The Molecular Biophysics of Evolutionary and Physiological Adaptation

Thesis by
Griffin Daniel Chure

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
Defended May 1, 2020

Griffin Daniel Chure
ORCID: 0000-0002-2216-2057

ACKNOWLEDGEMENTS

This is a test acknowledgements file

ABSTRACT

This is a test abstract file.

PUBLISHED CONTENT AND CONTRIBUTIONS

Chure, G., Kaczmarek, Z.A., and Phillips, R. (2019b). Physiological Adaptability and Parametric Versatility in a Simple Genetic Circuit. BioRxiv 2019.12.19.878462.

Hirokawa, S., Chure, G., Belliveau, N.M., Lovely, G.A., Anaya, M., Schatz, D.G., Baltimore, D., and Phillips, R. (2019). Sequence-Dependent Dynamics of Synthetic and Endogenous RSSs in V(D)J Recombination. BioRxiv 791954.

Chure, G., Razo-Mejia, M., Belliveau, N.M., Einav, T., Kaczmarek, Z.A., Barnes, S.L., and Phillips, R. (2019a). Predictive Shifts in Free Energy Couple Mutations to Their Phenotypic Consequences. Proceedings of the National Academy of Sciences 116.

Phillips, R., Belliveau, N.M., Chure, G., Garcia, H.G., Razo-Mejia, M., and Scholes, C. (2019). Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. Annu. Rev. Biophys. 48, 121–163.

Chure, G.* , Lee, H.J.* , Rasmussen, A., and Phillips, R. (2018). Connecting the Dots between Mechanosensitive Channel Abundance, Osmotic Shock, and Survival at Single-Cell Resolution. Journal of Bacteriology 200. * contributed equally

Razo-Mejia, M.* , Barnes, S.L.* , Belliveau, N.M.* , Chure, G.* , Einav, T.* , Lewis, M., and Phillips, R. (2018). Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. Cell Systems 6, 456-469.e10. * contributed equally

TABLE OF CONTENTS

Acknowledgements iii
Abstract iv
Published Content and Contributions v
Table of Contents vi
List of Illustrations vii
List of Tables viii
Chapter I: Through the Intramolecular Grape Vine: Signal Processing Via Allosteric Transcription Factors 1
1.1 Introduction 1
1.2 Results 3
References 8
Chapter II: Questionnaire 9
Chapter III: Consent Form 10

LIST OF ILLUSTRATIONS

Number *Page*

1.1 This is the short cap 4

1.2 States and weights for the simple repression motif. 6

LIST OF TABLES

Number *Page*

Chapter 1

THROUGH THE INTRAMOLECULAR GRAPE VINE: SIGNAL PROCESSING VIA ALLOSTERIC TRANSCRIPTION FACTORS

A version of this chapter originally appeared as Razo-Mejia, M.* , Barnes, S.L.* , Belliveau, N.M.* , Chure, G.* , Einav, T.* , Lewis, M., and Phillips, R. (2018). Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. Cell Systems 6, 456-469.e10. M.R.M, S.L.B, N.M.B, G.C., and T.E. contributed equally to this work from the theoretical underpinnings to the experimental design and execution.

1.1 Introduction

Understanding how organisms sense and respond to changes in their environment has long been a central theme of biological inquiry. At the cellular level, this interaction is mediated by a diverse collection of molecular signaling pathways. A pervasive mechanism of signaling in these pathways is allosteric regulation, in which the binding of a ligand induces a conformational change in some target molecule, triggering a signaling cascade . One of the most important examples of such signaling is offered by transcriptional regulation, where a transcription factor's propensity to bind to DNA will be altered upon binding to an allosteric effector.

Despite allostery's ubiquity, we lack a formal, rigorous, and generalizable framework for studying its effects across the broad variety of contexts in which it appears. A key example of this is transcriptional regulation, in which allosteric transcription factors can be induced or corepressed by binding to a ligand. An allosteric transcription factor can adopt multiple conformational states, each of which has its own affinity for the ligand and for its DNA target site. *In vitro* studies have rigorously quantified the equilibria of different conformational states for allosteric transcription factors and measured the affinities of these states to the ligand . In spite of these experimental observations, the lack of a coherent quantitative model for allosteric transcriptional regulation has made it impossible to predict the behavior of even a simple genetic circuit across a range of regulatory parameters.

The ability to predict circuit behavior robustly— that is, across both broad ranges of parameters and regulatory architectures —is important for multiple reasons. First, in the context of a specific gene, accurate prediction demonstrates that all components relevant to the gene's behavior have been identified and characterized to sufficient quantitative precision. Second, in the context of genetic circuits in general, robust prediction validates the model that generated the prediction. Possessing a validated model also has implications for future work. For example, when we have sufficient confidence in the model, a single data set can be used to accurately extrapolate a system's behavior in other conditions. Moreover, there is an essential distinction between a predictive model, which is used to predict a system's behavior given a set of input variables, and a retroactive model, which is used to describe the behavior of data that has already been obtained. We note that even some of the most careful and rigorous analysis of transcriptional regulation often entails only a retroactive reflection on a single experiment. This raises the fear that each regulatory architecture may require a unique analysis that cannot carry over to other systems, a worry that is exacerbated by the prevalent use of phenomenological functions (e.g. Hill functions or ratios of polynomials) that can analyze a single data set but cannot be used to extrapolate a system's behavior in other conditions .

This work explores what happens when theory takes center stage, namely, we first write down the equations governing a system and describe its expected behavior across a wide array of experimental conditions, and only then do we set out to experimentally confirm these results. Building upon previous work and the work of Monod, Wyman, and Changeux , we present a statistical mechanical rendering of allostery in the context of induction and corepression (shown schematically in Fig. [1.1] (A) and henceforth referred to as the MWC model) and use it as the basis of parameter-free predictions which we then test experimentally. More specifically, we study the simple repression motif – a widespread bacterial genetic regulatory architecture in which binding of a transcription factor occludes binding of an RNA polymerase, thereby inhibiting transcription initiation. The MWC model stipulates that an allosteric protein fluctuates between two distinct conformations – an active and inactive state – in thermodynamic equilibrium . During induction, for example, effector binding increases the probability that a repressor will be in the inactive state, weakening its ability to bind

to the promoter and resulting in increased expression. To test the predictions of our model across a wide range of operator binding strengths and repressor copy numbers, we design an *E. coli* genetic construct in which the binding probability of a repressor regulates gene expression of a fluorescent reporter.

In total, the work presented here demonstrates that one extremely compact set of parameters can be applied self-consistently and predictively to different regulatory situations including simple repression on the chromosome, cases in which decoy binding sites for repressor are put on plasmids, cases in which multiple genes compete for the same regulatory machinery, cases involving multiple binding sites for repressor leading to DNA looping, and induction by signaling . Thus, rather than viewing the behavior of each circuit as giving rise to its own unique input-output response, the MWC model provides a means to characterize these seemingly diverse behaviors using a single unified framework governed by a small set of parameters.

1.2 Results

Characterizing Transcription Factor Induction using the Monod-Wyman-Changeux (MWC) Model

We begin by considering a simple repression genetic architecture in which the binding of an allosteric repressor occludes the binding of RNA polymerase (RNAP) to the DNA . When an effector (hereafter referred to as an "inducer" for the case of induction) binds to the repressor, it shifts the repressor's allosteric equilibrium towards the inactive state as specified by the MWC model . This causes the repressor to bind more weakly to the operator, which increases gene expression. Simple repression motifs in the absence of inducer have been previously characterized by an equilibrium model where the probability of each state of repressor and RNAP promoter occupancy is dictated by the Boltzmann distribution (we note that non-equilibrium models of simple repression have been shown to have the same functional form that we derive below). We extend these models to consider allostery by accounting for the equilibrium state of the repressor through the MWC model.

Thermodynamic models of gene expression begin by enumerating all possible states of the promoter and their corresponding statistical weights. As

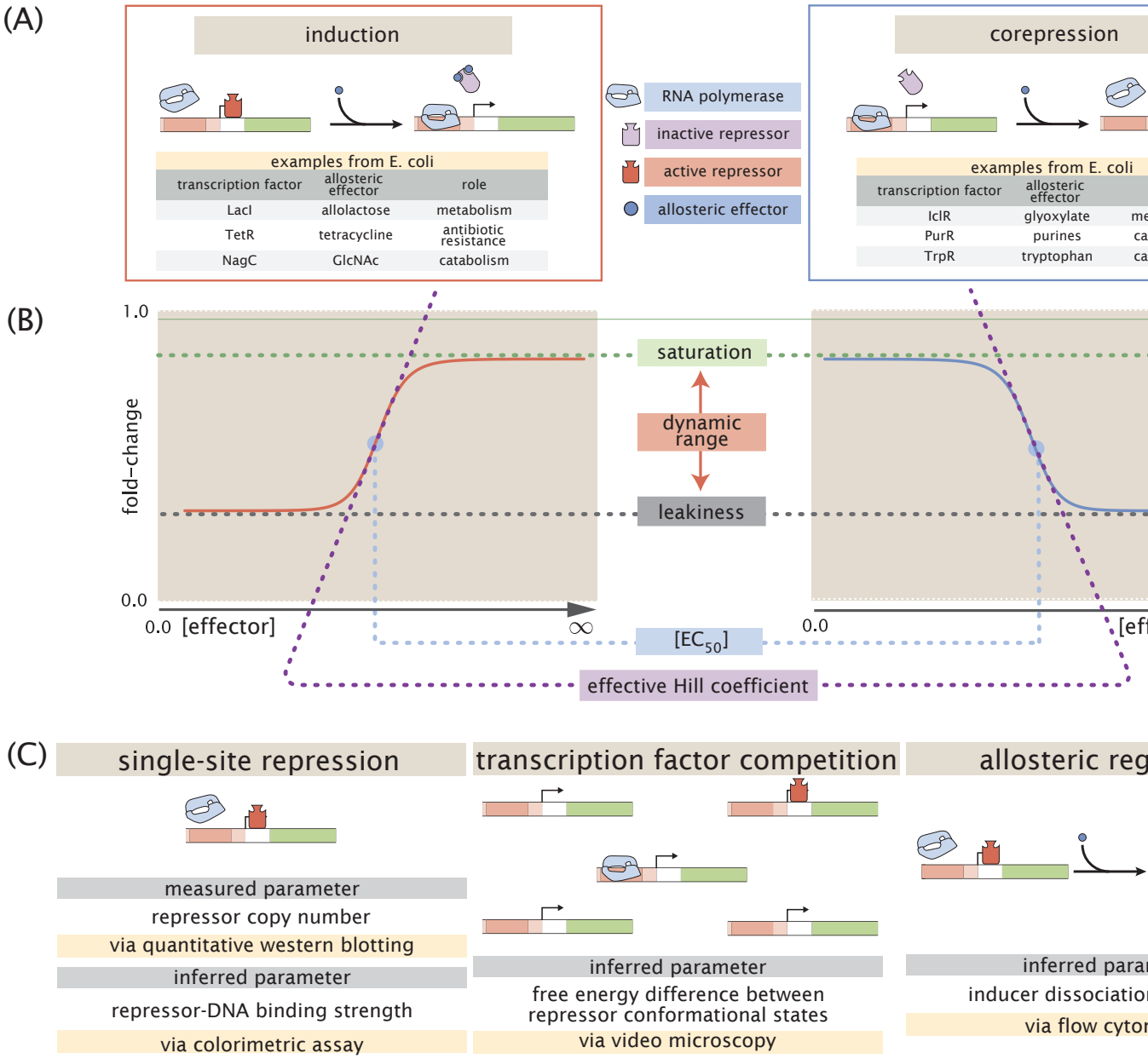

Figure 1.1: (A) We consider a promoter regulated solely by an allosteric repressor. When bound, the repressor prevents RNAP from binding and initiating transcription. Induction is characterized by the addition of an effector which binds to the repressor and stabilizes the inactive state (defined as the state which has a low affinity for DNA), thereby increasing gene expression. In corepression, the effector stabilizes the repressor's active state and thus further reduces gene expression. We list several characterized examples of induction and corepression that support different physiological roles in *E. coli* . (B) A schematic regulatory response of the two architectures shown in Panel (A) plotting the fold-change in gene expression as a function of effector concentration, where fold-change is defined as the ratio of gene expression in the presence versus the absence of repressor. We consider the following key phenotypic properties that describe each response curve: the minimum response (leakiness), the maximum response (saturation), the difference between the maximum and minimum response (dynamic range), the concentration of ligand which generates a fold-change halfway between the minimal and maximal response (EC_{50}), and the log-log slope at the mid-

shown in [1.2], the promoter can either be empty, occupied by RNAP, or occupied by either an active or inactive repressor. The probability of binding to the promoter will be affected by the protein copy number, which we denote as P for RNAP, R_A for active repressor, and R_I for inactive repressor. We note that repressors fluctuate between the active and inactive conformation in thermodynamic equilibrium, such that R_A and R_I will remain constant for a given inducer concentration . We assign the repressor a different DNA binding affinity in the active and inactive state. In addition to the specific binding sites at the promoter, we assume that there are N_{NS} non-specific binding sites elsewhere (i.e. on parts of the genome outside the simple repression architecture) where the RNAP or the repressor can bind. All specific binding energies are measured relative to the average non-specific binding energy. Thus, $\Delta\varepsilon_P$ represents the energy difference between the specific and non-specific binding for RNAP to the DNA. Likewise, $\Delta\varepsilon_{RA}$ and $\Delta\varepsilon_{RI}$ represent the difference in specific and non-specific binding energies for repressor in the active or inactive state, respectively.

Thermodynamic models of transcription posit that gene expression is proportional to the probability that the RNAP is bound to the promoter p_{bound}, which is given by

$$p_{\text{bound}} = \frac{\frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_P}}{1 + \frac{R_A}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}} + \frac{R_I}{N_{NS}} e^{-\beta\Delta\varepsilon_{RI}} + \frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_P}}, \tag{1.1}$$

with $\beta = \frac{1}{k_B T}$ where k_B is the Boltzmann constant and T is the temperature of the system. As $k_B T$ is the natural unit of energy at the molecular length scale, we treat the products $\beta\Delta\varepsilon_j$ as single parameters within our model. Measuring p_{bound} directly is fraught with experimental difficulties, as determining the exact proportionality between expression and p_{bound} is not straightforward. Instead, we measure the fold-change in gene expression due to the presence of the repressor. We define fold-change as the ratio of gene expression in the presence of repressor relative to expression in the absence of repressor (i.e. constitutive expression), namely,

$$\text{fold-change} \equiv \frac{p_{\text{bound}}(R > 0)}{p_{\text{bound}}(R = 0)}. \tag{1.2}$$

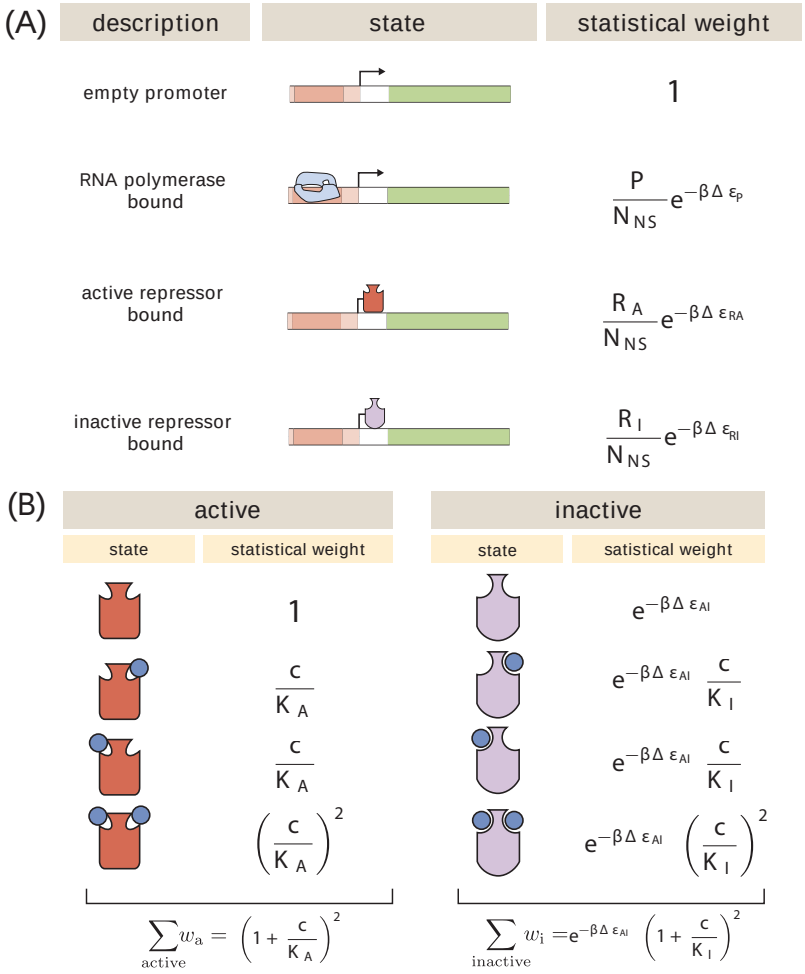

Figure 1.2: States and weights for the simple repression motif.** RNAP (light blue) and a repressor compete for binding to a promoter of interest. There are R_A repressors in the active state (red) and R_I repressors in the inactive state (purple). The difference in energy between a repressor bound to the promoter of interest versus another non-specific site elsewhere on the DNA equals $\Delta\varepsilon_{RA}$ in the active state and $\Delta\varepsilon_{RI}$ in the inactive state; the P RNAP have a corresponding energy difference $\Delta\varepsilon_P$ relative to non-specific binding on the DNA. N_{NS} represents the number of non-specific binding sites for both RNAP and repressor. A repressor has an active conformation (red, left column) and an inactive conformation (purple, right column), with the energy difference between these two states given by $\Delta\varepsilon_{AI}$. The inducer (blue circle) at concentration c is capable of binding to the repressor with dissociation constants K_A in the active state and K_I in the inactive state. The eight states for a dimer with $n = 2$ inducer binding sites are shown along with the sums of the active and inactive states.

We can simplify this expression using two well-justified approximations: (1) $\frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_P} \ll 1$ implying that the RNAP binds weakly to the promoter ($N_{NS} = 4.6 \times 10^6$, $P \approx 10^3$, $\Delta \varepsilon_P \approx -2$ to $-5\, k_B T$, so that $\frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_P} \approx 0.01$) and (2) $\frac{R_I}{N_{NS}} e^{-\beta \Delta \varepsilon_{RI}} \ll 1 + \frac{R_A}{N_{NS}} e^{-\beta \Delta \varepsilon_{RA}}$ which reflects our assumption that the inactive repressor binds weakly to the promoter of interest. Using these approximations, the fold-change reduces to the form

$$\text{fold-change} \approx \left(1 + \frac{R_A}{N_{NS}} e^{-\beta \Delta \varepsilon_{RA}}\right)^{-1} \equiv \left(1 + p_A(c) \frac{R}{N_{NS}} e^{-\beta \Delta \varepsilon_{RA}}\right)^{-1}, \tag{1.3}$$

where in the last step we have introduced the fraction $p_{\text{act}}(c)$ of repressors in the active state given a concentration c of inducer, such that $R_A(c) = p_{\text{act}}(c) R$. Since inducer binding shifts the repressors from the active to the inactive state, $p_{\text{act}}(c)$ grows smaller as c increases.

REFERENCES

Chapter 2

QUESTIONNAIRE

Chapter 3

CONSENT FORM