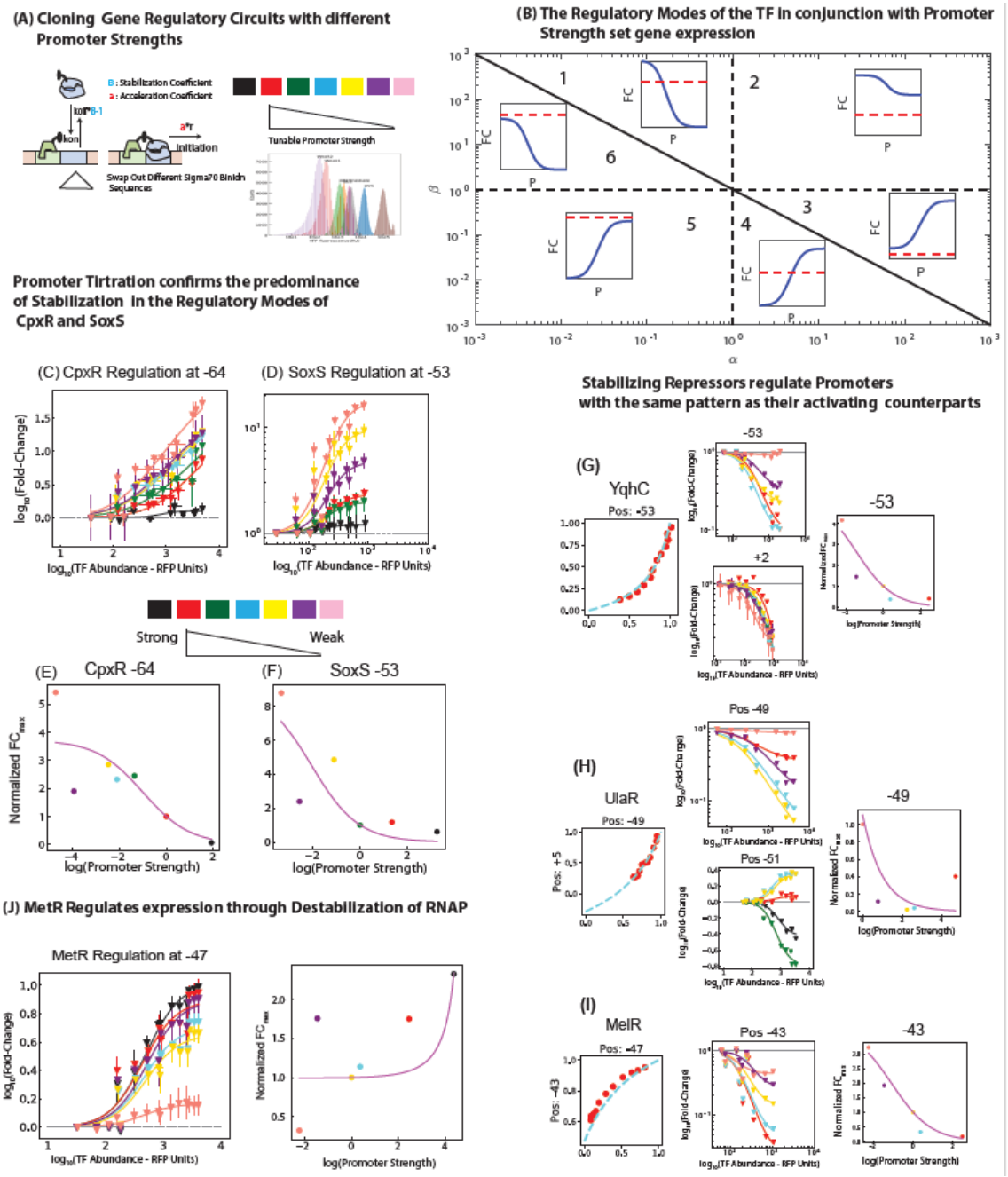


Figure 4.3: The Promoter Titration Effect Uncovers the Degree of Stabilization Across Repressing and Activating TFs (A) To determine if inference of stabilization from the manifold approach could be used to predict expression of novel gene circuits, synthetic promoters were designed with a range of basal promoter expression over 5 orders of magnitude. These promoters were based of the LacUV5 promoter sequence with targeted mutations in the -35 and -10 regions. **(B)** Diagram showing the Fold-Change as a function of promoter strength based on the TF's location in the phase space of regulatory Modes. The incoherent regimes (1,4) demonstrate the potential to go from activation to repression as promoter strength changes. **(C)** CpxR regulation at the 7 promoters of varying strength confirms stabilization at -64, confirming the inference in Chapter 1 and demonstrates the predictive nature of the manifold. (**F**) Stabilization is evident at this location when plotting the FC_{max} as a function of promoter strength with stronger promoters having weaker values. **(D)** SoxS regulation at -53 also follows the stabilization model except for the second weakest (purple) promoter (**F**). (**G,H,I**) Stabilizing Repression is confirmed for TFs from the manifold analysis. **(I)** MelR at -47 follows the rank order perfectly. (**G,H**) For UlaR, regulation at -49 follows the trend of stabilization along with YqhC at -53 For UlaR, regulation at -51 demonstrates a transition from strong to weak promoters as seen in the FC_{max} vs Promoter Strength plot, although the rank order of promoter strength does not match one to one with the expected stabilization response. (**J**) Activation through destabilization of RNA is seen for MetR regulation at position -47, demonstrating a non-canonical form of activation. The promoter titration data resembles regimes 3 and 4 from the from **(B)**.

As promoter titration confirmed the degree of stabilization for these two activators, we next profiled the 3 repressors from the manifold analysis that suggested the presence of stabilization in these regulators. Indeed, we see the same pattern of regulation as a function of promoter strength for these repressors as the stabilizing activators (Figure 4.3 G-I). For the YqhC TF, position -53 shows the same pattern as the activating TFs, with the strongest promoters with the lowest fold-changes (greatest repression) and the weaker promoter having attenuated repression (Figure 4.3G). This observation confirms the

presence of stabilization for this TF, as the weaker promoter is compensated by the TFs stabilizing property, resulting in muted repression. To confirm the promoter titration response was not due to some property independent of position, we cloned the YqhC binding sequence downstream at +2 (see Figure 4.3G). At this position, we find that the response of the TF on different promoters confirms no separation in the fold-change with different promoters, highlighting that the stabilization is position specific (as seen in the manifold analysis). This stabilization is also confirmed for MelR regulation -43 and UlaR -49, where the manifold approach informed us of strong relative stabilization (Figure 4.3 H and I). For UlaR, profiling the TF's response on the ranked promoters 2 bps away (-51) reveals the phenomenon seen in regime 4 (Figure 4.3H), where the stabilization is apparent in weaker promoters and leads to 2-fold activation - in contrast to repression for stronger promoters. Overall, we see that the promoter titration approach follows the expected response for a stabilizing repressor. The implication of this incoherent regulation is apparent***: to realize the attenuated repression at weak promoters, the TFs ability to stabilize RNAP tempers its effect in slowing it through initiation, effectively compensating for the delayed kinetics of promoter escape by increasing the effective concentration of RNAP at the promoter.***

The thermodynamic model informs us of the possibility of regulation where the net regulatory outcome is activation, but the TF destabilizes RNAP at the promoter. This is featured in the inset plot (regime 3), clearly demonstrating that a seemingly counter-intuitive notion of regulation is possible. The TF MetR at -47 regulates in this fashion with the promoter titration response shown in Figure 4.3J. This activation operates in a

distinct regulatory regime than the stabilizing activators CpxR and SoxS: the strongest promoters have the larger activation response, followed by the weaker promoters, with the ranked weakest promoter (salmon data) demonstrating barely any response to MetR. The activation is realized by the large relative effect on the kinetics of promoter escape, which allows for the TF to increase gene expression despite impairment of RNAP recruitment to the promoter (Figure 4.3J). This find is important, as it demonstrates TF activation outside the purview of the canonical activation of Class I and Class II, where stabilization of RNAP plays a central role. Instead, the role of MetR shares similar characteristics to the regime of repression by kinetic competition (Lanzer & Bujard, 1988), as increasing promoter strength "crowds out" the regulator, leading to lower concentration of bound TF at the promoter. This point demonstrates that understanding the mode of regulation is far more meaningful than the conventional labeling of activators or repressors of gene expression, providing a basis for rational appraisal of TF regulatory function and promoter activity from a first principles approach.

4.3 - Methods

4.3.1. Designing Promoter Sequences to Titrate Basal Promoter Expression.

To probe the role of stabilization in gene expression, we designed promoters with varying degrees of basal expression spanning several orders of magnitude. The promoter used in our throughput gene circuits was a variant of the *LacUV5* promoter, the natural promoter that regulates the *lac* operon. Based on prior work testing different variants of this promoter (Brewster, Jones, & Phillips, 2012), we used 4 variants from this work, with

two additional promoters designed from considering the sequence energy matrix of the Sigma70 promoter (LaFleur, Hossain, & Salis, 2022). These mutations encompassed bases in both the- -35 and -10 sequences: we made no changes to the spacer nucleotides as well as the sequences between the -41 to -35 and the -10 to +1 regions. The sequences are listed in Table 6.2.

4.3.2. Derivation and Fitting of the Manifold Model

To uncover the role of stabilization in the regulatory response of the 23 TFs, we made use of the manifold model that was applied to the CpxR position sweep data in Chapter 1 to uncover the regulatory modes. The derivation in Methods 2.3.2.1 holds, however, instead of assuming that the reference position operated through destabilization ($\beta \sim 0$), we allowed for the possibility of no or appreciable stabilization at the reference position. Specifically, the goal was to discover the ratio of the TF occupancy by stabilization between the two positions, assessing the relative use of stabilization between both positions. The problem of inferring this ratio from the position regulatory data reduces to fitting N-1 Kappa (K) values for a dataset of N positions. Rather than use a single reference position, we leveraged the full structure of the position regulation dataset for a given TF to fit the N-1 Kappa values from the $\binom{N}{2}$ available regulatory positions. (See Figure 4.4). The result was a robust fitting procedure, with information integrated across multiple datasets to determine the optimal value of Kappa. The values from this fitting procedure were reported in Figure 4.2 for the TFs YqhC, MelR, and UlaR.

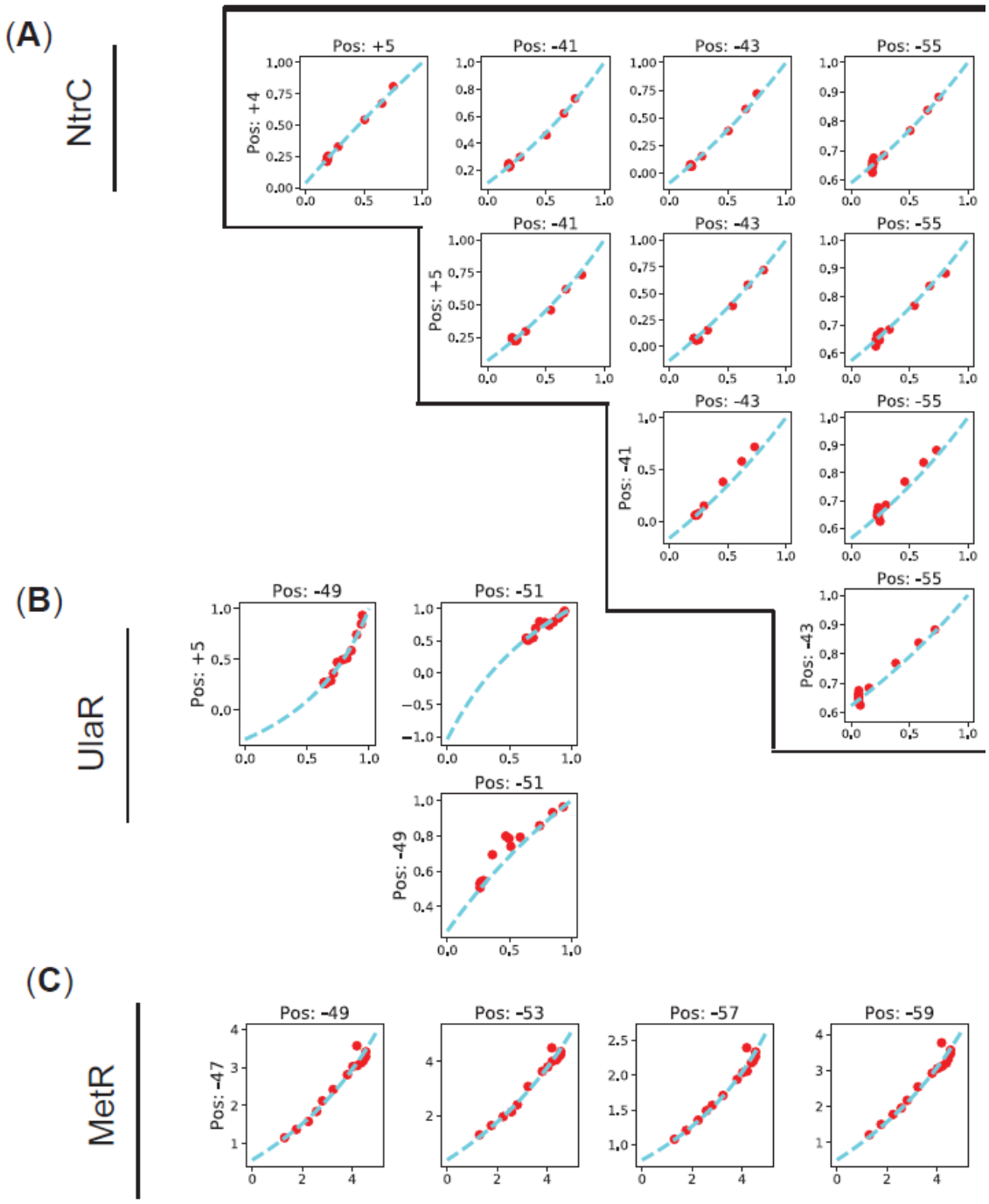

(A)
NtrC
Pos: +5
Pos: -41
Pos: -43
Pos: -55
Pos: +4
Pos: -41
Pos: -43
Pos: -55
Pos: +5
Pos: -43
Pos: -55
Pos: -41
Pos: -55
Pos: -43
(B)
UlaR
Pos: -49
Pos: -51
Pos: +5
Pos: -51
Pos: -49
(C)
MetR
Pos: -49
Pos: -53
Pos: -57
Pos: -59
Pos: -47

Figure 4.4. Assessment of Stabilization for Activating and Repressing TFs Using the Concentration Manifold Approach for Select TFs. Examples of the pairwise plots used in the manifold approach for 3 out of the 23 TFs assayed. Binding positions for a given TF were only considered if the fold-change in expression was robust (2-fold activation and repression). **(A)** For the NtrC TF, the 4 positions that met this threshold (+5, -41, -43, -55) were binned according to their mCherry expression, and fold-change values for one of the pairwise positions were plotted as a function of the other. The Y-label is the reference position (values plotted on the X-axis) and the title label denotes the response TF (values plotted on the Y-axis). The dashed cyan line represents the model fit for the Kappa (K) value as detailed in this section. For this TF, we see that the profiles are straight lines for almost all pairwise positions – indicating no curvature, and therefore no apparent stabilization in the regulatory response. **(B)** For the UlaR TF, we find strong curvature at position -49 indicating stabilization in the regulatory response. This was confirmed by promoter titration experiments (See Results Section 4.2.3). **(C)** Despite activation by the MetR TF, analysis of pairwise curves indicates near straight lines (K ~ 1) for regulating position, indicating that stabilization does not play a prominent role in setting this activation. This was confirmed by promoter titration, which revealed a destabilizing effect in regulation of RNAP.

Chapter V - Discussion

5.1 - Implications of Coherent and Incoherent Regulatory Regimes in CpxR Regulation

In this thesis, we characterize the influence of TF binding sequence, position, and promoter strength on the regulatory activity of TFs at the promoter. Using a synthetic biology approach, we circumvent expression of TFs from their native genetic context by using a library of *E.coli* strains that expresses TFs under the control of a synthetic promoter. Crucially, this allows us to induce expression of TFs that would only be possible under certain cell states using standard cell culture procedure and growth media, providing tight control on growth conditions. Additionally, it allows us to precisely tune the concentration for the majority of *E.coli* TFs *in vivo* and isolate the fundamental regulatory effect of the TF on gene expression. Using rationally designed gene constructs that allows us to alter the basal promoter strength, binding location and sequence composition of TF operator sites, we probe the elements of cis-regulatory architecture on the TFs ability to modulate two steps of the transcription process: (de)stabilization of RNAP at the promoter, and (de)acceleration of the rate of transcription from bound RNAP – collectively referred to as the regulatory modes.

In Chapter 1, we uncover the modes of regulation for the CpxR TF through systematic dissection of the regulatory response across multiple binding position. Despite CpxR binding to a slew of endogenous promoters in *E.coli*, its regulatory mechanism is unclear with annotated regulatory functions of activation and repression from several promoters

(Raivio & Silhavy, 1997; Salgado et al., 2023) . We profiled the regulatory activity of this TF across several binding positions and used a thermodynamic model of gene regulation to decouple TF occupancy from its maximal regulatory output – allowing us to delineate the TFs mechanistic role at the promoter in terms of its propensity to stabilize RNAP and/or accelerate initiation. We find that CpxR demonstrates strong propensity to stabilize: the stabilization coefficient inferred from the regulation at 12 positions was significant. An important observation was that the degree of stabilization of RNAP did not correlate with the magnitude of activation: the strongest activation position and the sole repressive position had similar stabilization coefficients. However, the difference between these positions arises primarily from differential effects on the initiation rate between positions: the strongest activation position had an initiation rate almost 12-fold higher while the repression position showed deceleration in transcription initiation by almost 20-fold. Taken together, we find that CpxR engages in two broad regulatory regimes where the two regulatory modes work in concert for strongly activation positions – a phenomenon we term coherent regulation - and where the modes work in opposing fashion for weaker activation positions - a phenomenon we term incoherent regulation.

5.2 - CpxR in the Context of Class I and Class II Activation

The position regulation of CpxR can be viewed in the context of well-developed models of activation. Regulation by stabilization is a predominant feature of the CRP TF, one of the most well studied positive regulators in *E.coli* (Hochschild & Dove, 1998; C. L. Lawson et al., 2004; Schultz et al., 1991). CRP regulation at different promoters involves a proximal and distal binding positions: both of which utilize stabilization of RNAP as

part of the regulatory response (Farrell et al., 1996). Both positions are situated in proximity to the Sigma70 promoter and allow the CRP TF to facilitate protein-protein contacts with the alpha-CTD of RNAP(Borukhov & Lee, 2005). These contacts are mediated by specific acidic patch domains on the CRP molecule: The AR1 patch makes specific contact with the alpha-CTD of RNAP and promotes stabilization of RNAP while the AR2 region contacts the alpha-NTD of RNAP and allows for efficient isomerization of the closed RNAP-DNA complex(Estrem et al., 1999; Lewis & Adhya, 2015). The activation regions engage in non-overlapping regulation, altering the rates of specific kinetic steps of the transcription process. While no structural studies exist to confirm the mechanism of CpxR stabilization and acceleration, scanning mutagenesis of the CpxR protein sequence that result in abrogation of CpxR activation could be used to isolate regions involved in stabilization and acceleration of RNAP. Crucially, as the acceleration seen in the CpxR data was position specific, critical residues involved in modulation of post-recruitment steps would not influence the positions where stabilization solely sets the activation response – thereby isolating residues involved in modulation of the two steps of the transcription process.

5.3 - Pitfalls of the Inference Procedure Used in Uncovering the Regulatory Modes of the CpxR TF

A key concept we introduced to infer the regulatory modes of the CpxR TF was the concentration manifold approach that compared the Fold-Change at a given concentration at one position to the same measure at another position. This formalism allows for the inference of stabilization with two important assumptions: The reference position

operated via destabilization ($\beta \sim 0$) in realizing its regulatory function and the binding affinity did not change between positions. The first assumption is predicated on a wealth of annotated data suggesting TF binding in proximal downstream positions is predominantly repressive (See Figure 3.1): proximal binding downstream to the promoter sequence could be considered to influence gene expression via steric hindrance with RNAP (Lanzer & Bujard, 1988; Lutz & Bujard, 1997). However, we find that the assumption is not always justified, as in the case of MraZ regulation at +1 in Chapter 2, which records activation across both binding sequences. The influence of position on binding affinity is nuanced, and we make recourse to the observation that prokaryotic TF binding sequences tend to have large information footprints, with sequences considerably larger than their eukaryotic counterparts (Jolma et al., 2015; Morgunova & Taipale, 2017). This suggests that sequence of the surrounding nucleotides should not affect the primary determinants of TF-DNA interactions TFs with large operator sequences, suggesting the binding affinity should be relatively stable across different locations.

An important limitation of the approach of the concentration manifold is the inability to infer destabilization: this is evident for CpxR regulation at -74, where the inference determines the FC_{max} (maximal regulatory output) of the thermodynamic model with precision, but fails to localize the stabilization and acceleration coefficients in the phase plot due the lack of curvature in the concentration manifold. The inference cloud spans multiple values of stabilization and acceleration values consistent with the FC_{max} contour line and encompasses both coherent and incoherent regimes. This limitation effectively precludes characterization of multiple *E.coli* regulators that operate through

destabilization and motivates the design of specific gene circuits to address this deficiency as carried out in Section 4.2.3.

5.4 - Final Summary of CpxR Data

The presence of incoherent regulation observed at several binding positions for the CpxR TF highlights 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). This opens questions on how pervasive this feature is across the diverse regulators in *E.coli*. Addressing how different regulators behave according to their mode of regulation would fill a vital gap in current information about TF regulatory function and allow for a more complete characterization of TFs. This knowledge of TF regulatory function in conjunction with methods of profiling TFs through genome-wide occupancy techniques, would provide an edge in the challenge of uncovering an "expression code": i.e., a set of rules that govern the magnitude and duration of gene expression from natural promoters (Gertz, Siggia, & Cohen, 2009). However, understanding these rules requires a dissection of the role of TF-identity, TF binding sequence, position, and the strength of the regulated promoter using methods capable of teasing out the relative contributions of each to the measured expression levels from the regulated promoter.

5.5 – Decoupling the Role of TF Binding Position and Sequence in Setting TF Regulatory Modes

Building on these conclusions, we extended our synthetic approach to characterize 93 *E.coli* TFs evaluating the relative contributions of sequence and binding position to properties of the TF regulatory function. Using the synthetic gene constructs we designed that allowed for integration of TF binding sequences at defined positions relative to the promoter, 186 TF binding sequences at 3 defined positions were measured in the respective TF-library strain as described in the first part of the Results. Our measurements mark the first quantitative regulatory profiles for several of these regulators and demonstrate rules regarding position regulation. We show that most of the regulation at immediately downstream of the promoter (+1), is repression that is consistent with previously reported functions for several of these factors regulating their endogenous promoters.

Armed with the quantitative profiles of TFs regulating distinct sequences and positions, we developed a rigorous statistical approach that allowed us to assess the importance of these promoter features to the maximal regulatory response of the promoter- thereby indirectly testing its effect in setting the regulatory modes of the TF. As Figure 1.2 details, the FC_{max} parameter encapsulates the TFs ability to (de)stabilize RNAP and (de)accelerate the rate of initiation: under certain conditions the maximal regulatory output is simply the product of the two regulatory modes. By contextualizing the thermodynamic model in an inference scheme, we assessed the model evidence for each feature (sequence or position), comparing the evidence for a model where that feature

was dispensable (independent) in setting the regulatory modes. Strikingly, we found that for 16 TFs with robust regulation – defined as having two fold-activation across two binding sequences and at least 2 binding position, a thermodynamic model where position and sequence did not contribute to setting the FC_{max} parameter could explain the TF specific regulatory data well – effectively demonstrating that the maximal regulatory output (and therefore stabilization and acceleration) could be explained independently from these features. This discovery presents evidence that TF identity, rather than the extrinsic features of the TF binding sequence and position, play an outsized role in setting TF regulatory modes: the model suggests, based on the quantitative profiles of regulation across distinct positions and sequences for a given TF, that these features were important in setting TF occupancy – the binding affinity - to capture the data.

However, for TFs where the features were important in setting FC_{max} , one observation is clear: the position axis captures the variability of the data, indicating its relative importance over binding sequence in determining the regulatory modes. This suggests that the position and sequence axes are decoupled in setting the TF regulatory modes - thereby reducing the possible feature space necessary for understanding this aspect of TF regulation. Furthermore, this work highlights the importance of statistical thermodynamic models where the promoter features are grounded in biochemically meaningful parameters that can be tested and rigorously evaluated for contributions to the TF's regulatory response.

5.6 - Summary of Assessment on the Contributions of Position and Sequence to TF Regulatory Modes

This conclusion demonstrates the need for stronger focus on the roles of TF regulatory domains in determining the mechanism of TF response. While recent work looking at systematic characterization of regulatory domains in TFs provides an important step in delineating structural features with specific regulatory function(DelRosso et al., 2023), further investigation is required – particularly in the area identifying conserved motifs associated with regulatory patterns. For certain TFs, systematic interrogation of TF coding regions via targeted mutagenesis approaches have implicated important residues crucial for activating and repressing functions (Lehming et al., 1987; Monsalve et al., 1996; Roy et al., 2004), however, systematic characterization of regulatory domains across a large set of regulators is rare and serves to underscore the dearth of information on structural motifs that might confer particular regulatory properties for the majority of *E.coli* TFs. Importantly such investigation might uncover conserved features in protein structure that act as determinants in the TFs ability to stabilize and accelerate RNAP, bridging the gap between specific regulatory output and sequence features for a host of different TFs.

Our considerations on the role of sequence and position on TF ability to modulate the kinetic steps of transcription have several limitations that preclude us from making more definitive conclusions across a broader selection of the 93 TFs. While we performed the statistical procedure for all the TFs profiled, we limited the assessment of the importance of these features to 23 TFs, that had robust regulation (which we defined as more than 2-

fold activation or repression) across both binding sequences and at least 2 positions. The lack of strong regulation across several of these TFs is worth mentioning, as in total we had 50 TFs that had robust regulation in at least one of the curves. A potential reason for the weak regulation across several TFs is the mCherry fusion used in our strain construction. Despite the presence of a linker sequence buffering the C-terminal domain with the mCherry protein, the fusion could have compromised TF function in either its capacity to bind DNA or exert regulatory control. As the regulatory domains are often poorly characterized in TFs, ruling out these possibilities is challenging and requires direct assays to measure DNA binding of these library strains to their sequences. One way to determine if the mCherry fusion impaired its function is to engineer several of the weakly regulating TFs in the synthetic circuit without the mCherry cassette. While the ability to measure TF abundance at single cell resolution will be lost, profiling the function of select TFs in an unfused library would check if the TFs simply regulated weakly or if the function had indeed been compromised.

5.7 - Implications of Conserved Regulatory Modes Across TFs with Diverse Regulatory Outcomes

As the assessment on the role of position and sequence in setting TF regulatory modes revealed a handful of TFs that implicated position, we measured the regulation of 23 of the TFs across several proximal binding positions to identify locations that demonstrate potential stabilization. We made use of the concentration manifold approach that allowed us to probe the prevalence of stabilization across the multiple binding positions for CpxR. In this case, as some of the TFs did not demonstrate proximal regulation at the +1

position, we did not make the same assumptions about the steric hindrance of downstream positions and instead used the totality of the regulatory data across all regulatory positions to delineate the amount of curvature in the response. Crucially, the presence of noticeable curvature did not allow us to directly determine the actual value of the stabilization coefficient as we did for the CpxR TF, but earmark binding positions where stabilization was present through the curvature in the data via the parameter $Kappa\ (K)$. Profiling the proximal regulation of these TFs in Chapter 3, we uncovered the degree of stabilization in the regulatory response, with several of the stabilizing TFs functioning as repressors. Previous work has uncovered repressors that regulate by stabilizing RNAP-DNA intermediates(Choy et al., 1997; Rojo et al., 1998; Schröder & Wagner, 2000), with the effect of delaying the progression of RNAP to transcription elongation and suggests the possibility that these repressors regulate by similar mechanisms.

A chief aspect to consider in elucidating the regulatory mode of stabilization through the manifold approach is if the information we gleaned could be used to predict TF regulation on novel promoters. Application of the inference in the manifold to predict TF regulatory function on different promoters would be an important step in demonstrating the utility of the thermodynamic model and crucially the case for classifying TFs according to the regulatory modes as outlined in this thesis. Indeed, the presence of both activators (CpxR) and repressors in our date regulating with a stabilizing component from the manifold analysis motivated us to test the role of stabilizing TFs on promoter strength: TFs that regulated by stabilization should show strong dependence of promoter

strength on the maximal regulatory effect (FC_{max}. See Figure 4.3). We see this hypothesis play out for the TFs YqhC, MngR, and MelR at specific binding locations. A crucial feature is the shared pattern of regulation as a function of promoter strength with the stabilizing activators CpxR and SoxS.

The activator MetR, however, displays a striking pattern when regulating promoters of different strength. Interpreted through the thermodynamic model, this pattern suggests that the TF destabilizes RNAP, with parallels to the classical repression model of kinetic competition(Lanzer & Bujard, 1988). However, the TF compensates for this inhibition by greatly speeding up the initiation rate, allowing for activation of gene expression. It is known that the model TF CRP engages in something similar when regulating Class II promoters (Busby & Ebright, 1997, 1999; Schultz et al., 1991) as binding at these promoters does occlude the -35 promoter sequence; however, this occlusion is made up by stabilizing contacts further upstream at the UP element, effectively increasing the concentration of bound RNAP. This line of reasoning suggests that the MetR binding at position -47 could present a similar occlusion of the -35 sequence, but with the absence of any stabilizing contacts with the RNAP holoenzyme to make up for it. Overall, the work in this chapter highlights the different regulatory modes exhibited by TFs in *E.coli* and demonstrates the convention of classifying TFs as activators or repressors fails to account for conserved modes of action by these regulators.

A major assumption guiding the promoter titration experiments is tied into our definition of promoter strength. As the basal promoter strength represents a fundamental parameter that shapes the net regulatory outcome, this quantity can arise from two potential sources

as depicted in Figure 4.3A: a change in the binding affinity to the TF (either through modulation of RNAP binding and unbinding to the DNA), or through changes in the initiation rate of transcription. As the fundamental contribution of the -35 and -10 sequences to bacterial transcription have been well elucidated(Brewster et al., 2012; Mejía-Almonte et al., 2020), the mutations we made in the -35 and -10 sequences were done to change the affinity of the sigma subunit of the RNAP holoenzyme as the interaction of this subunit with the DNA is a primary determinant of the occupancy. The thermodynamic model offers clear prediction on how stabilizing activators and repressors should behave when regulating promoters with different affinities for RNAP. However, the sigma subunit not only plays a role in recruitment of the holoenzyme to promoters, but dissociation of this subunit from the holoenzyme has been shown to be integral in transitioning RNAP from abortive transcription to productive elongation (Kapanidis et al., 2006; Tang et al., 2009). Therefore, changes to sequence composition of the -35 and -10 could in principle alter this kinetic step, changing both the affinity and the initiation rate. While sequences further downstream have been characterized as promoting the RNAP unwinding of DNA (Josaitis, Gaal, & Gourse, 1995), the previous point does indeed present a possibility: for TFs UlaR, YqhC, and SoxS, some of the designed promoters did not regulate in line with expectation suggesting a possible change in factors other than the affinity of RNAP to the promoter. However, most of the promoters across both stabilizing and destabilizing TFs did show the expected trend according to the thermodynamic model, indicating that changes to RNAP binding to DNA were indeed a substantial part of the effect of these targeted promoter mutations.

5.8 – Extension of the Regulatory Modes of Individual TFs to Combinatorial TF Activity

A natural extension to this work is determining how combinations of TFs with known stabilization and acceleration profiles regulate gene expression. The thermodynamic model provides an intuitive framework to probe these combinatorial regulation schemes with the extension from single TF to pairwise TF regulation determined by three additional parameters – the pairwise TF interaction parameter, the parameter that delineates the free energy of the ternary TF1-TF2-RNAP complex, and the initiation rate (acceleration) coefficient for the co-bound state. These parameters can be inferred from the measurement data of the combined circuit by using the TF specific χ and FC_{max} measured from the individual response functions. An interesting feature of combined regulation can be gleaned from the regulation of pairwise TFs that engage in stabilization of RNAP when regulating in isolation. This pairwise regulatory scheme has interesting implications for combinations of stabilizing activators and repressors that we delineate in the Results Section 4.2.3. In Figure 5.1, we demonstrate predictions from the thermodynamic models for TFs regulating incoherently – utilizing stabilization yet decelerating RNAP.

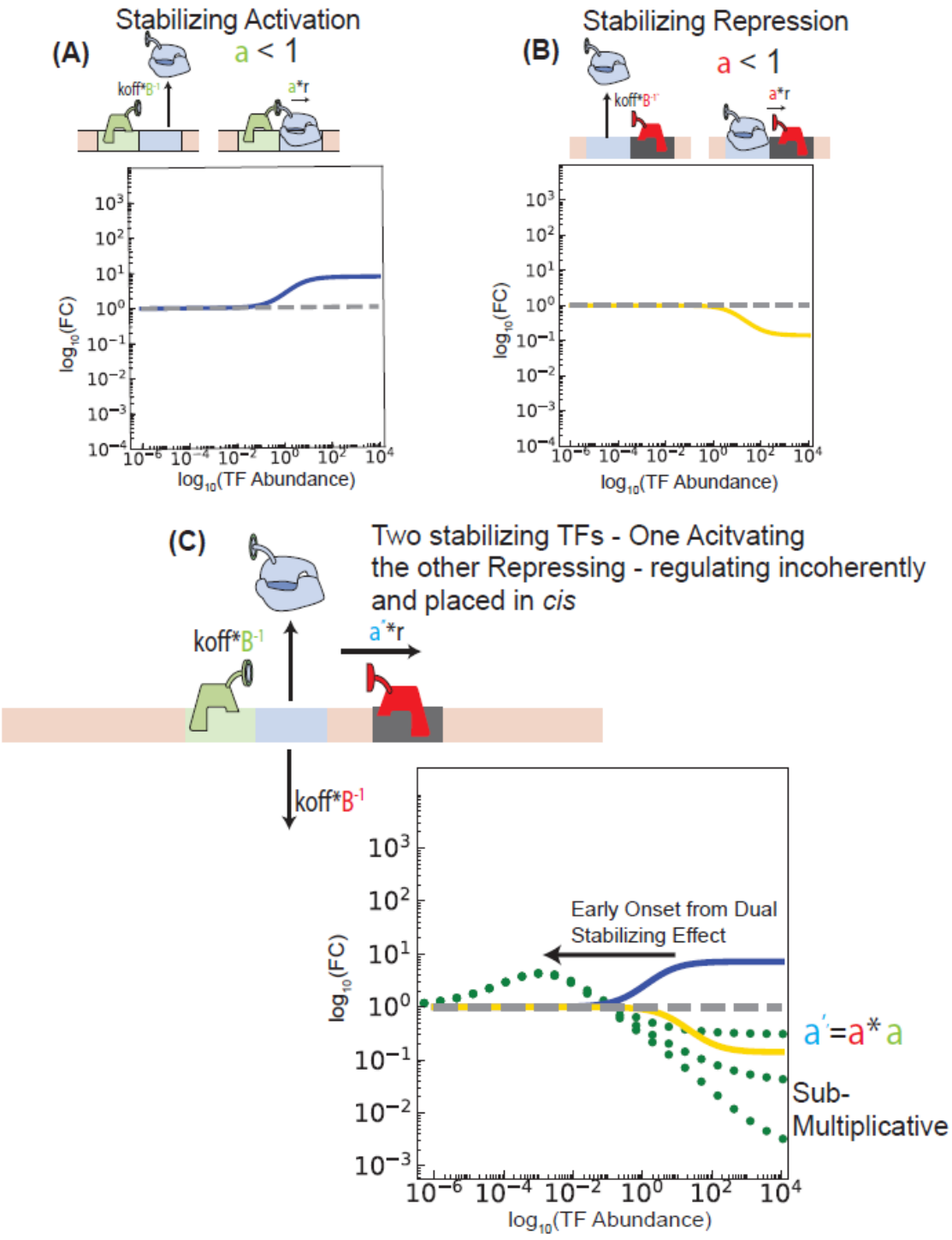

Stabilizing Activation
(A)
a < 1
koff*B-1
a*r
log10(FC)
log10(TF Abundance)
Stabilizing Repression
(B)
a < 1
koff*B-1
a*r
log10(FC)
log10(TF Abundance)
(C)
Two stabilizing TFs - One Acitvating
the other Repressing - regulating incoherently
and placed in cis
koff*B-1
a'*r
koff*B-1
Early Onset from Dual
Stabilizing Effect
a'=a*a
Sub-
Multiplicative
log10(FC)
log10(TF Abundance)

Figure 5.1: Combinatorial Regulation of TFs Regulating by Distinct Modes Leads to Dynamic Regulatory Outcomes. Pairwise regulation of TFs that regulate incoherently (opposing use of stabilization and acceleration) can generate diverse regulatory behavior when placed in *cis*. **(A)** Activator that features strong stabilization but weak deceleration of RNAP (incoherent activation). **(B)** For repression the strong stabilization is offset by strong deceleration – with a net outcome of repression (incoherent repression). **(C)** Predicted Fold-Changes for the co-regulation state. The individual activation and repression curves are show for reference. (solid lines). The possible co-regulatory profiles (dotted lines) are shown where the fold-change is computed by dividing the regulatory response for the co-regulated promoter by the regulatory response of the single repression curve **(B)**. For simplicity, the full two-dimensional response is not shown (independent titration of both TFs): only the diagonal of the co-regulated response matrix is shown. Note the presence of non-monotonic response for the dotted curves (initial activation at lower TF concentrations followed by repression).

In Panel A, this incoherent regulation is shown for activation, where the strong stabilization ($\beta > 1$) translates to a higher effective concentration of RNAP at the promoter, offsetting the reduced initiation rate ($\alpha < 1$) by the TF. We show that this scenario is biologically plausible in Section 2.2.3, where regulation of CpxR at binding positions -72 and -80 conforms to this regulation (See Figure 2.4A, Table 2.1). For the repressing TFs (Panel B), the decrease in the initiation rate is substantial, which despite stabilizing RNAP at the promoter, results in repression (solid yellow line). This features in the regulation for TFs YqhC, MelR, and UlaR that was delineated in Section 4.2.3. Co-regulation of these two TFs according to the thermodynamic model demonstrates a diverse range of possible outcomes as shown in Panel C. Here we plot the fold-change in expression of the co-regulated promoter (Y-axis) against the TF abundance for both the

activator and repressor (X-axis). The fold-change is measured by normalizing the co-regulation response to the response from the repressing TF. In this scenario, we see regulation at lower TF occupancy than the individual TFs (dotted curves), due to the combined stabilizing effect of both factors operating in *cis*. This translates to higher effective concentrations of both activator and repressor at the promoter, skewing the distribution of states to the individually bound TF and co-bound activator and repressor states (See **Figure 1.2** – Thermodynamic Model of Gene Regulation).

The thermodynamic prediction of pairwise TF regulation highlights an important feature: the dynamic regulatory response as seen in the non-monotonic profile (dotted curves). The initial activation of gene expression across all the curves is due to the distribution of bound states mediated by the stabilizing response. The activating TF, due to its higher stabilization coefficient, sets the initial regulatory response as the distribution of promoter states is dominated by the activator bound state. However, increasing TF abundance translates to a shift in the states to co-bound repressor and activator, due to the increase in repressor concentration as well as the stabilizing role of the repressor at the promoter (promoting a higher effective concentration of the TF). As both TFs regulate by stabilizing RNAP, the net effect results in saturation of RNAP occupancy, leaving the regulatory response to be set by the acceleration coefficient (changes to the initiation rate). As both TFs decelerate transcription initiation ($\alpha < 1$), this provides the inflection point in the activation response (Figure 5.1C) and marks the transition to the repressive regime with increasing TF abundance.

These predictions of pairwise TF regulation convey the utility in uncovering the regulatory modes of individual TFs, with the aim of applying this information to explaining the behavior of natural promoters. Dynamic outcomes of gene expression as demonstrated in this pairwise framework do not result from temporal expression patterns of individual regulators but rather their combined regulatory output in terms of their propensity to stabilize RNAP and modulate the rate of transcription initiation. As an increasing number of *E.coli* promoters are being identified as regulated by more than one TF(Han, Li, Filko, Li, & Zhang, 2023; Salgado et al., 2023; Shimada, Ogasawara, Kobayashi, Kobayashi, & Ishihama, 2021), our ability to quantitatively predict expression patterns and therefore the state of the cell, is increasingly reliant on developing useful models of combinatorial TF regulation (Harden et al., 2023; Martinez-Corral et al., 2023; Reiter, Wienerroither, & Stark, 2017; Scholes et al., 2017). However, knowing how individual TFs operate on the kinetic steps of transcription is the requisite information necessary for building more complex schemes of promoter regulation, and provides a powerful justification for interrogating TFs under this line of inquiry.

5.9 - The Role of TF Regulatory Modes in Setting the Timing of Gene Expression

Gene expression is a concerted process: factors necessary for the cell to achieve a particular state are often under the control of more than one promoter, requiring expression of these factors to be coordinated in time. Timing in gene regulation plays an important role in how cells make transitions from one state to another, and several studies have emphasized the importance of cis-acting features at the promoter in mediating the flow of temporal information across gene regulatory networks(Ali & Brewster, 2022;

Miano et al., 2023). Kinetic measurements of promoter dynamics have demonstrated that promoters regulating genes are expressed in temporal order that relates to the order of their products in a metabolic pathway or to key structural components necessary in protein complexes (Kalir et al., 2001; Zaslaver et al., 2004), optimizing gene expression to ensure cellular resources are directed at the right moment (Co, Lagomarsino, Caselle, & Osella, 2017). While the importance of cis-regulatory determinants in mediating this temporal cascade is well attested, there is little systematic investigation on how different regulatory regimes set by the dual regulatory modes appreciably impacts timing in gene expression. This shortcoming is an important barrier, as a unifying framework on gene regulation must not only account for the quantitative amount of gene production, but also how different temporal regimes in expression are available to cells through control by distinct regulators. We show in Figure 5.2 preliminary evidence for the role of different stabilization and acceleration regulatory regimes in mediating the response time of the promoters that suggests further investigation is warranted. For this analysis, we utilize the full kinetic model incorporating the regulatory effect of the TF to determine the time for a promoter to produce half its maximum gene product (the first passage time). In the incoherent regimes of strong stabilization and deceleration, the first passage

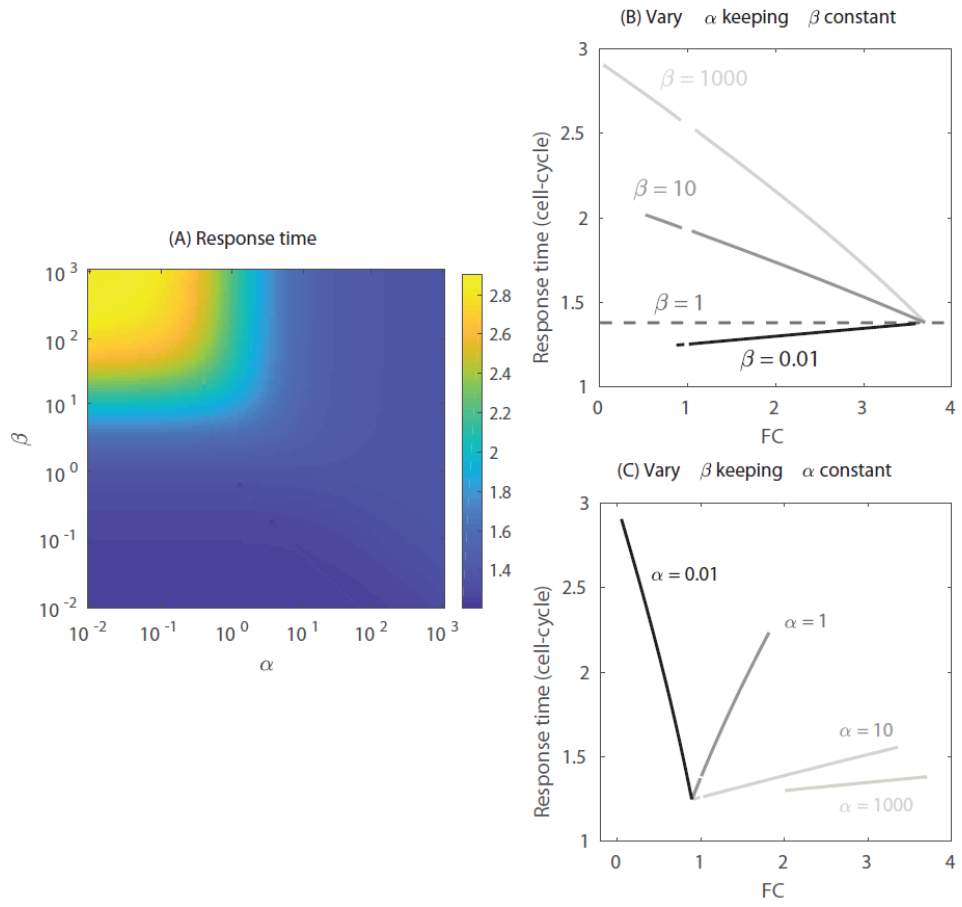


Figure 5.2: The Role of TF Regulatory Modes in Setting the First Passage Time (A) To compute the response time , we considered the full kinetic model of TF regulation using the Chemical Master Equation. The quantity of interest was the probability of having m mRNAs at time t. Taking the expected values of this distribution, we solved for the mean mRNA as a function of the promoter occupancy and time. We then solved the ordinary differential equations in MATLAB using the ode45 solver. We assume that at t=0, the gene is constitutively expressed – at time t=0, the TFs are switched to an active regime and regulate the target promoter. The response time is the time taken to reach half the protein expression of the regulated state. **(B,C)** Role of Stabilization and Acceleration on the promoter response time. For each plot, we kept one of the regulatory modes constant while varying the other to isolate the effect of each mode.

time is longest, registering higher than two cell cycles (Figure 5.2A). This feature of stabilization is relevant to our data in Chapter 3 (Figure 4.3D) as several repressors we profiled regulate in this regime. In Figure 5.2B, we show that changing the degree of stabilization (as acceleration is kept constant) impacts the first passage time and vice versa (Figure 5.2C) The roles of stabilization and acceleration function oppositely in setting the timing of a promoter as increasing the initiation rate of shortens the time to expression. This presents a regulatory strategy for increasing the steady state levels and the timing response of genes simultaneously. Although integrating a TF that functions coherently ($\alpha > 1 \; and \; \beta > 1$) will generate a strong promoter response, it does so at the expense of several cell cycles. However, integrating a TF that functions in the incoherent or acceleration only regime ($\alpha > 1 \; and \; \beta < 1 \; or \; \alpha > 1 \; and \; \beta \sim 1$) could allow for an increasing amount of gene expression with the benefit of a faster response. This breakdown of the impact of the regulatory modes on timing of expression has interesting implications for gene regulatory networks, particularly for TFs that regulate first in a gene regulatory pathway. It would be interesting to determine if these TFs function using an incoherent or an initiation only strategy in setting the response of their downstream targets, as this ensures robust expression with a fast turnaround time.

5.10 - Final Thoughts

The work presented in this thesis demonstrates the dual approach of using theory and experiment to understand gene regulation: we framed the process of gene regulation using a thermodynamic model that explicitly accounts for TF regulation on distinct steps of the kinetic process with the aim of answering what determinants set these features.

With an intuitive model in hand, we leveraged a synthetic biology approach to test the influence of binding sequence and position and tease out the contributions of these features in setting the TFs regulatory function. Central to the argument of this thesis is the utility of this interplay between modelling and experiment. What started out as characterizing TF function transitioned into making testable predictions on how distinct promoters would be regulated based on acquired information on the regulatory modes of *E.coli* TFs. Our goal was to uncover the precise contributions of cis-regulatory elements in setting TF activity, and this work makes the case that relatively simple models of gene regulation capture these contributions quite well. This suggest that future studies in gene expression could benefit from a "bottoms-up" approach: where individual contributions of gene regulation (TF identity being an important one) are dissected and used to inform thinking on how natural promoters are constructed from fundamental components. Uncovering a "genetic code" or regulatory grammar that can map how features of the promoter such as TF identity, binding site position and sequence set the quantitative characteristics of gene expression might hinge on our ability to utilize these simple models effectively.

Appendix

6.1 – Table of sequences used in measuring the isolated regulatory output of 93 *E.coli* TFs.

Table 6.1: List of Sequences Used for the Cloning and Measurement of TF regulatory Function. The information provided includes the TF and the operator sequences used in the quantitative profiling of TF function in Chapter 3. The promoters in which the sequences were found are listed in the 2nd column. Blank entries for the promoter column indicate that the binding sequence does not situate proximally to any annotated promoter in the *E.coli* genome.

TF	Promoter	Sequence (5' to 3')
Fis	_dusBp	GCCGATTATTTACGC
Fis	_cbpAp6	GTTTAAAATATGTTC
Fis	_dusBp	GTGTAAATTTTAGTC
BasR	_tomBp1	AACATTAAGGAGTAATTGAA
BasR	_putAp	TATTTCACGATAACGTTAAG
FNR	_grcAp1	TTGATTTAAATCAA
FNR	_narKp1_nirBp	TTGATTTACATCAA
GlpR	_glpABCp	ATATGCGCGAAATCAAACAA
GlpR	_glpDp	TATGTTCGATAACGAACATT
ArgP	_argOp	TTTCTGATTGCCAATTAATA
ArgP	_argOp	TTATCAATTTCCGCTAATAA
MalT	_malEp	GGGGAGGAGG
MalT	_malPp	CAGGATGAGG
ArgR	_argFp	AGTGAATTTTAATTCAAT
ArgR	_argIp	ATTGAATTTTAATTCATT
TrpR	_mtrp2	GTACTCGTGTACTGGTAC
TrpR	_trpLp	GAACTAGTTAACTAGTAC
ArcA	_nuoAp2	GTTGATATTATGTAA
ArcA	_hyaAp	TTCACTACGCTTTATTAACAAT

Ada	_aidBp	TGCTGGATAAGAATGTTTTAGCAA
Ada		AACCGGAATATGAAAGCAAAGCGC
BetI	_betIp	TATATTGAACGTCCAATCAA
IclR	_aceBp	TGGAAATTGTTTTTG
IclR	_aceBp	TTTAATT
PurR	_purHp	GCGCAAACGTTTTCGT
PurR	_purRp	AGGCAAACGTTTACCT
NhaR	_osmCp1	ATTCGGAATATCCTGCT
NhaR	_nhaAp1	GCTCGTAAAAAACGAAT
RcnR	_rcnRp	TACTACCCCCCAGTA
RcnR	_rcnRp	TACTCCCCCCCAGTA
ExuR	_exuRp2	AAAGTGGTATAACAAATA
ExuR	_exuRp2	CAGTGTTCGACCTGTTAG
FabR	_fabAp	AGCGTACACGTGTTAGCT
FabR	_fabBp	GGCGTACAAGTGTACGCT
MelR	_melAp_melAp_melRp	ATCTGAGTTTATGGGAAT
MelR	_melRp	TTCTGATATTCCAGGAAA
PdhR	_pdhRp	AATTGGTAAGACCAATT
PdhR	_cyoAp	CATTGGTTATACCAATT
IscR	_iscRp	ATAGTTGACCAATTTACTCGGGAAT
IscR	_iscRp	ATACCCGACTAAATCAGTCAAGTAA
NarP	_fdnGp	TTTTCTACCGCTATTG
NarP	_nirBp	CAATATACCCATTAAG
Lrp	_argOp	CTTATTAGTTTTTCTG
Lrp	_argOp	CAGATAATACTCTCTG
SoxS	_ybjCp	GTTGCATTAATTGCTAAAAG
SoxS	_sodAp	AAGTACGGCATTGATAATCA
CytR	_cytRp	TTTTCACGCGCTATCTGCTAA
CytR	_tsxp2	AAACGAAACATATTTTTGTGAGCAAT
MraZ	_mraZp	TGGGA
MraZ	_mraZp	TGGGG
OmpR	_tppBp	GTAACAGATTATTACAAAGG
OmpR	_ompCp1	TTTACATTTTGAAACATCTA
LacI	_lacZp1	GGCAGTGAGCGCAACGCAATT
LacI	_lacZp1	AATTGTGAGCGGATAACAATT
Cra	_pykFp	TCTTGAATGGTTTCAGCA
Cra	_hypFp2_hypFp3	TGATGAACCGTTTCATCA
RcdA	_bluFp	ATGTGTACAG
RcdA	_sulAp	CTATGCACAT
NagC	_nagBp	CTTATTTTATCATTCAAAAAATC
NagC	_nagBp	TTTAATTCGTATCGCAAATTAAA

PhoB	_phoBp	TTTTCATAAATCTGTCATA
PhoB	_pstSp	CTGTCATAAAACTGTCATA
GntR	_gntTp1	TGACGTTACCCATAACAAAT
GntR	_gntKp	CAATGTTACCGATAACAGTT
FeaR	_feaBp1	AGCTTGTCACAGCTGACAAAG
FeaR	_tynAp1	AAGCTGGCACACCTGCCAAAC
UlaR	_ulaAp	CATTTAATCATAAATAATCT
UlaR	_ulaAp	TGATTACTTTTGAAAATTAG
RutR	_carAp1	TTGACCATTTGGTCCA
RutR	_rutRp2	TGGACTAAACGGTCAAAA
YqjI	_yqjHp	ATTTAGATATATCTGATTA
YqjI		ATATCAGATATAAATTAGATATATCTAATTA
PaaX	_paaXp	TGATTCATAAATCACAACAATAACAACAGAC TGAAGCG
PaaX	_paaAp	TGATTCGTGTTTTTAATTAATTCACGAAAACT GGAATCG
CynR	_cynTp	ATAAGTAAA
CynR	_cynTp	TTACCTTAT
Cbl		AGTAATTGCATATTTAATCTTTCCTTAGCCGT TTTTTTGC
Cbl	_ssuEp	CTTTAGTTTATTTTCAGAAAAAGATACACAAA ATTTATAT
HipB	_relAp1	TATCCTGTAACCCTGCAACGCTGGCTCGGGA TA
HipB	_hipBp	TTATCCCGTAGAGCGGATAA
CaiF	_fixAp	TTTCAATATTGGTGAG
CaiF	_fixAp	TTTCAATATTGGTGAT
AgaR	_agaSp	TTTCTTTTTCTTTCGTTTTG
AgaR	_agaSp	TTTCGTTTTATTTCTTTTTT
UxuR		AATGTGATTAACCAGGTC
UxuR	_uxuAp	AAATTGGTCAACCATTGT
LeuO	_casp	TATATTCACAGTATGAATATT
LeuO	_casp	AAAATTCATGGTAATTATTAT
NarL	_focAp1_focAp3	TAAATATCGGTAATTT
NarL	_narGp	GAATTTACTTTATTTT
NsrR	_fliAp1	ATAACTCATAT
NsrR	_nrfAp	GTGGTATTTTACATGCACT
DcuR	_frdAp	GTTTAGTAATTAAATTA
DcuR	_dctAp	CAAAACTTTAAAAAGTG
Nac	_codBp	ATATAAAAAATATAT
Nac		ATAAATTTTCTCAGG

CRP	_mtlAp	AAATGTGACACTACTCACATTT
CRP	_hptp	ATGTGTGATCGTCATCACAATT
CRP	_mtlAp	TTATGTGATTGATATCACACAA
CusR	_cusCp	ATGACAATTTTGTCAT
CusR	_yedXp	ATTACAAAATTGTAAT
AraC	_araBp_araBp	TAGCATTTTTATCCATA
AraC	_araEp	CAGCAATTTAATCCATA
IlvY	_ilvCp	GTTGCAAAAATTGCAAT
IlvY	_ilvCp	ATATCAATTTCCGCAAT
YqhC	_yqhDp	ATTTCTCCAGCACTCTGGAGAAAT
YqhC	_yqhDp	AATAGGCAAGACATTGGCAGAAAT
Fur	_gpmAp	AATGAGAATTATTATCATT
Fur	_ryhBp	AATGAGAATGATTATTATT
UhpA	_uhpTp	GCCTGGACGCTATCTCAGGC
UhpA	_uhpTp	AACTAAGAAATTTTCCAGGT
CsgD	_fliAp1	AAACGGATAATCAT
CsgD	_fliFp1	AAAGGCGCTAAATA
GlrR	_rpoEp2b	TGTTTAATTTACTAAACA
GlrR	_rpoEp2b	TGTTCAGGGAGAGTATTCAT
AcrR	_acrAp	TACATACATTTGTGAATGTATGTA
AcrR	_marRp	TTCATTCATT
FadR	_fabBp	GGCTGATCGGACTTGTT
FadR	_fadIp	ATCAGGTCAGACCACTT
MngR	_mngRp	CATAATGGTTAAATTCGTATTA
MngR	_mngRp	TATTTGTATTGTATTTGTATTA
CysB	_cysKp1	CCGCCCCTTATTCCATCTTGCATGTCATTATTT CCCTTCTGT
CysB	_hslJp	GATAATCCCCAATGACCTTTCATCCTCTATTC TTAAAATAGT
MarA	_purAp	TTTGAGTGCAAAAAGTGCTG
MarA	_fprp	TGATTTGATCGATTGAGCCT
RhaS	_rhaBp	ATCTTTCCCTGGTTGCC
RhaS	_rhaTp	TTAATGCCGTGATTGCC
RhaR	_rhaSp	CGTCGAAAATTTAAGGT
RhaR	_rhaSp	ATCTTGAAAAATCGACG
DeoR	_deoCp2	ATGTTAGAATACTAAC
DeoR	_deoCp1	GTGTTAGAATTCTAAC
YdeO	_gadEp3	ATTTCAAACATTATCATGGCTGATAT
YdeO	_slpp	ATATCACCTCAGAATCAGATGAAAA
PutA	_putPp1	GGTGCA
PutA	_putAp	GTTGCA

AllR	_gclp	TTGGAAAAATTTTCCAA
AllR		TTGGAAAAATATTCCAA
MarR	_marRp	ATTACTTGCCAGGGCAACTAAT
MarR	_marRp	ATACTTGCCTGGGCAATATTAT
TorR	_tnaCp	CGATTCACAT
TorR	_torRp	ATATGAACAG
NtrC	_glnAp1	AGCACTATATTGGTGCA
NtrC	_glnLp	TGCACTAAAATGGTGCA
MetR		ATGCAAATTTTTCAC
MetR	_glyAp	CTGAAAAGAAATCCG
ArsR	_arsRp	TAAGTCATATATGTTTTTGACTTA
GalR	_galPp	GTGTAAGCGATTACAC
GalR	_galEp2_galEp1_galEp2	GTGTAAACGATTCCAC
EvgA	_safAp	TGCCTACAGCTGTAAGAA
EvgA	_yfdXp	TTCCTACAATTGTAAGAC
LldR	_lldPp	CACAATTGGCAGTGCCACTTTTA
LldR	_lldPp_lldPp	AAGAATTGGCCCTACCAATTCTT
NrdR	_nrdDp	ACCCAATATGTTGTAT
NrdR	_nrdAp	CCCCTATATATAGTGT
TyrR	_aroFp	TGTAAATTTATCTATACA
TyrR	_mtrp2	TGTAAAATAATATATACA
MetJ	_metFp	TTTACATCTGGACGTC
MetJ	_folEp	ATAACGATGTTTTAAC
MurR	_murQp1_murQp2	AGACTCAATGATTCC
MurR	_murQp1_murQp2	GGGATATATGATTCC
CueR		CCTTCCCGTAAGGGGAAGG
CueR	_copAp	CCTTCCCCTTGCTGGAAGG
PhoP	_iraMp2	TGTTTAACATCCATTGA
PhoP	_safAp	TGATTAACGATTTTTAA
RbsR	_rbsDp	TCAGCGAAACGTTTCGCTGA
RbsR	_purHp	TTGCGCAAACGTTTTCGTTA
BaeR	_spyp	CTTTTTTTCTCCATAATTGG
BaeR	_spyp	GGTATAGTTCTTCATAATCT
SutR	_sutRp	CGTTATAACGCACAAGTG
SutR	_sutRp	ACCTCGTGCGTTATAGCGTACA
QseB	_qseBp2	AAAATAAAATTTA
QseB	_qseBp2	CAATTACGGATTT
Rob	_inaAp	ACGACACGTTTCATTAAGAT
Rob	_fumCp1	ATGGCACGAAAGACCAAACA
XylR	_araCp	CTGCCGTGATTATAGACA

XylR	_araCp	TTGTTACGCGTTTTTGTC
PepA	_carAp1	ATTTGGTCCACTTTTTTCTGCTCGTTT
PepA	_pepAp1	CGTCTTTTGCAGGATTTTAGCTTGTTT
CpxR	_ppiAp1	GTAAAATTAGGTAAA
CpxR	_cpxPp	GTAAACTCTCTATCGTTGA
GadX	_gadEp3	ATGGCTGATATTTTCCGTAG
GadX	_gadEp3	TTAGCCATTTCAAACATTAT
ModE	_modAp	CGTTATATTGTCGCCTACATAACG
ModE	_napFp2_napFp1	CGCTATATAAATATATTTATAACC
GcvA	_gcvTp	TTGAAATAAACTAAT
GcvA	_gcvTp	ATTATATTTTCTAAT
TreR	_treBp	TGCGTTTCGGGAACGTT
TreR	_treBp	ATCGTTTCCTGAACGAT
Dan	_ttdAp	GTTAATT
Dan	_ttdAp	GTTGATT
AscG	_ascFp	TGAAACCGGTTTCT
AscG	_ascFp	GGTGACCGGTTTCA

Table 6.2: LacUV5 Promoter Variants Used in Construction of Promoters with Different Basal Expression Levels (Arranged from Strongest to Weakest).

-35	Spacer	-10
TTGACA	CTTTATGCTTCCGGCTCG	TATAAT
TTTACA	CTTTATGCTTCCGGCTCG	TATAAT
TTGACA	CTTTATGCTTCCGGCTCG	AATAAT
TTTACC	CTTTATGCTTCCGGCTCG	TATAAT
ATTACA	CTTTATGCTTCCGGCTCG	TATAAT
TTAAGA	CTTTATGCTTCCGGCTCG	TATAAT
TTAAGA	CTTTATGCTTCCGGCTCG	AATAAT

Bibliography

Ackers, G. K., Johnson, A. D., & Shea, M. A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci U S A, 79*(4), 1129-1133. doi:10.1073/pnas.79.4.1129

Afif, H., Allali, N., Couturier, M., & Van Melderen, L. (2001). The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. *Mol Microbiol, 41*(1), 73-82. doi:10.1046/j.1365-2958.2001.02492.x

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., & Harrison, S. C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science, 242*(4880), 899-907. doi:10.1126/science.3187531

Ali, M. Z., & Brewster, R. C. (2022). Controlling gene expression timing through gene regulatory architecture. *PLoS Comput Biol, 18*(1), e1009745. doi:10.1371/journal.pcbi.1009745

Ali, M. Z., Parisutham, V., Choubey, S., & Brewster, R. C. (2020). Inherent regulatory asymmetry emanating from network architecture in a prevalent autoregulatory motif. *Elife, 9*. doi:10.7554/eLife.56517

Allen, B. L., & Taatjes, D. J. (2015). The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol, 16*(3), 155-166. doi:10.1038/nrm3951

Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., & Iyer, L. M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev, 29*(2), 231-262. doi:10.1016/j.femsre.2004.12.008

Aravind, L., & Koonin, E. V. (1999). DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res, 27*(23), 4658-4670. doi:10.1093/nar/27.23.4658

Arthur, B. P., François, J., & Jacques, M. (1959). The genetic control and cytoplasmic expression of "Inducibility" in the synthesis of β-galactosidase by E. coli. *Journal of Molecular Biology, 1*(2), 165-178. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022283659800450. doi:https://doi.org/10.1016/S0022-2836(59)80045-0

Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., & Teichmann, S. A. (2004). Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol, 14*(3), 283-291. doi:10.1016/j.sbi.2004.05.004

Baleja, J. D., & Sykes, B. D. (1991). Comparison of the structures of operator DNA free and in complex with lambda repressor. *Biochem Cell Biol, 69*(2-3), 202-205. doi:10.1139/o91-030

Barker, A., Fickert, R., Oehler, S., & Müller-hill, B. (1998). Operator search by mutant Lac repressors. *J Mol Biol, 278*(3), 549-558. doi:10.1006/jmbi.1998.1729

Bartman, C. R., Hsu, S. C., Hsiung, C. C., Raj, A., & Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell, 62*(2), 237-247. doi:10.1016/j.molcel.2016.03.007

Bashor, C. J., Patel, N., Choubey, S., Beyzavi, A., Kondev, J., Collins, J. J., & Khalil, A. S. (2019). Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science, 364*(6440), 593-597. doi:10.1126/science.aau8287

Bateman, E. (1998). Autoregulation of eukaryotic transcription factors. *Prog Nucleic Acid Res Mol Biol, 60*, 133-168. doi:10.1016/s0079-6603(08)60892-2

Baumgart, L. A., Lee, J. E., Salamov, A., Dilworth, D. J., Na, H., Mingay, M., . . . O'Malley, R. C. (2021). Persistence and plasticity in bacterial gene regulation. *Nat Methods, 18*(12), 1499-1505. doi:10.1038/s41592-021-01312-2

Becker, N. A., Peters, J. P., Maher, L. J., 3rd, & Lionberger, T. A. (2013). Mechanism of promoter repression by Lac repressor-DNA loops. *Nucleic Acids Res, 41*(1), 156-166. doi:10.1093/nar/gks1011

Beer, M. A., & Tavazoie, S. (2004). Predicting gene expression from sequence. *Cell, 117*(2), 185-198. doi:10.1016/s0092-8674(04)00304-6

Beerli, R. R., & Barbas, C. F., 3rd. (2002). Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol, 20*(2), 135-141. doi:10.1038/nbt0202-135

Benson, N., Sugiono, P., & Youderian, P. (1988). DNA sequence determinants of lambda repressor binding in vivo. *Genetics, 118*(1), 21-29. doi:10.1093/genetics/118.1.21

Berg, O. G., Winter, R. B., & von Hippel, P. H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry, 20*(24), 6929-6948. doi:10.1021/bi00527a028

Berthoumieux, S., de Jong, H., Baptist, G., Pinel, C., Ranquet, C., Ropers, D., & Geiselmann, J. (2013). Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol Syst Biol, 9*, 634. doi:10.1038/msb.2012.70

Besse, M., von Wilcken-Bergmann, B., & Müller-Hill, B. (1986). Synthetic lac operator mediates repression through lac repressor when introduced upstream and downstream from lac promoter. *Embo j, 5*(6), 1377-1381. doi:10.1002/j.1460-2075.1986.tb04370.x

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., & Phillips, R. (2005). Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev, 15*(2), 116-124. doi:10.1016/j.gde.2005.02.007

Boedicker, J. Q., Garcia, H. G., Johnson, S., & Phillips, R. (2013). DNA sequence-dependent mechanics and protein-assisted bending in repressor-mediated loop formation. *Phys Biol, 10*(6), 066005. doi:10.1088/1478-3975/10/6/066005

Boeger, H., Griesenbeck, J., & Kornberg, R. D. (2008). Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell, 133*(4), 716-726. doi:10.1016/j.cell.2008.02.051

Boettiger, A. N., & Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science, 325*(5939), 471-473. doi:10.1126/science.1173976

Borukhov, S., & Lee, J. (2005). RNA polymerase structure and function at lac operon. *C R Biol, 328*(6), 576-587. doi:10.1016/j.crvi.2005.03.007

Bragdon, M. D. J., Patel, N., Chuang, J., Levien, E., Bashor, C. J., & Khalil, A. S. (2023). Cooperative assembly confers regulatory specificity and long-term genetic circuit stability. *Cell, 186*(18), 3810-3825.e3818. doi:10.1016/j.cell.2023.07.012

Brewster, R. C., Jones, D. L., & Phillips, R. (2012). Tuning promoter strength through RNA polymerase binding site design in Escherichia coli. *PLoS Comput Biol, 8*(12), e1002811. doi:10.1371/journal.pcbi.1002811

Brewster, R. C., Weinert, F. M., Garcia, H. G., Song, D., Rydenfelt, M., & Phillips, R. (2014). The transcription factor titration effect dictates level of gene expression. *Cell, 156*(6), 1312-1323. doi:10.1016/j.cell.2014.02.022

Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D. K., Carmi, M., & Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol Cell, 79*(3), 459-471.e454. doi:10.1016/j.molcel.2020.05.032

Brückner, R., & Titgemeyer, F. (2002). Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol Lett, 209*(2), 141-148. doi:10.1111/j.1574-6968.2002.tb11123.x

Buescher, J. M., Liebermeister, W., Jules, M., Uhr, M., Muntel, J., Botella, E., . . . Sauer, U. (2012). Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. *Science, 335*(6072), 1099-1103. doi:10.1126/science.1206871

Bulcha, J. T., Giese, G. E., Ali, M. Z., Lee, Y. U., Walker, M. D., Holdorf, A. D., . . . Walhout, A. J. M. (2019). A Persistence Detector for Metabolic Network Rewiring in an Animal. *Cell Rep, 26*(2), 460-468.e464. doi:10.1016/j.celrep.2018.12.064

Bundalovic-Torma, C., Whitfield, G. B., Marmont, L. S., Howell, P. L., & Parkinson, J. (2020). A systematic pipeline for classifying bacterial operons reveals the evolutionary landscape of biofilm machineries. *PLoS Comput Biol, 16*(4), e1007721. doi:10.1371/journal.pcbi.1007721

Burda, Z., Krzywicki, A., Martin, O. C., & Zagorski, M. (2011). Motifs emerge from function in model gene regulatory networks. *Proc Natl Acad Sci U S A, 108*(42), 17263-17268. doi:10.1073/pnas.1109435108

Busby, S., & Ebright, R. H. (1997). Transcription activation at class II CAP-dependent promoters. *Mol Microbiol, 23*(5), 853-859. doi:10.1046/j.1365-2958.1997.2771641.x

Busby, S., & Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *J Mol Biol, 293*(2), 199-213. doi:10.1006/jmbi.1999.3161

Cai, L., Friedman, N., & Xie, X. S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature, 440*(7082), 358-362. doi:10.1038/nature04599

Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., & Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature, 453*(7194), 544-547. doi:10.1038/nature06965

Chen, W. C. W., Gaidukov, L., Lai, Y., Wu, M. R., Cao, J., Gutbrod, M. J., . . . Lu, T. K. (2022). A synthetic transcription platform for programmable gene expression in mammalian cells. *Nat Commun, 13*(1), 6167. doi:10.1038/s41467-022-33287-9

Choi, J. K., & Kim, Y. J. (2009). Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat Genet, 41*(4), 498-503. doi:10.1038/ng.319

Choy, H. E., Hanger, R. R., Aki, T., Mahoney, M., Murakami, K., Ishihama, A., & Adhya, S. (1997). Repression and activation of promoter-bound RNA polymerase activity by Gal repressor. *J Mol Biol, 272*(3), 293-300. doi:10.1006/jmbi.1997.1221

Co, A. D., Lagomarsino, M. C., Caselle, M., & Osella, M. (2017). Stochastic timing in gene expression for simple regulatory strategies. *Nucleic Acids Res, 45*(3), 1069-1078. doi:10.1093/nar/gkw1235

Coomer, M. A., Ham, L., & Stumpf, M. P. H. (2022). Noise distorts the epigenetic landscape and shapes cell-fate decisions. *Cell Syst, 13*(1), 83-102.e106. doi:10.1016/j.cels.2021.09.002

Dalma-Weiszhausz, D. D., & Brenowitz, M. (1996). Interactions between RNA polymerase and the positive and negative regulators of transcription at the Escherichia coli gal operon. *Biochemistry, 35*(12), 3735-3745. doi:10.1021/bi952408s

de Boer, C. G., Vaishnav, E. D., Sadeh, R., Abeyta, E. L., Friedman, N., & Regev, A. (2020). Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat Biotechnol, 38*(1), 56-65. doi:10.1038/s41587-019-0315-8

Dekel, E., & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature, 436*(7050), 588-592. doi:10.1038/nature03842

DelRosso, N., Tycko, J., Suzuki, P., Andrews, C., Aradhana, Mukund, A., . . . Bintu, L. (2023). Large-scale mapping and mutagenesis of human transcriptional effector domains. *Nature, 616*(7956), 365-372. doi:10.1038/s41586-023-05906-y

Déthiollaz, S., Eichenberger, P., & Geiselmann, J. (1996). Influence of DNA geometry on transcriptional activation in Escherichia coli. *Embo j, 15*(19), 5449-5458.

Dey, S., Soltani, M., & Singh, A. (2020). Enhancement of gene expression noise from transcription factor binding to genomic decoy sites. *Sci Rep, 10*(1), 9126. doi:10.1038/s41598-020-65750-2

Djordjevic, M., & Bundschuh, R. (2008). Formation of the open complex by bacterial RNA polymerase--a quantitative model. *Biophys J, 94*(11), 4233-4248. doi:10.1529/biophysj.107.116970

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science, 297*(5584), 1183-1186. doi:10.1126/science.1070919

English, M. A., Gayet, R. V., & Collins, J. J. (2021). Designing Biological Circuits: Synthetic Biology Within the Operon Model and Beyond. *Annu Rev Biochem, 90*, 221-244. doi:10.1146/annurev-biochem-013118-111914

Estrem, S. T., Ross, W., Gaal, T., Chen, Z. W., Niu, W., Ebright, R. H., & Gourse, R. L. (1999). Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev, 13*(16), 2134-2147. doi:10.1101/gad.13.16.2134

Fakhouri, W. D., Ay, A., Sayal, R., Dresch, J., Dayringer, E., & Arnosti, D. N. (2010). Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo. *Mol Syst Biol, 6*, 341. doi:10.1038/msb.2009.97

Farrell, S., Simkovich, N., Wu, Y., Barberis, A., & Ptashne, M. (1996). Gene activation by recruitment of the RNA polymerase II holoenzyme. *Genes Dev, 10*(18), 2359-2367. doi:10.1101/gad.10.18.2359

Feklístov, A., Sharon, B. D., Darst, S. A., & Gross, C. A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol, 68*, 357-376. doi:10.1146/annurev-micro-092412-155737

Ferguson, M. L., Le Coq, D., Jules, M., Aymerich, S., Radulescu, O., Declerck, N., & Royer, C. A. (2012). Reconciling molecular regulatory mechanisms with noise patterns of bacterial metabolic promoters in induced and repressed states. *Proc Natl Acad Sci U S A, 109*(1), 155-160. doi:10.1073/pnas.1110541108

Fitzgerald, D. M., Bonocora, R. P., & Wade, J. T. (2014). Comprehensive mapping of the Escherichia coli flagellar regulatory network. *PLoS Genet, 10*(10), e1004649. doi:10.1371/journal.pgen.1004649

François, J., & Jacques, M. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology, 3*(3), 318-356. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022283661800727. doi:https://doi.org/10.1016/S0022-2836(61)80072-7

Friedman, D. I. (1988). Integration host factor: a protein for all reasons. *Cell, 55*(4), 545-554. doi:10.1016/0092-8674(88)90213-9

Galagan, J. E., Minch, K., Peterson, M., Lyubetskaya, A., Azizi, E., Sweet, L., . . . Schoolnik, G. K. (2013). The Mycobacterium tuberculosis regulatory network and hypoxia. *Nature, 499*(7457), 178-183. doi:10.1038/nature12337

Galburt, E. A. (2018). The calculation of transcript flux ratios reveals single regulatory mechanisms capable of activation and repression. *Proc Natl Acad Sci U S A, 115*(50), E11604-e11613. doi:10.1073/pnas.1809454115

Garcia, H. G., Brewster, R. C., & Phillips, R. (2016). Using synthetic biology to make cells tomorrow's test tubes. *Integr Biol (Camb), 8*(4), 431-450. doi:10.1039/c6ib00006a

Garcia, H. G., Sanchez, A., Boedicker, J. Q., Osborne, M., Gelles, J., Kondev, J., & Phillips, R. (2012). Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep, 2*(1), 150-161. doi:10.1016/j.celrep.2012.06.004

Gertz, J., Siggia, E. D., & Cohen, B. A. (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature, 457*(7226), 215-218. doi:10.1038/nature07521

Giordano, T. J., Deuschle, U., Bujard, H., & McAllister, W. T. (1989). Regulation of coliphage T3 and T7 RNA polymerases by the lac repressor-operator system. *Gene, 84*(2), 209-219. doi:10.1016/0378-1119(89)90494-0

Gitai, Z., Dye, N., & Shapiro, L. (2004). An actin-like gene can determine cell polarity in bacteria. *Proc Natl Acad Sci U S A, 101*(23), 8643-8648. doi:10.1073/pnas.0402638101

Gogos, J. A., Hsu, T., Bolton, J., & Kafatos, F. C. (1992). Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger domain. *Science, 257*(5078), 1951-1955. doi:10.1126/science.1290524

Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell, 123*(6), 1025-1036. doi:10.1016/j.cell.2005.09.031

Goldman, S. R., Ebright, R. H., & Nickels, B. E. (2009). Direct detection of abortive RNA transcripts in vivo. *Science, 324*(5929), 927-928. doi:10.1126/science.1169237

Görke, B., & Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol, 6*(8), 613-624. doi:10.1038/nrmicro1932

Gossen, M., & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A, 89*(12), 5547-5551. doi:10.1073/pnas.89.12.5547

Guharajan, S., Chhabra, S., Parisutham, V., & Brewster, R. C. (2021). Quantifying the regulatory role of individual transcription factors in Escherichia coli. *Cell Rep, 37*(6), 109952. doi:10.1016/j.celrep.2021.109952

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol, 19*(10), 621-637. doi:10.1038/s41580-018-0028-8

Halford, S. E., & Marko, J. F. (2004). How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res, 32*(10), 3040-3052. doi:10.1093/nar/gkh624

Hall, B. G. (1989). Selection, adaptation, and bacterial operons. *Genome, 31*(1), 265-271. doi:10.1139/g89-044

Han, Y., Li, W., Filko, A., Li, J., & Zhang, F. (2023). Genome-wide promoter responses to CRISPR perturbations of regulators reveal regulatory networks in Escherichia coli. *Nat Commun, 14*(1), 5757. doi:10.1038/s41467-023-41572-4

Harden, T. T., Vincent, B. J., & DePace, A. H. (2023). Transcriptional activators in the early Drosophila embryo perform different kinetic roles. *Cell Syst, 14*(4), 258-272.e254. doi:10.1016/j.cels.2023.03.006

Hazkani-Covo, E., & Graur, D. (2005). Evolutionary conservation of bacterial operons: does transcriptional connectivity matter? *Genetica, 124*(2-3), 145-166. doi:10.1007/s10709-005-0950-5

He, X., Samee, M. A., Blatti, C., & Sinha, S. (2010). Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Comput Biol, 6*(9). doi:10.1371/journal.pcbi.1000935

Heiderscheit, E. A., Eguchi, A., Spurgat, M. C., & Ansari, A. Z. (2018). Reprogramming cell fate with artificial transcription factors. *FEBS Lett, 592*(6), 888-900. doi:10.1002/1873-3468.12993

Hensel, Z., Feng, H., Han, B., Hatem, C., Wang, J., & Xiao, J. (2012). Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis. *Nat Struct Mol Biol, 19*(8), 797-802. doi:10.1038/nsmb.2336

Hinman, V., & Cary, G. (2017). The evolution of gene regulation. *Elife, 6*. doi:10.7554/eLife.27291

Hirsh, J., & Schleif, R. (1973). In vivo experiments on the mechanism of action of L-arabinose C gene activator and lactose repressor. *J Mol Biol, 80*(3), 433-444. doi:10.1016/0022-2836(73)90414-2

Hochschild, A., & Dove, S. L. (1998). Protein-protein contacts that activate and repress prokaryotic transcription. *Cell, 92*(5), 597-600. doi:10.1016/s0092-8674(00)81126-5

Hsu, L. M., Cobb, I. M., Ozmore, J. R., Khoo, M., Nahm, G., Xia, L., . . . Ahn, C. (2006). Initial transcribed sequence mutations specifically affect promoter escape properties. *Biochemistry, 45*(29), 8841-8854. doi:10.1021/bi060247u

Inukai, S., Kock, K. H., & Bulyk, M. L. (2017). Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev, 43*, 110-119. doi:10.1016/j.gde.2017.02.007

Ishihama, A. (2012). Prokaryotic genome regulation: a revolutionary paradigm. *Proc Jpn Acad Ser B Phys Biol Sci, 88*(9), 485-508. doi:10.2183/pjab.88.485

Jacques, M., Jean-Pierre, C., & François, J. (1963). Allosteric proteins and cellular control systems. *Journal of Molecular Biology, 6*(4), 306-329. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022283663800911. doi:https://doi.org/10.1016/S0022-2836(63)80091-1

Jensen, D., & Galburt, E. A. (2021). The Context-Dependent Influence of Promoter Sequence Motifs on Transcription Initiation Kinetics and Regulation. *J Bacteriol, 203*(8). doi:10.1128/jb.00512-20

Jensen, D., Manzano, A. R., Rammohan, J., Stallings, C. L., & Galburt, E. A. (2019). CarD and RbpA modify the kinetics of initial transcription and slow promoter escape of the Mycobacterium tuberculosis RNA polymerase. *Nucleic Acids Res, 47*(13), 6685-6698. doi:10.1093/nar/gkz449

Jolma, A., Yin, Y., Nitta, K. R., Dave, K., Popov, A., Taipale, M., . . . Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature, 527*(7578), 384-388. doi:10.1038/nature15518

Jones, D. L., Brewster, R. C., & Phillips, R. (2014). Promoter architecture dictates cell-to-cell variability in gene expression. *Science, 346*(6216), 1533-1536. doi:10.1126/science.1255301

Josaitis, C. A., Gaal, T., & Gourse, R. L. (1995). Stringent control and growth-rate-dependent control have nonidentical promoter sequence requirements. *Proc Natl Acad Sci U S A, 92*(4), 1117-1121. doi:10.1073/pnas.92.4.1117

Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet, 6*(6), 451-464. doi:10.1038/nrg1615

Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., . . . Alon, U. (2001). Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science, 292*(5524), 2080-2083. doi:10.1126/science.1058758

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., & Ebright, R. H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science, 314*(5802), 1144-1147. doi:10.1126/science.1131399

Keseler, I. M., Gama-Castro, S., Mackie, A., Billington, R., Bonavides-Martínez, C., Caspi, R., . . . Karp, P. D. (2021). The EcoCyc Database in 2021. *Front Microbiol, 12*, 711077. doi:10.3389/fmicb.2021.711077

Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N. C., Joung, J. K., & Collins, J. J. (2012). A synthetic biology framework for programming eukaryotic transcription functions. *Cell, 150*(3), 647-658. doi:10.1016/j.cell.2012.05.045

Kinney, J. B., Murugan, A., Callan, C. G., Jr., & Cox, E. C. (2010). Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci U S A, 107*(20), 9158-9163. doi:10.1073/pnas.1004290107

Kleinschmidt, C., Tovar, K., Hillen, W., & Porschke, D. (1988). Dynamics of repressor-operator recognition: the Tn10-encoded tetracycline resistance control. *Biochemistry, 27*(4), 1094-1104. doi:10.1021/bi00404a003

Kominek, J., Doering, D. T., Opulente, D. A., Shen, X. X., Zhou, X., DeVirgilio, J., . . . Hittinger, C. T. (2019). Eukaryotic Acquisition of a Bacterial Operon. *Cell, 176*(6), 1356-1366.e1310. doi:10.1016/j.cell.2019.01.034

LaFleur, T. L., Hossain, A., & Salis, H. M. (2022). Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat Commun, 13*(1), 5159. doi:10.1038/s41467-022-32829-5

Lanzer, M., & Bujard, H. (1988). Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A, 85*(23), 8973-8977. doi:10.1073/pnas.85.23.8973

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., . . . Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. *Nature, 565*(7738), 251-254. doi:10.1038/s41586-018-0836-1

Lawson, C. L., Swigon, D., Murakami, K. S., Darst, S. A., Berman, H. M., & Ebright, R. H. (2004). Catabolite activator protein: DNA binding and transcription activation. *Curr Opin Struct Biol, 14*(1), 10-20. doi:10.1016/j.sbi.2004.01.012

Lawson, R. C., Jr., & York, S. S. (1987). Stoichiometry of lac repressor binding to nonspecific DNA: three different complexes form. *Biochemistry, 26*(15), 4867-4875. doi:10.1021/bi00389a039

Lee, T. H., & Maheshri, N. (2012). A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol Syst Biol, 8*, 576. doi:10.1038/msb.2012.7

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., . . . Young, R. A. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science, 298*(5594), 799-804. doi:10.1126/science.1075090

Lehming, N., Sartorius, J., Niemöller, M., Genenger, G., v Wilcken-Bergmann, B., & Müller-Hill, B. (1987). The interaction of the recognition helix of lac repressor with lac operator. *Embo j, 6*(10), 3145-3153. doi:10.1002/j.1460-2075.1987.tb02625.x

Levo, M., & Segal, E. (2014). In pursuit of design principles of regulatory sequences. *Nat Rev Genet, 15*(7), 453-468. doi:10.1038/nrg3684

Lewis, D. E., & Adhya, S. (2015). Molecular Mechanisms of Transcription Initiation at gal Promoters and their Multi-Level Regulation by GalR, CRP and DNA Loop. *Biomolecules, 5*(4), 2782-2807. doi:10.3390/biom5042782

Li, H. S., Israni, D. V., Gagnon, K. A., Gan, K. A., Raymond, M. H., Sander, J. D., . . . Khalil, A. S. (2022). Multidimensional control of therapeutic human cell function with synthetic gene circuits. *Science, 378*(6625), 1227-1234. doi:10.1126/science.ade0156

Li, X. Y., & McClure, W. R. (1998). Characterization of the closed complex intermediate formed during transcription initiation by Escherichia coli RNA polymerase. *J Biol Chem, 273*(36), 23549-23557. doi:10.1074/jbc.273.36.23549

Lim, H. N., Lee, Y., & Hussein, R. (2011). Fundamental relationship between operon organization and gene expression. *Proc Natl Acad Sci U S A, 108*(26), 10626-10631. doi:10.1073/pnas.1105692108

Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., & Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry, 45*(22), 6873-6888. doi:10.1021/bi0602718

Liu, R., Liang, L., Freed, E. F., Choudhury, A., Eckert, C. A., & Gill, R. T. (2020). Engineering regulatory networks for complex phenotypes in E. coli. *Nat Commun, 11*(1), 4050. doi:10.1038/s41467-020-17721-4

Lutz, R., & Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res, 25*(6), 1203-1210. doi:10.1093/nar/25.6.1203

Ma, H. W., Buer, J., & Zeng, A. P. (2004). Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics, 5*, 199. doi:10.1186/1471-2105-5-199

Maamar, H., Raj, A., & Dubnau, D. (2007). Noise in gene expression determines cell fate in Bacillus subtilis. *Science, 317*(5837), 526-529. doi:10.1126/science.1140818

MacArthur, B. D., Sevilla, A., Lenz, M., Müller, F. J., Schuldt, B. M., Schuppert, A. A., . . . Lemischka, I. R. (2012). Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. *Nat Cell Biol, 14*(11), 1139-1147. doi:10.1038/ncb2603

Macneil, L. T., & Walhout, A. J. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res, 21*(5), 645-657. doi:10.1101/gr.097378.109

Madan Babu, M., & Teichmann, S. A. (2003). Functional determinants of transcription factors in Escherichia coli: protein families and binding sites. *Trends Genet, 19*(2), 75-79. doi:10.1016/s0168-9525(02)00039-2

Mangan, S., & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A, 100*(21), 11980-11985. doi:10.1073/pnas.2133841100

Martínez-Antonio, A., Janga, S. C., & Thieffry, D. (2008). Functional organisation of Escherichia coli transcriptional regulatory network. *J Mol Biol, 381*(1), 238-247. doi:10.1016/j.jmb.2008.05.054

Martinez-Corral, R., Park, M., Biette, K. M., Friedrich, D., Scholes, C., Khalil, A. S., . . . DePace, A. H. (2023). Transcriptional kinetic synergy: A complex landscape revealed by integrating modeling and synthetic biology. *Cell Syst, 14*(4), 324-339.e327. doi:10.1016/j.cels.2023.02.003

McAdams, H. H., & Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A, 94*(3), 814-819. doi:10.1073/pnas.94.3.814

Mejía-Almonte, C., Busby, S. J. W., Wade, J. T., van Helden, J., Arkin, A. P., Stormo, G. D., . . . Collado-Vides, J. (2020). Redefining fundamental concepts of transcription initiation in bacteria. *Nat Rev Genet, 21*(11), 699-714. doi:10.1038/s41576-020-0254-8

Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., & Albà, M. M. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics, 18*(2), 333-334. doi:10.1093/bioinformatics/18.2.333

Miano, A., Rychel, K., Lezia, A., Sastry, A., Palsson, B., & Hasty, J. (2023). High-resolution temporal profiling of E. coli transcriptional response. *Nat Commun, 14*(1), 7606. doi:10.1038/s41467-023-43173-7

Minezaki, Y., Homma, K., & Nishikawa, K. (2005). Genome-wide survey of transcription factors in prokaryotes reveals many bacteria-specific families not found in archaea. *DNA Res, 12*(5), 269-280. doi:10.1093/dnares/dsi016

Molina, N., Suter, D. M., Cannavo, R., Zoller, B., Gotic, I., & Naef, F. (2013). Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proc Natl Acad Sci U S A, 110*(51), 20563-20568. doi:10.1073/pnas.1312310110

Monsalve, M., Calles, B., Mencía, M., Salas, M., & Rojo, F. (1997). Transcription activation or repression by phage psi 29 protein p4 depends on the strength of the RNA polymerase-promoter interactions. *Mol Cell, 1*(1), 99-107. doi:10.1016/s1097-2765(00)80011-8

Monsalve, M., Mencia, M., Rojo, F., & Salas, M. (1996). Activation and repression of transcription at two different phage phi29 promoters are mediated by interaction of the same residues of regulatory protein p4 with RNA polymerase. *Embo j, 15*(2), 383-391.

Moore, M., Klug, A., & Choo, Y. (2001). Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci U S A, 98*(4), 1437-1441. doi:10.1073/pnas.98.4.1437

Morgunova, E., & Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol, 47*, 1-8. doi:10.1016/j.sbi.2017.03.006

Moses, V., & Prevost, C. (1966). Catabolite repression of beta-galactosidase synthesis in Escherichia coli. *Biochem J, 100*(2), 336-353. doi:10.1042/bj1000336

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., & Darst, S. A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science, 296*(5571), 1285-1290. doi:10.1126/science.1069595

Naqvi, S., Kim, S., Hoskens, H., Matthews, H. S., Spritz, R. A., Klein, O. D., . . . Wysocka, J. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nat Genet, 55*(5), 841-851. doi:10.1038/s41588-023-01366-2

Newby, G. A., Kiriakov, S., Hallacli, E., Kayatekin, C., Tsvetkov, P., Mancuso, C. P., . . . Khalil, A. S. (2017). A Genetic Tool to Track Protein Aggregates and Control Prion Inheritance. *Cell, 171*(4), 966-979.e918. doi:10.1016/j.cell.2017.09.041

Niland, P., Hühne, R., & Müller-Hill, B. (1996). How AraC interacts specifically with its target DNAs. *J Mol Biol, 264*(4), 667-674. doi:10.1006/jmbi.1996.0668

Okuda, S., Kawashima, S., Kobayashi, K., Ogasawara, N., Kanehisa, M., & Goto, S. (2007). Characterization of relationships between transcriptional units and operon structures in Bacillus subtilis and Escherichia coli. *BMC Genomics, 8*, 48. doi:10.1186/1471-2164-8-48

Omelchenko, M. V., Makarova, K. S., Wolf, Y. I., Rogozin, I. B., & Koonin, E. V. (2003). Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. *Genome Biol, 4*(9), R55. doi:10.1186/gb-2003-4-9-r55

Omont, N., & Képès, F. (2004). Transcription/replication collisions cause bacterial transcription units to be longer on the leading strand of replication. *Bioinformatics, 20*(16), 2719-2725. doi:10.1093/bioinformatics/bth317

Osbourn, A. E., & Field, B. (2009). Operons. *Cell Mol Life Sci, 66*(23), 3755-3775. doi:10.1007/s00018-009-0114-3

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat Genet, 31*(1), 69-73. doi:10.1038/ng869

Paget, M. S. (2015). Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules, 5*(3), 1245-1265. doi:10.3390/biom5031245

Pandelakis, M., Delgado, E., & Ebrahimkhani, M. R. (2020). CRISPR-Based Synthetic Transcription Factors In Vivo: The Future of Therapeutic Cellular Programming. *Cell Syst, 10*(1), 1-14. doi:10.1016/j.cels.2019.10.003

Parisutham, V., Chhabra, S., Ali, M. Z., & Brewster, R. C. (2022). Tunable transcription factor library for robust quantification of regulatory properties in Escherichia coli. *Mol Syst Biol, 18*(6), e10843. doi:10.15252/msb.202110843

Patange, S., Girvan, M., & Larson, D. R. (2018). Single-cell systems biology: probing the basic unit of information flow. *Curr Opin Syst Biol, 8*, 7-15. doi:10.1016/j.coisb.2017.11.011

Paulsson, J. (2004). Summing up the noise in gene networks. *Nature, 427*(6973), 415-418. doi:10.1038/nature02257

Pedraza, J. M., & van Oudenaarden, A. (2005). Noise propagation in gene networks. *Science, 307*(5717), 1965-1969. doi:10.1126/science.1109090

Perez-Rueda, E., Hernandez-Guerrero, R., Martinez-Nuñez, M. A., Armenta-Medina, D., Sanchez, I., & Ibarra, J. A. (2018). Abundance, diversity and domain architecture variability in prokaryotic DNA-binding transcription factors. *PLoS One, 13*(4), e0195332. doi:10.1371/journal.pone.0195332

Piatek, A., & Mahfouz, M. M. (2017). Targeted genome regulation via synthetic programmable transcriptional regulators. *Crit Rev Biotechnol, 37*(4), 429-440. doi:10.3109/07388551.2016.1165180

Pinto, C., Melo-Miranda, R., Gordo, I., & Sousa, A. (2021). The Selective Advantage of the lac Operon for Escherichia coli Is Conditional on Diet and Microbiota Composition. *Front Microbiol, 12*, 709259. doi:10.3389/fmicb.2021.709259

Poss, Z. C., Ebmeier, C. C., & Taatjes, D. J. (2013). The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol, 48*(6), 575-608. doi:10.3109/10409238.2013.840259

Price, M. N., Arkin, A. P., & Alm, E. J. (2006). The life-cycle of operons. *PLoS Genet, 2*(6), e96. doi:10.1371/journal.pgen.0020096

Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature, 335*(6192), 683-689. doi:10.1038/335683a0

Raivio, T. L., & Silhavy, T. J. (1997). Transduction of envelope stress in Escherichia coli by the Cpx two-component system. *J Bacteriol, 179*(24), 7724-7733. doi:10.1128/jb.179.24.7724-7733.1997

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol, 4*(10), e309. doi:10.1371/journal.pbio.0040309

Raj, A., & van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell, 135*(2), 216-226. doi:10.1016/j.cell.2008.09.050

Raser, J. M., & O'Shea, E. K. (2004). Control of stochasticity in eukaryotic gene expression. *Science, 304*(5678), 1811-1814. doi:10.1126/science.1098641

Reiter, F., Wienerroither, S., & Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev, 43*, 73-81. doi:10.1016/j.gde.2016.12.007

Revyakin, A., Liu, C., Ebright, R. H., & Strick, T. R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science, 314*(5802), 1139-1143. doi:10.1126/science.1131398

Rigali, S., Derouaux, A., Giannotta, F., & Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem, 277*(15), 12507-12515. doi:10.1074/jbc.M110968200

Rocha, E. P., & Danchin, A. (2003). Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet, 34*(4), 377-378. doi:10.1038/ng1209

Rodionov, D. A., Mironov, A. A., Rakhmaninova, A. B., & Gelfand, M. S. (2000). Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria. *Mol Microbiol, 38*(4), 673-683. doi:10.1046/j.1365-2958.2000.02115.x

Rojo, F., Mencía, M., Monsalve, M., & Salas, M. (1998). Transcription activation and repression by interaction of a regulator with the alpha subunit of RNA polymerase: the model of phage phi 29 protein p4. *Prog Nucleic Acid Res Mol Biol, 60*, 29-46. doi:10.1016/s0079-6603(08)60888-0

Rosenfeld, N., Elowitz, M. B., & Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *J Mol Biol, 323*(5), 785-793. doi:10.1016/s0022-2836(02)00994-4

Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., & Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science, 307*(5717), 1962-1965. doi:10.1126/science.1106914

Rosinski, J. A., & Atchley, W. R. (1999). Molecular evolution of helix-turn-helix proteins. *J Mol Evol, 49*(3), 301-309. doi:10.1007/pl00006552

Roy, S., Semsey, S., Liu, M., Gussin, G. N., & Adhya, S. (2004). GalR represses galP1 by inhibiting the rate-determining open complex formation through RNA polymerase contact: a GalR negative control mutant. *J Mol Biol, 344*(3), 609-618. doi:10.1016/j.jmb.2004.09.070

Rubinstein, N. D., Zeevi, D., Oren, Y., Segal, G., & Pupko, T. (2011). The operonic location of auto-transcriptional repressors is highly conserved in bacteria. *Mol Biol Evol, 28*(12), 3309-3318. doi:10.1093/molbev/msr163

Salgado, H., Gama-Castro, S., Lara, P., Mejia-Almonte, C., Alarcón-Carranza, G., López-Almazo, A. G., . . . Collado-Vides, J. (2023). RegulonDB v12.0: a comprehensive resource of transcriptional regulation in E. coli K-12. *Nucleic Acids Res*. doi:10.1093/nar/gkad1072

Sanchez, A., Choubey, S., & Kondev, J. (2013). Stochastic models of transcription: from single molecules to single cells. *Methods, 62*(1), 13-25. doi:10.1016/j.ymeth.2013.03.026

Sastry, A. V., Gao, Y., Szubin, R., Hefner, Y., Xu, S., Kim, D., . . . Palsson, B. O. (2019). The Escherichia coli transcriptome mostly consists of independently regulated modules. *Nat Commun, 10*(1), 5536. doi:10.1038/s41467-019-13483-w

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol, 47*, 597-626. doi:10.1146/annurev.mi.47.100193.003121

Scholes, C., DePace, A. H., & Sánchez, Á. (2017). Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle. *Cell Syst, 4*(1), 97-108.e109. doi:10.1016/j.cels.2016.11.012

Schröder, O., & Wagner, R. (2000). The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. *J Mol Biol, 298*(5), 737-748. doi:10.1006/jmbi.2000.3708

Schultz, S. C., Shields, G. C., & Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science, 253*(5023), 1001-1007. doi:10.1126/science.1653449

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., . . . Widom, J. (2006). A genomic code for nucleosome positioning. *Nature, 442*(7104), 772-778. doi:10.1038/nature04979

Seungsoo, K., & Joanna, W. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. *Mol Cell, 83*(3), 373-392. Retrieved from https://www.sciencedirect.com/science/article/pii/S1097276522012151. doi:https://doi.org/10.1016/j.molcel.2022.12.032

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., . . . Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol, 30*(6), 521-530. doi:10.1038/nbt.2205

Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli.

Shih, Y. L., & Rothfield, L. (2006). The bacterial cytoskeleton. *Microbiol Mol Biol Rev, 70*(3), 729-754. doi:10.1128/mmbr.00017-06

Shimada, T., Ogasawara, H., Kobayashi, I., Kobayashi, N., & Ishihama, A. (2021). Single-Target Regulators Constitute the Minority Group of Transcription Factors in Escherichia coli K-12. *Front Microbiol, 12*, 697803. doi:10.3389/fmicb.2021.697803

Siebenlist, U., Simpson, R. B., & Gilbert, W. (1980). E. coli RNA polymerase interacts homologously with two different promoters. *Cell, 20*(2), 269-281. doi:10.1016/0092-8674(80)90613-3

Simpson, R. B. (1979). The molecular topography of RNA polymerase-promoter interaction. *Cell, 18*(2), 277-285. doi:10.1016/0092-8674(79)90047-3

So, L. H., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R., & Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. *Nat Genet, 43*(6), 554-560. doi:10.1038/ng.821

Stefan, K., Zhongge, Z., & Terence, H. (2009). Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell, 139*(7), 1366-1375. Retrieved from https://www.sciencedirect.com/science/article/pii/S0092867409015050. doi:https://doi.org/10.1016/j.cell.2009.12.001

Stormo, G. D. (2000). DNA binding sites: representation and discovery. *Bioinformatics, 16*(1), 16-23. doi:10.1093/bioinformatics/16.1.16

Stülke, J., & Hillen, W. (1999). Carbon catabolite repression in bacteria. *Curr Opin Microbiol, 2*(2), 195-201. doi:10.1016/s1369-5274(99)80034-4

Suvorova, I. A., Tutukina, M. N., Ravcheev, D. A., Rodionov, D. A., Ozoline, O. N., & Gelfand, M. S. (2011). Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria. *J Bacteriol, 193*(15), 3956-3963. doi:10.1128/jb.00277-11

Swank, Z., Laohakunakorn, N., & Maerkl, S. J. (2019). Cell-free gene-regulatory network engineering with synthetic transcription factors. *Proc Natl Acad Sci U S A, 116*(13), 5892-5901. doi:10.1073/pnas.1816591116

Swapna, G., Kumari, V., & Nagaraja, V. (2015). Different Modes of Transactivation of Bacteriophage Mu Late Promoters by Transcription Factor C. *PLoS One, 10*(6), e0129504. doi:10.1371/journal.pone.0129504

Takada, H., Kijima, K., Ishiguro, A., Ishihama, A., & Shimada, T. (2023). Genomic SELEX Reveals Pervasive Role of the Flagella Master Regulator FlhDC in Carbon Metabolism. *Int J Mol Sci, 24*(4). doi:10.3390/ijms24043696

Tan, D., Chen, R., Mo, Y., Gu, S., Ma, J., Xu, W., . . . Huang, W. (2021). Quantitative control of noise in mammalian gene expression by dynamic histone regulation. *Elife, 10*. doi:10.7554/eLife.65654

Tang, G. Q., Roy, R., Bandwar, R. P., Ha, T., & Patel, S. S. (2009). Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proc Natl Acad Sci U S A, 106*(52), 22175-22180. doi:10.1073/pnas.0906979106

Thieffry, D., & Romero, D. (1999). The modularity of biological regulatory networks. *Biosystems, 50*(1), 49-59. doi:10.1016/s0303-2647(98)00087-2

Utley, R. T., Ikeda, K., Grant, P. A., Côté, J., Steger, D. J., Eberharter, A., . . . Workman, J. L. (1998). Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature, 394*(6692), 498-502. doi:10.1038/28886

van Dijk, D., Sharon, E., Lotan-Pompan, M., Weinberger, A., Segal, E., & Carey, L. B. (2017). Large-scale mapping of gene regulatory logic reveals context-dependent repression by transcriptional activators. *Genome Res, 27*(1), 87-94. doi:10.1101/gr.212316.116

van Hijum, S. A., Medema, M. H., & Kuipers, O. P. (2009). Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol Mol Biol Rev, 73*(3), 481-509, Table of Contents. doi:10.1128/mmbr.00037-08

Vilar, J. M., & Leibler, S. (2003). DNA looping and physical constraints on transcription regulation. *J Mol Biol, 331*(5), 981-989. doi:10.1016/s0022-2836(03)00764-2

Vlková, M., & Silander, O. K. (2022). Gene regulation in Escherichia coli is commonly selected for both high plasticity and low noise. *Nat Ecol Evol, 6*(8), 1165-1179. doi:10.1038/s41559-022-01783-2

Wade, J. T., Castro Roa, D., Grainger, D. C., Hurd, D., Busby, S. J., Struhl, K., & Nudler, E. (2006). Extensive functional overlap between sigma factors in Escherichia coli. *Nat Struct Mol Biol, 13*(9), 806-814. doi:10.1038/nsmb1130

Wanner, B. L., Kodaira, R., & Neidhardt, F. C. (1978). Regulation of lac operon expression: reappraisal of the theory of catabolite repression. *J Bacteriol, 136*(3), 947-954. doi:10.1128/jb.136.3.947-954.1978

Wasserman, W. W., & Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet, 5*(4), 276-287. doi:10.1038/nrg1315

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., . . . Hughes, T. R. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell, 158*(6), 1431-1443. doi:10.1016/j.cell.2014.08.009

Wilson, C. J., Zhan, H., Swint-Kruse, L., & Matthews, K. S. (2007). The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci, 64*(1), 3-16. doi:10.1007/s00018-006-6296-z

Yankulov, K., Blau, J., Purton, T., Roberts, S., & Bentley, D. L. (1994). Transcriptional elongation by RNA polymerase II is stimulated by transactivators. *Cell, 77*(5), 749-759. doi:10.1016/0092-8674(94)90058-2

Zaslaver, A., Mayo, A. E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., . . . Alon, U. (2004). Just-in-time transcription program in metabolic pathways. *Nat Genet, 36*(5), 486-491. doi:10.1038/ng1348

Zhao, S., Hong, C. K. Y., Myers, C. A., Granas, D. M., White, M. A., Corbo, J. C., & Cohen, B. A. (2023). A single-cell massively parallel reporter assay detects cell-type-specific gene regulation. *Nat Genet, 55*(2), 346-354. doi:10.1038/s41588-022-01278-7