ON THE NATURE OF GENE REGULATORY DESIGN – THE BIOPHYSICS OF TRANSCRIPTION FACTOR BINDING SHAPES GENE REGULATION

by

Claudia Igler

May, 2019

A thesis presented to the
Graduate School
of the
Institute of Science and Technology Austria, Klosterneuburg, Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

The dissertation of Claudia Igler, titled *On the Nature of Gene Regulatory Design - The Biophysics of Transcription Factor Binding Shapes Gene Regulation*, is approved by:

Supervisor: Călin Guet, IST Austria, Klosterneuburg, Austria

Signature: _____

Co-Supervisor: Jon P. Bollback, University of Liverpool, Liverpool, UK

Signature: _____

Committee Member: Gasper Tkačik, IST Austria, Klosterneuburg, Austria

Signature: _____

Committee Member: Stephen T. Abedon, The Ohio State University, Columbus (Ohio), USA

Signature: _____

Exam Chair: Nick Barton, IST Austria, Klosterneuburg, Austria

Signature: _____

IST Austria Thesis, ISSN: 2663-337X

I hereby declare that this dissertation is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Claudia Igler

May 3, 2019

Abstract

Decades of studies have revealed the mechanisms of gene regulation in molecular detail. We make use of such well-described regulatory systems to explore how the molecular mechanisms of protein-protein and protein-DNA interactions shape the dynamics and evolution of gene regulation.

i) We uncover how the biophysics of protein-DNA binding determines the potential of regulatory networks to evolve and adapt, which can be captured using a simple mathematical model.
ii) The evolution of regulatory connections can lead to a significant amount of crosstalk between binding proteins. We explore the effect of crosstalk on gene expression from a target promoter, which seems to be modulated through binding competition at non-specific DNA sites.
iii) We investigate how the very same biophysical characteristics as in i) can generate significant fitness costs for cells through global crosstalk, meaning non-specific DNA binding across the genomic background.
iv) Binding competition between proteins at a target promoter is a prevailing regulatory feature due to the prevalence of co-regulation at bacterial promoters. However, the dynamics of these systems are not always straightforward to determine even if the molecular mechanisms of regulation are known. A detailed model of the biophysical interactions reveals that interference between the regulatory proteins can constitute a new, generic form of system memory that records the history of the input signals at the promoter.

We demonstrate how the biophysics of protein-DNA binding can be harnessed to investigate the principles that shape and ultimately limit cellular gene regulation. These results provide a basis for studies of higher-level functionality, which arises from the underlying regulation.

Acknowledgments

In order to be successful in life one needs luck and hard work; even though my time as a PhD student consisted of quite a bit of the latter, I was also incredibly lucky to have a group of truly amazing people to guide me along the way, work with me or simply support me.
I especially want to thank Călin Guet for helping me discover my passion for basic research (and the best organisms to use for it), as well as for shaping the way I think about science and how to conduct it, by being an admirable example and an amazing supervisor. I am immensely grateful for his patience, guidance and care, which have quite significantly contributed to where I am today. He encouraged me to push my boundaries and be critical of my work, but not too much of myself.
As the second keeper of my sanity and spirit, I want to thank Mato Lagator, who always provided me with advice, support and friendship, whenever I needed it. No one could wish for a better collaborator, one driven by honest scientific curiosity and always willing to lend an open ear.
Third, Gasper Tkačik has shown me what it means to have a truly interdisciplinary spirit, to perform meticulous modeling and when to use models in a meaningful way. It was a hard and frustrating learning experience, but one I highly appreciate nonetheless.
There are quite a number of other people who helped me find my way around an experimental lab, bacteria, phages, genetics or similar topics and I am grateful to all of them for their time and patience. Most notably among them I want to mention Jon Bollback, Steve Abedon, Diarmaid Hughes, Beth Lazazzera, Fitnat Yildiz, John Marko, Bella Tomanek, Rok Grah, Anna Staron, Ali Sezgin, Kathrin Tomasek, Maros Pleska, Tobi Bergmiller, Moritz Lang, Remy Chait, Torsten Waldminghaus and Florian Pauler.

Not least of all, I also want to thank my friends and family whom I have neglected during my scientific quest and who still supported me and listened to everything, even if I suspect that they often could not understand what I was saying. I will never be able to express how much I owe to my greatest supporter and invaluable source of strength and love and laughter, my bestie Sophie Diem; I will always be deeply grateful. Daniel Čapek, my cherished BFF, never failed to cheer me up and distract me from my frustrations. Claire Fourcade showed me how to blossom, no matter how hard the winter, and shared all my ups and downs. My twin brother Clemens and my sister-in-law Anna-Helene have always believed in me and wanted me to be happy above all else; always providing me with a reliable feeling of support. My parents and my grandmother never quite understood my passion for science, but they were there with me every step of the way regardless, and they helped me to become the person I am today - I am lucky to have them in my life. I want to thank Andi, Eike, Carola and Werner for their love and encouragement, which have enriched my life in many ways. Thank you Andi for believing in me in a way that I myself could not.

About the Author

Claudia Igler completed a BSc in Mathematics and an MSc in Biomedical Engineering at the Vienna University of Technology before joining IST Austria in September 2013.
Her main research interest lies in applying mathematical tools to biologically and medically relevant questions. During her time at University she joined a research project on 'Detecting epileptic seizures in EEG signals' at the Austrian Institute of Technology and helped to improve an algorithm on spike detection in EEG signals. During her PhD studies, Claudia has taken part in workshops on *Advanced Bacterial Genetics* (New York, 2015) and *Computational Synthetic Biology* (San Diego, 2016), and presented her research results at the *Population Genetics Meeting* in Edinburgh in 2015. In September 2018, she published the results of her main PhD project in the journal *Nature Ecology & Evolution*.

List of Publications Appearing in Thesis

Igler, C., Lagator, M., Tkačik, G., Bollback, J. P. & Guet, C. C. Evolutionary potential of transcription factors for gene regulatory rewiring. *Nat. Ecol. Evol.* **2,** (2018).

Table of Contents

List of Figures

List of Tables

List of Symbols/Abbreviations

DNA	deoxyribonucleic acid
mRNA	messenger RNA
tRNA	transfer RNA
E. coli	*Escherichia coli*
S. enterica	*Salmonella enterica*
TF	Transcription factor
RNAP	RNA polymerase
GFP	green fluorescent protein
bp	base pair
GRN	gene regulatory network
OD_{600}	optical density at 600nm
AU	arbitrary units
RFU	relative fluorescence units
NFU	normalized fluorescence units
aTc	anhydrotetracycline
IPTG	isopropyl β-D-1-thiogalactopyranoside
CAA	casamino acids
LB	lysogeny broth
FDR	false discovery rate
ANOVA	analysis of variance
ChIP	chromatin immunoprecipitation

1. Introduction

1.1 Gene regulation by transcription factors

The complexity and marvel of cellular life originates fundamentally from the regulation of the genes within a cell. It is gene regulation that allows genetically identical cells to develop varying shapes and functions - thereby also determining the complexity of an organism - and it is also gene regulation that governs if and how cells can respond to external signals. Based on some of the supposedly most simple organisms, François Jacob, André Lwoff, Jacques Monod and Mark Ptashne laid the foundation of gene regulatory research through the discoveries of transcriptional repressors in *E. coli*[1] and phage Lambda[2,3] over 50 years ago,. Their seminal work defined the individual units that ultimately determine the entire functioning of a cell. These transcriptional units consist of one or several genes under the control of an upstream promoter, which contains binding sites for RNA polymerase (RNAP) and transcriptional regulators, and is bounded by a downstream terminator[4]. The transcriptional regulators are proteins, which serve as receptors for signals from the external environment or the inside of the cell and transmit the signal information by changing their regulatory activity (usually through conformational changes). The presence or absence of regulatory proteins in a promoter region is then integrated in a manner resembling a logic unit to produce a specific transcription rate from the promoter.

Since these early discoveries the picture of how transcription factors (TFs) regulate gene expression in a positive (activators)[5] or negative (repressors) manner has become much clearer, but also much more complex[6]. The lac operon and the Lambda genetic switch are among the best characterized regulatory systems, but even those have revealed continuous surprises over the years[7,8]. Individual TFs can regulate gene expression by either hindering or helping RNAP binding or one of the following steps of transcription initiation, depending on their relative position in the promoter region: Repressors can hinder RNAP binding directly through steric occlusion at the promoter, or indirectly, obstructing the promoter through DNA looping or blocking activator binding. Activators on the other hand, can interact with RNAP in different ways (for example, by contacting the α-subunit or the σ-factor), or induce a conformational change in the promoter region[9]. Consequently, the mode of activation can differ from recruiting RNAP to its binding site to assisting the transition from closed to open complex formation, which precedes transcription initiation.

At least three-quarter of the promoters in *E. coli* are regulated by TFs and most of these promoter regions generally contain several binding sites for one or more TFs. Hence, signal integration in a promoter region is often a complex interaction between several players. TF binding of the same species can moreover be facilitated by binding cooperativity and DNA loop formation[10,11]. Binding sites for different TFs in close vicinity or even overlapping one another can lead to competition and steric hindrance between them[12,13], or speed up the dissociation from the binding site[14]. In addition,

there are also indirect mechanisms how regulatory proteins can affect transcription initiation from afar like induction of structural changes, polarity and strength changes in supercoiling[15,16] or transcriptional interference from another promoter[17].

Most of the players in transcriptional regulation (DNA, RNA, proteins) are present at only a few copies per cell, introducing stochasticity into the process of gene regulation, leading to a surprising amount of phenotypic variation in clonal cell populations[18]. While noise in gene expression can be detrimental for individual cells by preventing precisely tuned responses, surprisingly, this noise can also provide a fitness advantage to cell populations[19–22]. Examples of beneficial noise include the stochastic expression of stress-related genes and bet hedging, meaning that a part of the population displays a phenotype suboptimal for the current environment but potentially advantageous in a future environment[23]. So far, most observations of phenotypic variability in clonal cell populations have been linked to stochasticity in transcription and translation, but the underlying molecular mechanisms often remain unclear[23].

Overall, the picture of gene regulation that arose in the last 50+ years exhibits many more nuances than originally expected - even considering just the 'simplest' organisms. Although the molecular regulatory mechanisms have been elucidated in detail in isolated systems, it remains to be explored how those mechanisms shape the dynamics and evolution of complex regulatory systems.

1.2 From transcriptional units to networks

Individual transcriptional units generally do not work in isolation within their host cells but are connected into larger motifs, which are again interwoven to form large gene regulatory networks. Those networks determine cellular complexity, structure and functioning, much more so than the number of genes does[24]. While the picture of molecular interactions in individual units is becoming ever clearer, the emergent behavior of regulatory networks remains largely mysterious.

In order to simplify this daunting task, one approach is to break down the network into smaller building blocks, so-called network motifs[25]. One example of such motifs are autoregulatory circuits, where a TF controls its own expression directly or indirectly. However, even very simple genetic circuits generally do not allow for straightforward predictions of function just by looking at the wiring diagram, especially if they involve auto-regulation. Hence, even a limited number of components can result in a large diversity of complex behaviors depending on their connectivity and context[26]. This increasing complexity and unpredictability makes understanding gene regulatory networks a central quest in modern biology as it allows cells to respond and adapt to their environment in intricate ways.

A topic of specific interest in this context is the evolution of gene regulatory networks. Until now the main efforts have been put into finding global network properties that can describe their ability to physiologically and/or evolutionarily adapt. Two properties that have been discussed in depth in connection to network evolution are robustness and evolvability. One obstacle to unifying the results of different studies is the fact that both concepts have been defined in various ways and on various levels[27,28]. Accordingly, there has been some dispute on the apparent paradox that robustness (resistance to genetic change) has to be negatively correlated with evolvability (adaption in response to genetic change)[29]. This paradox has been addressed at the network level by different suggestions: i) network degeneracy: a partial overlap in the functioning of multi-functional components allows evolvability of a component while its degenerate counter-part still robustly fulfills its function[30]; or ii) the existence of neutral networks: genotype networks that code for the same phenotype make them robust but they also allow for a wide exploration of genotype space which makes them more evolvable as well, as they will eventually encounter new phenotypes[31,32]. However, a molecular understanding of these properties has not so far been accomplished.

Regardless of these specific definitions for robustness and evolvability, regulatory networks seem to be extremely plastic with regard to their regulatory connections[33,34], being able to tolerate a large amount of rewiring[35]. While several studies investigated the origin of gene duplications as catalyst for network rewiring[36,37], on a mechanistic level network adaptation has to occur through local rewiring of regulatory connections[38,39]. Again, analyses have been generally performed at the network level or – as that can be a quite complex task experimentally – through computational studies, while the molecular-level analyses have received little to no attention, especially from an experimental point of view.

1.3 Global cellular properties influence gene regulation

While most regulatory networks are generally studied in isolation from the rest of the cell, we should not ignore that the cellular machinery and genomic background can have a significant impact on local promoter regulation as well. Several studies have shown for example that global cellular properties affect gene expression levels due to changes in growth rate or macromolecular crowding[40,41].

The cellular cytoplasm is crowded with proteins and up to 50% of the DNA is occupied by non-specifically bound proteins[41]. Hence, macromolecular crowding and non-specific binding will both impact the DNA-target search of binding proteins[41–43]. TFs search for their binding targets through a combination of 3D-diffusion in the cytoplasm and 1D-sliding along the DNA[44], where the latter generally would speed up search times as compared to pure 3D diffusion, but also can be hindered by non-specific DNA binding[41,42]. Others suggest that non-specific binding does not affect the target search time significantly but rather the occupancy time at the target promoter[45,46]. Whether occupancy time is increased or decreased depends on the

mobility of the non-specifically bound protein (immobile or sliding)[45]. Immobile proteins on DNA can also lead to roadblocks for RNAP progression, thereby aborting the transcription process[47]. As roadblocks are dislodged with increasing efficiency for increasing TF dissociation rates, they might not form to a significant extent in the case of non-specific binding. Macromolecular crowding however can affect the association-dissociation equilibria of TFs at their target binding sites, which in simple regulatory systems can lead to increased or decreased gene expression (for activators and repressors, respectively)[46]. In more complicated networks, like the phage Lambda genetic switch, crowding results in highly non-linear changes in binding strength and impacts network performance in a counter-intuitive manner[46]. Macromolecular crowding was also found to have a larger impact on specific - than non-specific - DNA binding, decreasing 3D-diffusion but aiding 1D-search dynamics[43].

Another global cellular effect on gene regulation is determined by cell growth itself. Depending on the environmental conditions, growth rates can vary substantially, which leads to changes in cell size and macromolecular composition[48], as well as the number of chromosomes per cell[49,50]. Hence, growth rate affects gene expression indirectly by changing global cellular parameters like RNAP and ribosome abundance, gene copy number and protein dilution[40]. Increasing growth rates will lead to higher amounts of transcripts per cell for constitutively expressed genes but at the same time lower protein densities due to the strong increase in cell volume, for example. The effect of growth rate on gene expression depends on the details of promoter regulation, however, showing differences between activation and repression as well as for systems with cooperativity. Networks characterized by a bistable behavior can even exhibit qualitatively different behavior[40].

Gene regulation in bacterial cells is therefore critically influenced by constraints due to other cellular proteins and the background DNA, as well as the available transcriptional machinery, which is dictated by growth rate effects.

1.4 Mathematical models of gene regulation

Modeling approaches

Since the first experimental explorations of gene regulation, mathematical modeling has proven to be a useful and complementary tool to gain insight into mechanisms driving gene expression processes[51,52]. Models allow a precise description of regulatory interactions and systematic derivations of quantitative and qualitative network behavior, thereby providing powerful predictions and biological insights. Increases in genomic information and computational power in the last decades has spurred the development of an array of modeling approaches, varying in their mathematical complexity and biological detail[53,54]. Four of the most commonly employed model classes for gene regulation are: Boolean models, thermodynamic models, differential equation models and stochastic models[54].

Boolean models[54]

Boolean models rely on simplistic representations of complex biochemical networks by describing regulatory processes as logic gates and classifying the entities of the system using two states ('on' or 'off). Accordingly, these models do not require quantitative details of the system under investigation, making them easy to implement, analyze and interpret. Despite their simplicity, they have been employed successfully in many instances and can provide fundamental insights into underlying processes - in some cases coming even close to the accuracy achieved by more complex differential equation models[55]. Boolean models were for example applied to study 12 variants of the *lac* promoter, finding that few mutations were sufficient to significantly change the logic type of the input function[56].

Generally, this type of model provides a good starting point for investigations into networks that have been poorly studied experimentally, but they can give wrong or inaccurate predications if the network structure and concentrations of the molecular players are of importance.

Thermodynamic models[54]

Thermodynamic models make use of detailed knowledge of biophysical system parameters to predict gene expression levels from the combination of binding sites in the promoter region. One of the main assumptions of these models is that the prediction of gene expression can be replaced by calculating the probability of RNAP being bound to the promoter - which is increased by activator binding and decreased by repressor binding - from the equilibrium DNA occupancy of the involved binding proteins.

As a first step, all the possible system states (binding configurations of the involved TFs) have to be enumerated and their statistical mechanical weights (Boltzmann weights) assigned. The weights are determined by the concentrations of the involved TFs and their binding affinities to specific or non-specific DNA sequences. Cooperative and competitive interactions between TFs can be incorporated by modifying the Boltzmann weights. In a second step, the probability of RNAP being bound is obtained by summing the Boltzmann weights of all states where RNAP is bound and dividing it by the sum of all possible weights[57]. Even though these modeling approaches neglect processes that occur downstream of TF-DNA binding, they have proven to be very successful in diverse contexts from prokaryotic to eukaryotic regulatory networks. Application of a thermodynamic model to the Lambda phage genetic switch for example led to the suggestion that the involved repressors are non-specifically bound most of the time[58]. However, one major shortcoming of thermodynamic models is that they cannot capture temporally evolving dynamics of a system.

Another crucial concern for the suitability of these types of models concerns the second main assumption; namely that the promoter system is in equilibrium. The validity of this assumption can be judged by considering the timescales of processes

involved in regulation compared to those of transcription initiation at the promoter. If those timescales are clearly separated (regardless of which constitutes the fast and which the slow process), then the equilibrium binding probability will provide a good approximation for gene expression from this promoter. For example, the assumption is fulfilled if equilibration of TF binding states at the promoter is much faster than the protein transcription and degradation rates. However, the validity of equilibrium for promoter regulation has been experimentally challenged by showing that the *lac* repressor stays bound at its operator for several minutes[8]. More generally, many cell decisions seem to be determined by transient transcriptional dynamics rather than steady state[59–62]. The prevailing focus on steady state behavior - which is a feature of many mathematical models in gene regulation, not only thermodynamic models - might limit the phenotype space that can be studied and ignore biologically relevant mechanisms.

Differential equation models[54]

Differential equation models describe the time (or space) progression of biochemical species (e.g. mRNA and proteins) through a defined set of rules. These rules are inferred from chemical reaction equations and define how each species evolves over time as a function of other system components and the corresponding reaction rates. These types of models can also be combined with thermodynamic models: Thermodynamics are used to describe the details of protein-DNA interactions, whereas the time progression of proteins and RNA is described by the differential equations. This approach has been used for many prokaryotic promoter systems, for example the phage Lambda genetic switch[51,63] and the lac operon[64], to elucidate the emergence and stability of bistable systems.

One limitation of this type of models is the often large number of parameters that must be accounted for. This makes the approach unsuitable for poorly characterized systems and poses the danger of overfitting. Accordingly, the model fit might deliver good results but still miss important biological implications. Moreover, this leads to a computational challenge as even moderately complex networks can lead to an explosion of parameters if they are investigated in detail. Consequently, differential equation models are often not as fine-scaled as it is possible with thermodynamic models. Nevertheless, as they are well-suited to capture the dynamics of biological systems, while still allowing for a reasonable amount of detail, they provide a unique trade-off between molecular accuracy and temporal (spatial) evolution.

Stochastic models[65,66]

Stochastic models constitute the model class of highest complexity as they are able to capture stochasticity in gene expression as well as the full system dynamics. These models provide the full probability distributions of mRNA and proteins (as compared to only mean values in the previous model classes) as a function of time by solving either Langevin or Chemical Master equations.

Stochastic system descriptions tend to be very complicated even for simple systems, containing a large number of parameters. This also makes these models computationally expensive and hard to solve.

1.5 Design principles of regulatory architecture

A key challenge in understanding gene regulatory networks is the search for universal design principles that link biological function in a direct manner to a given DNA sequence[67]. It is still unclear how much of the diversity in biological designs that we observe in nature is due to 'historical accident' or due to a set of design rules that remain yet to be found[4]. Although we are still far from determining a predictive set of design principles, several studies give hope that there are network features, which have not come about by random accidents.

Comparisons to random models have revealed architectural design that are overrepresented in prokaryotic genomes. Increasing genomic information has revealed that most promoters are regulated by more than one TF, yet overall they are less heavily regulated than would be expected[13]. Moreover, many of the binding sites in co-regulated promoter regions overlap one another, hinting at an advantage of TF competition in gene regulation. Indeed, TF competition has been found to enhance the capacity of signal integration[12] and maximize information flow in signaling networks[68]. Generally, promoter architecture – meaning the arrangement, strength and multiplicity of operators - crucially determines noise in gene expression[69]. At a higher level, the overrepresentation of certain network motifs[25] – network building blocks of a specific regulatory connectivity – suggests that certain constraints are acting on their design.

Different roads have been taken to map out these design principles on different layers of the network architecture: General rules on the energetic bounds of specific TF-DNA binding have been drawn from constraints of the genomic background[24]. Similarly, optimizing information transmission in the face of transcriptional noise sets limitations on how genes are regulated[22]. High-throughput measurements have also spurred a wealth of studies trying to predict binding affinities of TFs, but the common divergence between *in vitro* and *in vivo* promoter occupancy in these approaches shows that there are important aspects missing[70]. Interactions between closely bound TFs and sequence context are possible candidates for these missing features, but even though our understanding in that regard is constantly increasing, there are no general design rules to be discerned so far. On the next level, rules for how individual transcriptional units are combined into small networks, called network motifs, are being studied mostly through computational studies as well as synthetically built networks[67]. These studies indicate that there is a finite space of solutions for a given target function, with only a handful of critical parameters determining design.

However, there are still various ways to implement a certain circuit diagram, and architecture alone is not enough to specify function[67]. A first step was taken to connect these motifs to cellular decision-making by studying them as dynamical systems and defining motifs as functional modules[71]. Combinations of different motifs however, can lead to behavior that is not easily anticipated by summing up the individual behaviors[72].

Most approaches at finding general design rules have focused on beneficial or overrepresented network features. Yet, instead of asking: 'Why are we observing certain types of regulatory structures?' it might be even more informative to ask, 'Why are we not observing other types?', meaning that it is worth focusing more on the constraints that limit the phenotypic space of regulatory systems. At the most fundamental level, this space will be delimited by the molecular interactions between regulatory components and the selection pressures acting upon them.

1.6 Biophysical constraints on regulatory architecture

Cellular behavior and adaptation arise from the timing and regulation of gene expression. Gene regulation on the other hand is crucially shaped at the molecular level by biophysical characteristics of protein-protein and protein-DNA interactions. By combining experiments with synthetic transcriptional systems and mathematical modeling, we will study how regulatory architecture is constrained through those biophysical characteristics. Specifically, we will investigate how molecular interactions shape regulatory connectivity, regulatory crosstalk, and gene expression dynamics, thereby ultimately delimiting the regulatory design space upon which selection can act.

1.7 A brief overview

Chapter 2

We investigate the mechanistic basis of rewiring gene regulation at a local level, meaning changes in connectivity between TFs and their target promoters, which drives gene regulatory network evolution, but has been neglected so far. A combination of well-known molecular systems and a simple thermodynamic model allows us to dissect the contribution of individual parameters to local rewiring. The results highlight that only a few, inherent biophysical properties of individual network components – TFs – crucially determine the potential for evolutionary changes.

Chapter 3

Regulatory rewiring between two promoters can lead to binding sequences that have substantial affinity for both repressors. We explore the consequences of molecular binding crosstalk between two related repressors at several mutated promoters. The

presence of a second repressor can lead to a counter-intuitive increase in promoter repression over repression achieved by increasing the concentration of the focal TF. A simplified thermodynamic model suggests the importance of global competition at non-specific binding sites between the two repressors in explaining this phenomenon.

Chapter 4

In temperate phages crosstalk between their repressors is directly related to organismal fitness. We explore the potential consequences of repressor crosstalk on phage lifestyle decisions and fitness.

Chapter 5

We study the impact of non-specific TF binding to background genomic DNA on cellular fitness. Certain growth conditions - which are relevant in natural bacterial environments - show a significant growth arrest in the presence on non-specific TF binding. This effect seems to be dependent on TF concentration, non-specific binding strength and binding cooperativity, as well as DNA concentration; possible putting global cellular constraints on gene regulatory parameters and horizontal gene transfer.

Chapter 6

Competition between transcription factors in a promoter region due to closely spaced binding sites is a common architecture at prokaryotic promoters, yet the implications of this architecture for signal integration remain largely unexplored. We investigate the importance of signal arrival timing and ordering at a promoter with two inputs and find substantial history-dependence in gene expression over several hours. Accordingly, we employ a differential equation model combined with a thermodynamic description of protein-DNA binding and protein-protein interactions to elucidate the underlying mechanisms. Interference between TFs at their binding sites seems to be a previously unrecognized means to encode memory in gene regulation and to produce gene expression variability that is not caused by stochasticity.

2. Evolutionary potential of transcription factors for gene regulatory rewiring

The following chapter has been published in Nature Ecology& Evolution:
Igler, C., Lagator, M., Tkačik, G., Bollback, J. P. & Guet, C. C. Evolutionary potential of transcription factors for gene regulatory rewiring. *Nat. Ecol. Evol.* **2,** (2018).

2.1 Abstract

Gene regulatory networks evolve through rewiring of individual components, that is, through changes in regulatory connections. However, the mechanistic basis of regulatory rewiring is poorly understood. Using a canonical gene regulatory system, we quantify the properties of transcription factors that determine the evolutionary potential for rewiring of regulatory connections: robustness, tunability, evolvability. *In vivo* repression measurements of two repressors at mutated operator sites reveal their contrasting evolutionary potential: while robustness and evolvability were positively correlated, both were in trade-off with tunability. Epistatic interactions between adjacent operators alleviated this trade-off. A thermodynamic model explains how the differences in robustness, tunability and evolvability arise from biophysical characteristics of repressor-DNA binding. The model also uncovers that the energy matrix, which describes how mutations affect repressor-DNA binding, encodes crucial information about the evolutionary potential of a repressor. The biophysical determinants of evolutionary potential for regulatory rewiring constitute a mechanistic framework for understanding network evolution.

2.2 Introduction

From the seminal discovery of repression and activation as the basic mechanisms of gene regulation[1,73], a fundamental picture has emerged where individual regulatory components - promoters and transcription factors (TFs) - are interconnected into gene regulatory networks (GRNs): global structures that determine cellular gene expression patterns. However, a mechanistic understanding of how GRNs evolve is still lacking. GRN evolution can be studied at two opposing levels of organization: (i) global emerging features of GRNs, such as functional redundancy, which can promote changes in network structure[30], or (ii) local rewiring, which leads to the formation of new regulatory connections within GRNs[38]. The principles of GRN evolution have been primarily studied globally, at the level of entire networks, through comparative genomic analyses[34,38] or *in silico*[74,75], in order to understand how global network features determine evolutionary properties like robustness[76] (phenotypic persistence in the face of mutation), tunability[77] (changes in gene expression levels), and evolvability[28] (capacity to acquire new regulatory connections). Yet, GRN structures can change solely through making and breaking of connections at the molecular level, that is, through local rewiring of individual components[33,35,39,78–80]. However, how characteristics of individual regulatory components impact GRN evolution by determining robustness, tunability and evolvability is unknown.

Local network rewiring, i.e. changes in the binding specificity of a TF, involves loss of binding, gain of binding and modifications in the strength of binding, which occur either through mutations in TFs or in DNA-binding sites of TFs (operators). Most experimental studies on network rewiring focused on mutations in proteins[81] or on the consequences of gene duplication events[36,37,82], showing that TF divergence affects GRN evolution[83]. However, in contrast to mutations in operators[84–86], mutational pathways of TFs are thought to be heavily constrained by epistasis between amino acids[87], the high frequency of deleterious mutations[88] and the strong pleiotropic effects of TFs[89], suggesting that operators are superior targets for modifying existing and acquiring novel network connections.

In contrast to previous studies on promoter evolution, which considered promoters independently of the associated TFs[86,90–92], we want to understand how the properties of a TF determine its evolutionary interactions with operator sites. To achieve this, we define the *evolutionary potential for local rewiring* with respect to point mutations in an operator, thus characterizing the evolutionary potential for an individual network component that does not itself change: the repressor. We combine three distinct properties, which have been previously used to describe network rewiring[35,93,94], to define the evolutionary potential of a repressor as the ability (i) to withstand operator mutations (*robustness*), (ii) to modify the strength of binding to existing operators (*tunability*), and (iii) to acquire binding to new operators (*evolvability*) (Fig. 2.1a). Using two of the best understood prokaryotic repressors - Lambda CI and P22 C2 - we study how characteristics of individual TFs determine the evolutionary potential for regulatory rewiring.

2.3 Results

Experimental system for quantitative measurements of evolutionary potential

We used homologous[95] elements of the bacteriophage Lambda and P22 genetic switches[2,96]. Specifically, we used Lambda CI and P22 C2 repressors, along with their respective P_R promoter regions. The P_R promoter region consists of RNA Polymerase (RNAP) binding sites and two operators, O_{R1} and O_{R2}, which regulate P_R expression through cooperative repressor binding (Fig. 2.1b). We experimentally studied changes in gene expression, and hence binding of the repressors, along the mutational path between the two promoters by directionally mutating the operator sequence of one repressor to that of the other (Fig. 2.1c). Throughout, we refer to systems containing matching (non-matching) repressors and promoters as *cognate* (non-cognate) (Fig. 2.1b). We created a library of O_{R1} operator mutants by selecting all base pairs known to have large impact on repressor binding[97,98], and that differed between Lambda and P22 O_{R1} sequences, resulting in six mutated positions (Fig. 2.1d, Table 2.1).Subsequently, we also investigated mutations in O_{R2}, even though repressor binding to this operator is considered to have only a minor direct impact on P_R repression[2]. All mutants were cloned into a very low copy number plasmid[99] and fluorescence as a proxy for P_R expression levels was measured in the presence and absence of repressor. This setup, which measures binding of two repressors along the

mutational path between the two operators, allowed us to study in a comparative manner how the evolutionary potential for regulatory rewiring depends on repressors themselves.

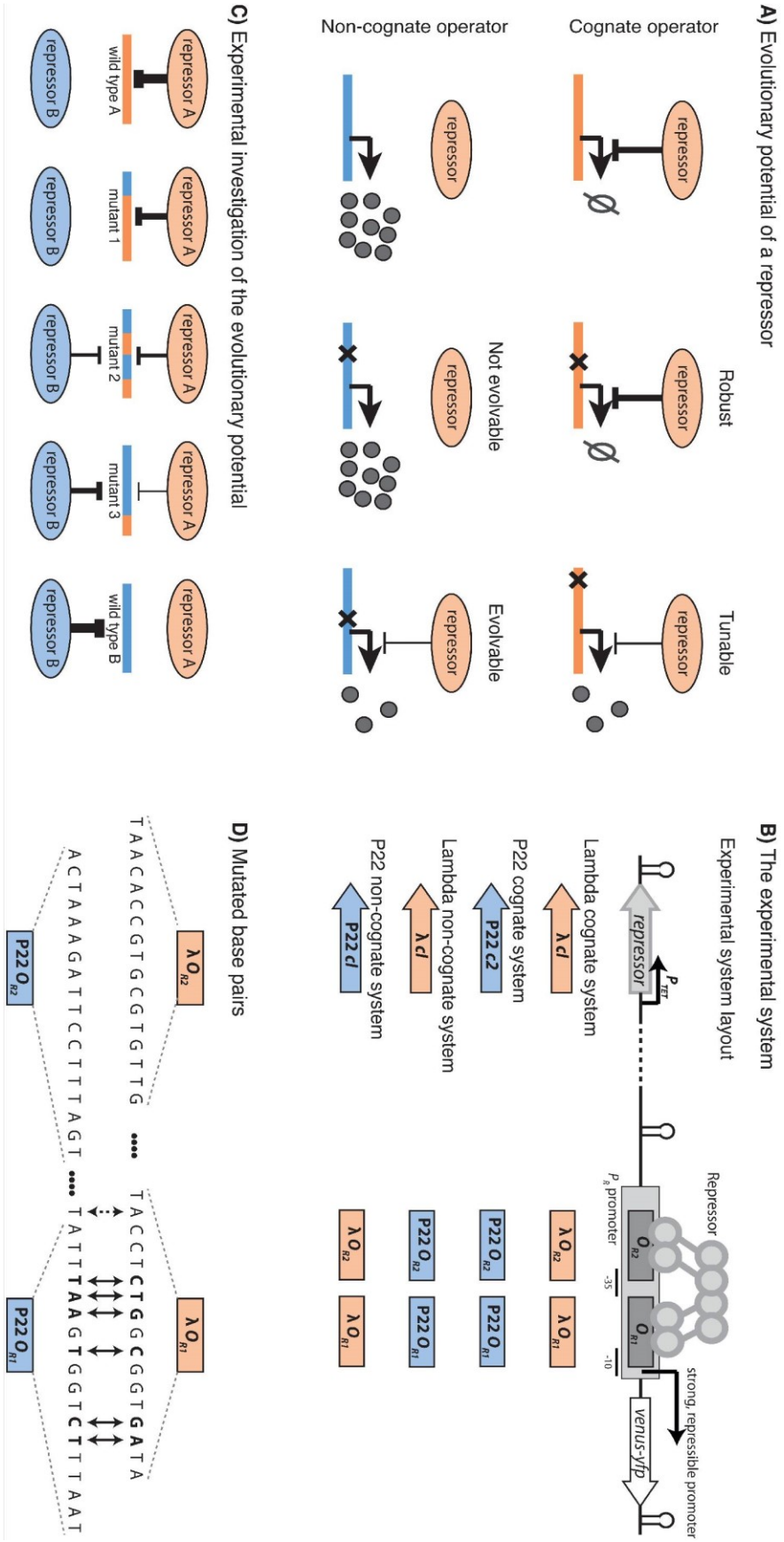

Figure 2.1. Experimental investigation of evolutionary potential of a repressor.

a) *Mutations (indicated by 'x') in the cognate operator can either have no effect on repressor binding (**robust**); alter repressor binding (**tunable**); or remove repressor binding (not shown). Mutations in the non-cognate site can either have no effect on repressor binding (**not evolvable**); or lead to gain of repressor binding (**evolvable**). Together, robustness, tunability and evolvability describe the evolutionary potential for regulatory rewiring.* ***b)*** *The synthetic template consists of a repressor controlled by an inducible P_{tet} promoter, and a strong P_R promoter - containing two repressor operators (O_{R1} and O_{R2}) and the RNA Polymerase (RNAP) binding sites - that controls the expression of a fluorescence marker venus-yfp.* ***c)*** *An increasing number of mutations (blue) are introduced into the cognate operator (orange) of repressor A. The thickness of the blunt-ended arrows indicates the strength of repression.* ***d)*** *Homology alignment of Lambda and P22 O_{R1} and O_{R2}, showing mutated sites in bold. Arrows show O_{R1} base pairs that were exchanged. The dashed arrow marks an additional site that was used to construct four cognate Lambda mutants, as one of the original positions abolished RNAP binding (Table 2.1).*

Evolutionary potential of repressors

To characterize the evolutionary potential of the two repressors, we experimentally measured their robustness, tunability and evolvability in terms of how repressor binding is affected by operator mutations. Robustness and tunability were quantified on the cognate promoter background. *Robustness* was the fraction of cognate operator mutants that maintained at least 90% repression. *Tunability* was the standard deviation in repression levels when repression was reduced but not completely lost (90-10%). From these definitions, it does not follow that robustness and tunability are necessarily negatively correlated: the expression variability (tunability) generated by non-robust mutations can be either large or small. *Evolvability* was the fraction of non-cognate operator mutants that could be repressed to at least 10%.

Lambda CI and P22 C2 have drastically different evolutionary potential (Fig. 2.2a), in spite of their shared ancestry[95]. These differences are particularly evident when considering the relationship between repression and the number of mutations in the operator (Fig. 2.2b). The high Lambda CI robustness to up to three mutations is surprising, since the O_{R1} site is almost fully conserved across at least twelve different lambdoid phages[100]. As this site is part of a complex promoter region in the phage, it could be conserved due to binding of RNAP or the second repressor in the switch (Cro). In contrast to Lambda CI, one to three mutations in the P22 cognate O_{R1} site led to a wide range of repression (0-100%).

At the non-cognate site, even introduction of single point mutations in P22 O_{R1} led to repression of at least 35% by Lambda CI (Fig. 2.2c). Gain of binding to the non-cognate site was much less frequent for P22 C2, and, except for one mutant, the range of repression was 0-20%, markedly lower than the 10-90% of Lambda CI (Fig. 2.2c).

Overall, Lambda CI had higher robustness as well as evolvability, suggesting that a repressor that is more robust to mutations in its cognate operator might also more readily acquire novel binding sites. At the same time, P22 C2 was more tunable, indicating a trade-off between robustness and tunability. The consistently stronger binding of Lambda CI compared to P22 C2 suggests that the evolutionary potential for regulatory rewiring is a property of the repressor, not of the operator.

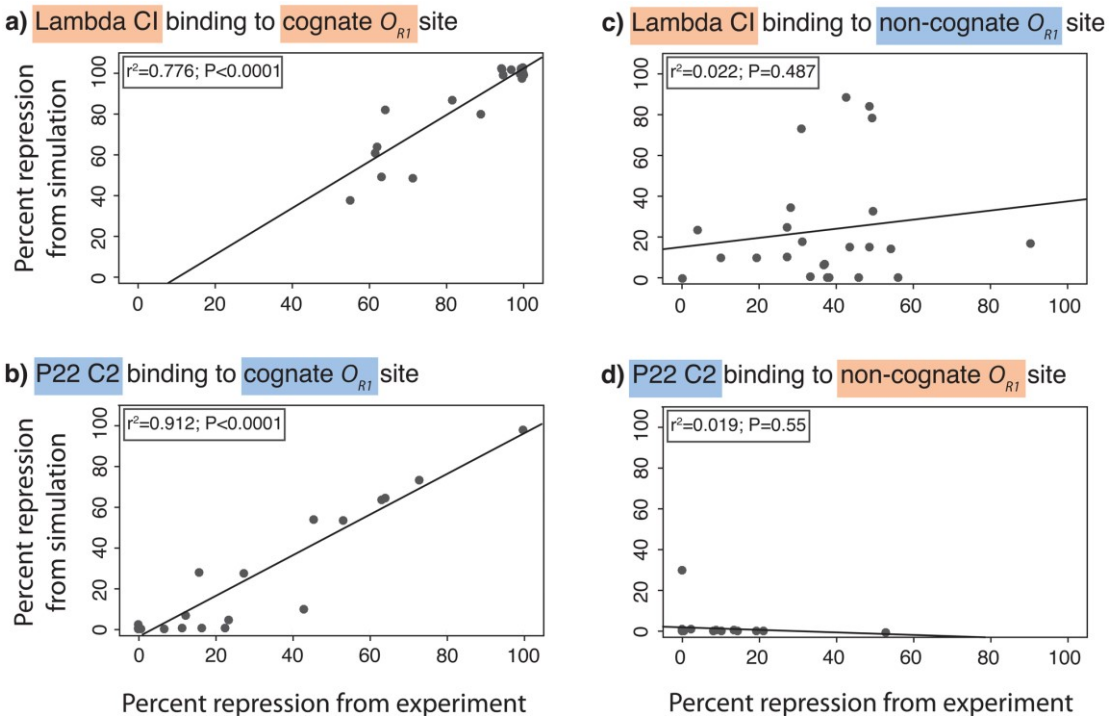

Figure 2.2. Lambda CI and P22 C2 have different evolutionary potential.

a) *Robustness, tunability and evolvability of Lambda CI and P22 C2.* ***b)*** *Loss of binding was determined by mutating away from the cognate site, making it more similar to the non-cognate site. The dotted line shows the 90% repression threshold used to evaluate robustness.* ***c)*** *Gain of binding was determined by mutating away from the non-cognate site making it more similar to the cognate one. The dotted line shows the 10% repression threshold for evolvability. Expression levels in the absence of repressor are shown in Table 2.2. Mutants that abolished RNAP binding are not shown, resulting in a different number of mutants in b) and c). Points show mean percent repression over three replicates, bars are standard errors of the mean. Lambda is orange, P22 is blue. Binding to the wild type cognate or non-cognate site is shown by a dark orange point.*

Thermodynamic model of evolutionary potential

In order to expand on the experimental findings and identify how evolutionary potential depends on the biophysical system parameters, we used a thermodynamic model of gene regulation[51,57] (Fig. 2.3a). While experimentally we determined the general trends underlying the evolutionary potential of the two repressors by introducing mutations in a directional manner, we used the model to comprehensively explore all possible mutations in the six selected O_{R1} positions.

The model — for which all parameter values except repressor concentrations were taken from literature (Table 2.3, Fig. 2.4) — accurately reproduced experimental observations in cognate mutants (Fig. 2.5). The poor model fit to non-cognate mutants is not surprising, as the model assumption of independent contribution of each position to the overall binding energy is known to be violated when mutated far away from the wild type sequence[101]. Nevertheless, the use of the model is justified because: (i) the model performs comparably for both repressors (Fig. 2.5), (ii) it provides a lower bound for the experimentally measured non-cognate repression, and (iii) only modest improvements are achievable by accounting for dinucleotide dependencies[102,103].

.

We simulated binding to all possible mutants at the six chosen positions (4095) and quantified the evolutionary potential of repressors: for tunability and evolvability we used the same definitions as in the experiments (Fig. 2.3b,c), but calculated them separately for each mutant class. We used a standard definition to quantify robustness in our simulations[76] (see Methods), which we could not apply to the experimental measurements due to the insufficient number of mutants connected by single mutations. Importantly, applying the experimental definition of robustness to the simulations identified consistent differences in robustness (51.9% for Lambda CI and 0.3% for P22 C2). Overall, model simulations corroborated the experimentally determined differences in the evolutionary potential of the two repressors: Lambda CI was more robust and more evolvable than P22 C2, but less tunable for up to three mutations (Fig. 2.3d).

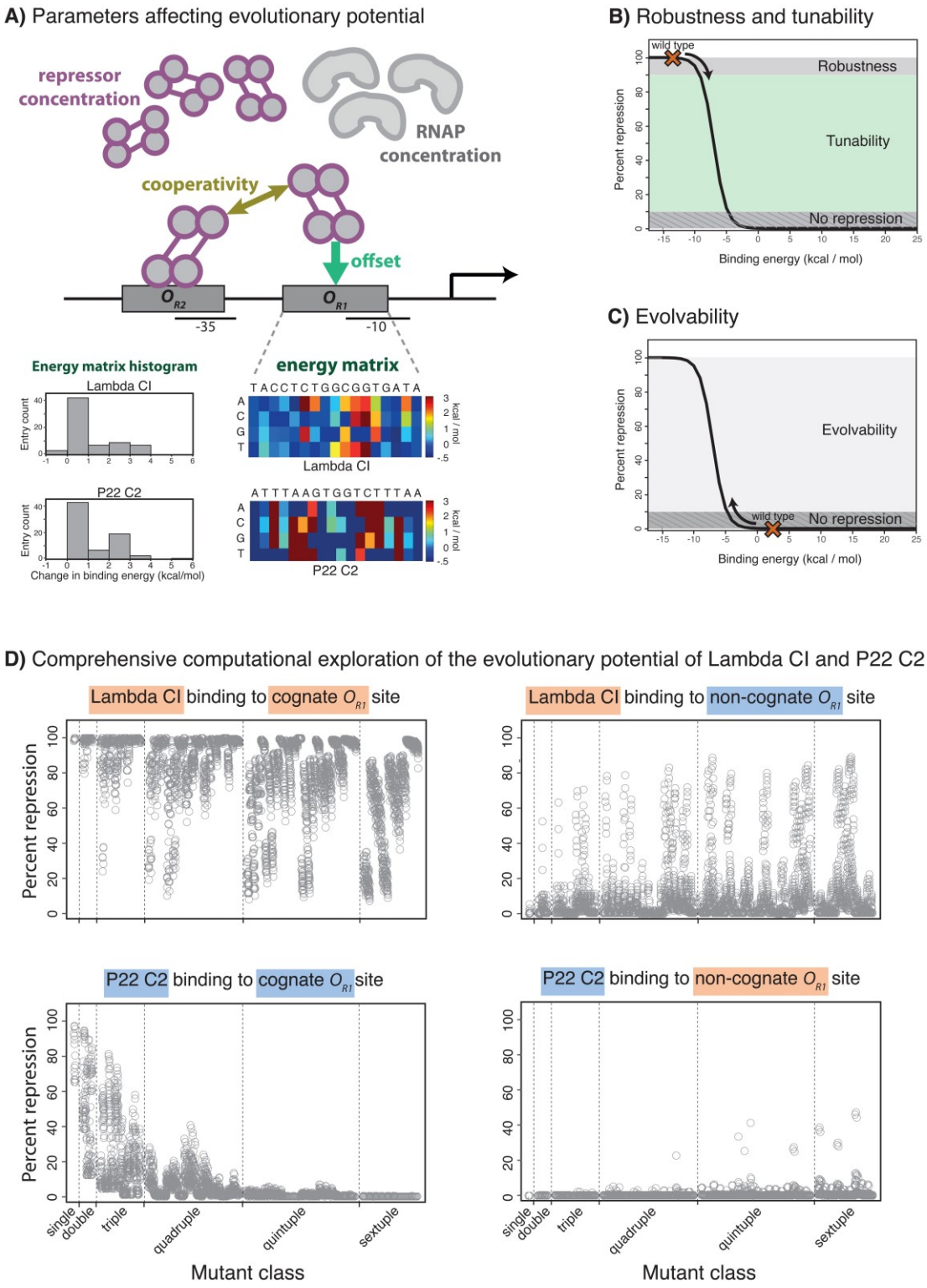

Figure 2.3. Thermodynamic model of gene expression.

a) *Gene expression is determined by: intra-cellular concentration of (i) repressor, and (ii) RNAP; iii) cooperativity of binding between two repressor dimers; iv) binding energy to the wild type operator (offset E_{WT}); and v) additional contribution of each mutation to the binding energy (energy matrix). Negative (positive) entries in the energy matrix show mutations that decrease (increase) binding energy, and hence increase (decrease) repression. Zero values denote the wild type sequence.* ***b), c)*** *The sigmoidal relationship between binding energy and repression, determined by the*

thermodynamic model, provides quantitative definitions of robustness, tunability and evolvability. ***d)*** *Comprehensive simulation of repression for all possible mutations in the six chosen positions in* O_{R1}.

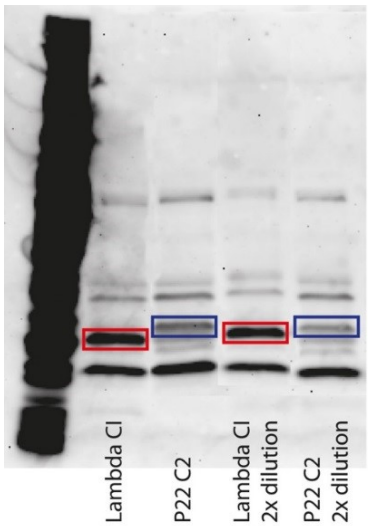

Figure 2.4. Cellular concentrations of the two repressors are 3-6 fold different.

Gel electrophoresis image of a Western blot used to calculate relative cellular concentrations of Lambda CI and P22 C2. Western blot was carried out in the presence of the system inducer, aTc. The left-most lane shows the molecular weight marker. The first visible band above the boxed bands shows the reference gene by which the values for Lambda CI and P22 C2 concentration were normalized. Red boxes mark Lambda CI bands; blue boxes mark P22 C2 bands. We calculated the relative difference in concentrations of the two repressors at full concentration and in a 2x dilution. Lambda CI had a higher cellular concentration by a factor of 3.8 at full concentration, and of 5.5 at 2x dilution.

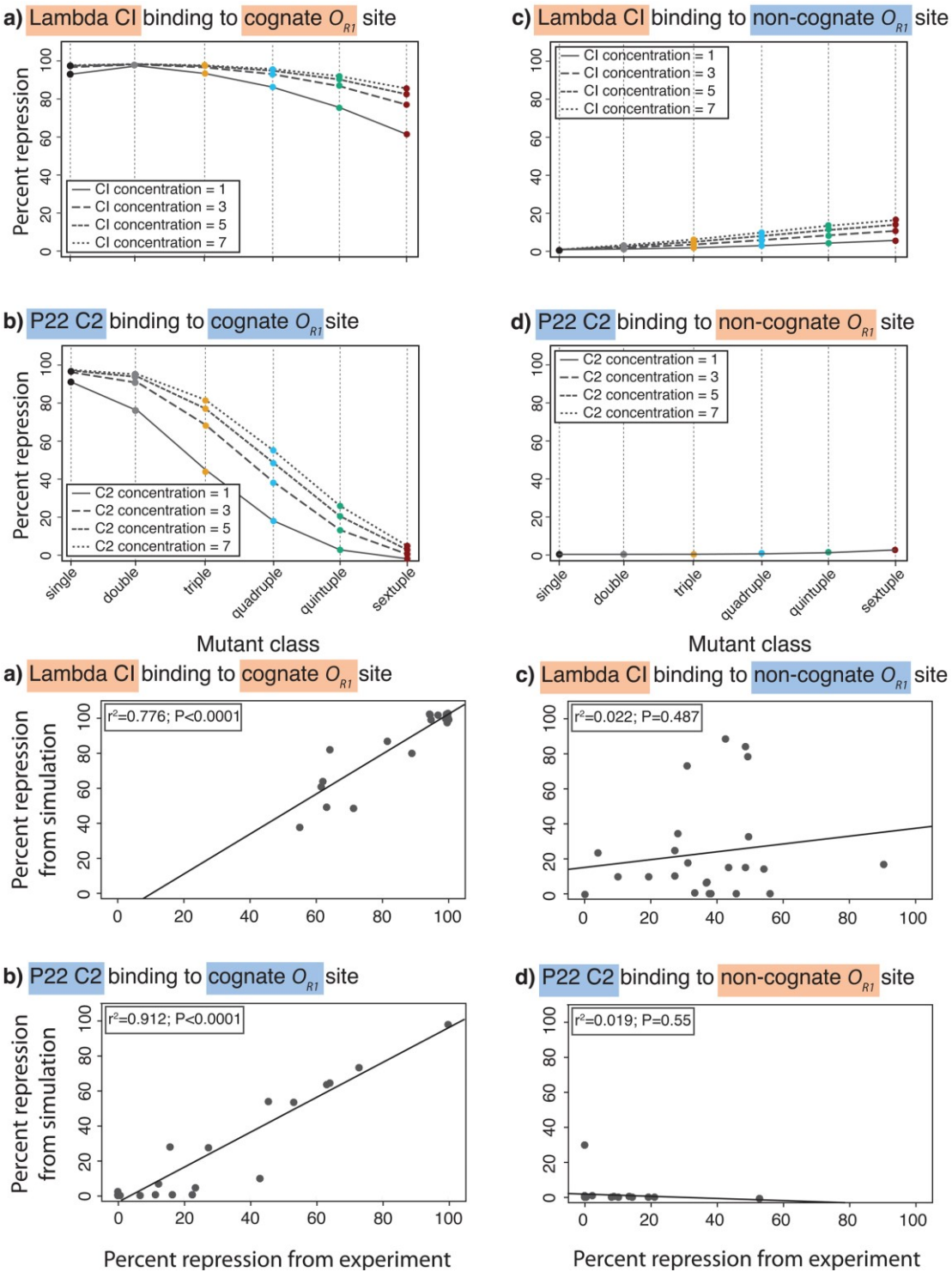

Figure 2.5. Model-to-data fit for O_{R1} mutant library.

Linear regression between experimental and model-derived percent repression of each experimentally tested O_{R1} mutant (from Fig. 2.1d) is shown for: ***a)*** *Lambda cognate mutants;* ***b)*** *P22 cognate mutants;* ***c)*** *Lambda non-cognate mutants; and* ***d)*** *P22 non-cognate mutants. Lambda is orange, P22 is blue. The model accurately describes binding of both repressors to their cognate sites (left side panels), but tends to underestimate their binding to non-cognate sites as the simulation values are generally lower than the experimental values (right side panels). Previously determined offset and energy matrices for Lambda CI[97] and P22 C2[98], as well as the*

strength of cooperativity between repressor dimers[12] were used in the model. Only the repressor concentrations were fitted from the data. The good fit for cognate mutants suggests a general agreement with previously published experimentally determined binding energies[97,98].

To confirm that the observed differences in evolutionary potential did not arise from the specific operator sites used in this study, we simulated evolvability of both repressors to 10^6 random operators. We found that Lambda CI bound a consistently higher portion of random sites (Fig. 2.6) irrespective of repressor and RNAP concentration, further supporting the view that evolutionary potential is a property of the repressor, not the operator.

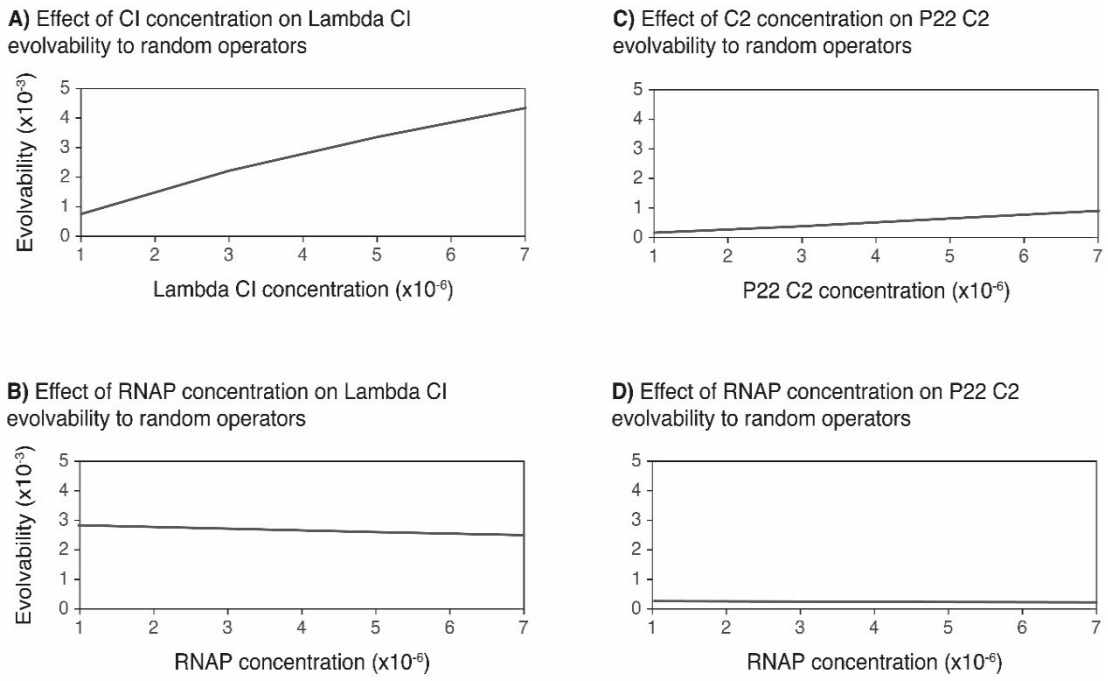

Figure 2.6. Evolvability on random operator sequences.

*Repression by Lambda CI and P22 C2 was modeled on 1 million random 17 or 18 base pair operator sequences in the lac operon promoter region, for which the repressor operator does not overlap with the RNAP binding site. Binding energies were calculated for random operators using Lambda and P22 energy matrices and for RNAP using the energy matrix from Kinney et al., 2010[117]. Lambda CI bound a higher portion of random sequences than P22 C2, irrespective of repressor (**a,c**) or RNAP concentrations (**b,d**), indicating an inherent difference in the evolvability of the two TFs.*

The thermodynamic model identifies several system parameters that affect the evolutionary potential of a repressor (Fig. 2.3a): (i) intra-cellular conditions, i.e. concentrations of repressor and RNAP, (ii) interactions arising from the promoter architecture, which in our system enable cooperative repressor binding, and (iii) intrinsic binding characteristics of the repressor itself. Repressor-specific binding

characteristics are captured in the total binding energy, E_{tot}, which is determined by the strength of repressor binding to its wild type operator (called 'offset', or E_{WT}), to which the effect of each mutation on binding is added, as defined by the 'energy matrix' (E_{seq}), so that $E_{tot} = E_{WT} + E_{seq}$. Hence, the 'offset' captures the overall propensity of a repressor to bind cognate DNA, while the 'energy matrix' describes how operator mutations affect repressor binding.

A) Effects of repressor concentration, RNAP concentration and cooperativity on evolutionary potential

	Robustness	Tunability	Evolvability
Repressor concentration (range: 1-7 x 10^{-6}M)	+ / +	− / ✱	+ / +
RNAP concentration (range: 1-7 x 10^{-6}M)	0 / −	+ / ✱	− / −
Cooperativity (range: 1-7 k_BT)	+ / +	− / ✱	n/a

B) Swapping model parameters between Lambda CI and P22 C2

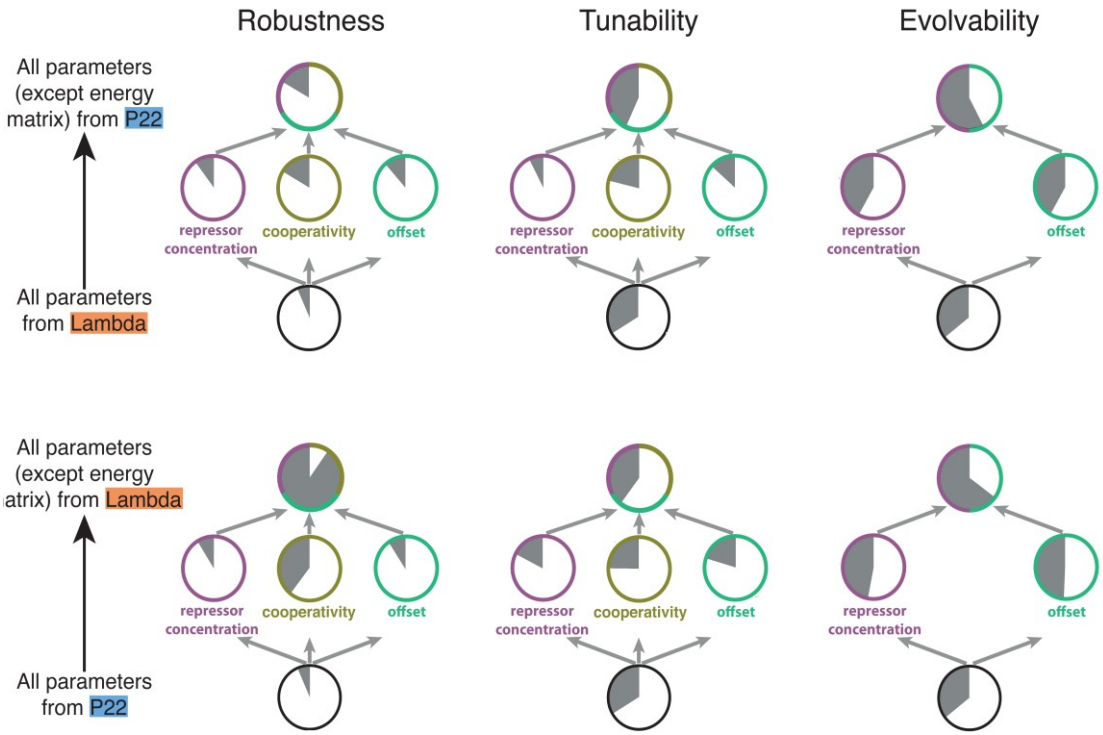

Figure 2.7. System parameters determine evolutionary potential.

a) *Correlation between each evolutionary property and a given system parameter: '+' indicates a positive correlation; '-' a negative correlation; '0' a negligible effect; and '*' a non-linear relationship. Lambda CI is orange, P22 C2 is blue.* ***b)*** *We swapped parameter values of repressor concentration, cooperativity and offset from one repressor to the other. 'Fraction of variance explained' (R^2) was calculated between the repressor with swapped parameter(s), and the other repressor with its original parameters. R^2 is shown as the grey portion of the pie charts: the fuller the pie chart, the more similar the evolutionary property between the two repressors. Starting from*

the original parameter values, each of the three parameters was swapped individually, and all three simultaneously.

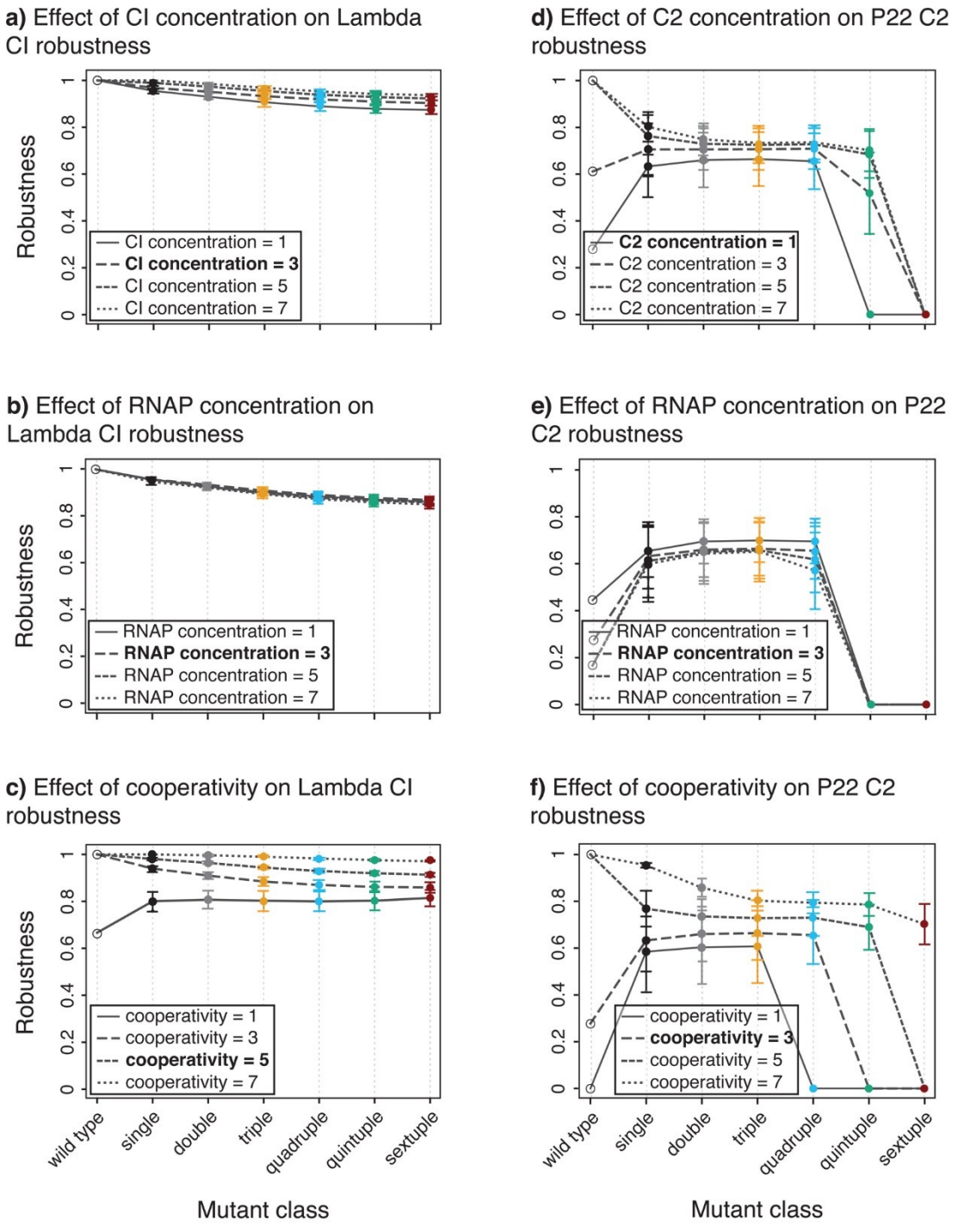

Figure 2.8. Robustness depends on repressor and RNAP concentrations as well as on cooperativity.

Points are mean robustness for all mutants with a given number of mutations, with colors indicating the mutant class. Bars are the variance of the mean frequency of neutral mutations for a given mutant class. ***a)*** *Effect of CI concentration on Lambda CI robustness;* ***b)*** *Effect of RNAP concentration on Lambda CI robustness;* ***c)*** *Effect of cooperativity on Lambda CI robustness;* ***d)*** *Effect of C2 concentration on P22 C2*

robustness; ***e)*** *Effect of RNAP concentration on P22 C2 robustness;* ***f)*** *Effect of cooperativity on P22 C2 robustness. Each line connects mean robustness across mutant classes for a given parameter value. Cooperativity values are given in kcal/mol. CI and RNAP concentration are given as x10^{-6}M. The parameter values for cooperativity and concentrations that were used in the model are bolded.*

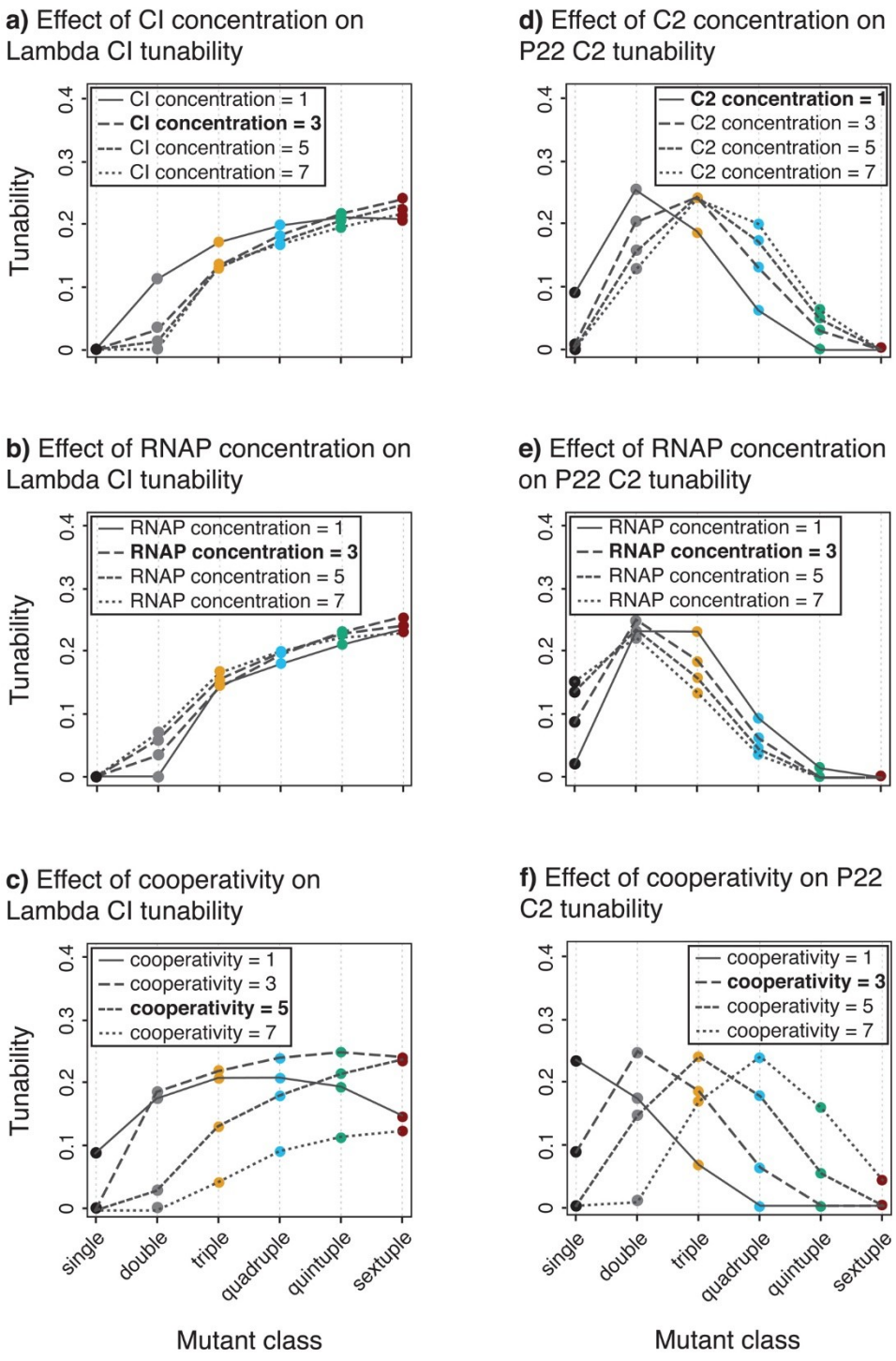

Figure 2.9. Tunability depends on repressor and RNAP concentrations as well as on cooperativity.

Points are tunability values for a given mutant class. Colors indicate the mutant class. ***a)*** *Effect of CI concentration on Lambda CI tunability;* ***b)*** *Effect of RNAP concentration on Lambda CI tunability;* ***c)*** *Effect of cooperativity on Lambda CI tunability;* ***d)*** *Effect of C2 concentration on P22 C2 tunability;* ***e)*** *Effect of RNAP concentration on P22 C2 tunability;* ***f)*** *Effect of cooperativity on P22 C2 tunability. Each line connects tunability measures across mutant classes for a given parameter value. Cooperativity values are given in kcal/mol. CI and RNAP concentration are given as x10^{-6}M. The parameter values for cooperativity and concentrations that were used in the model are bolded.*

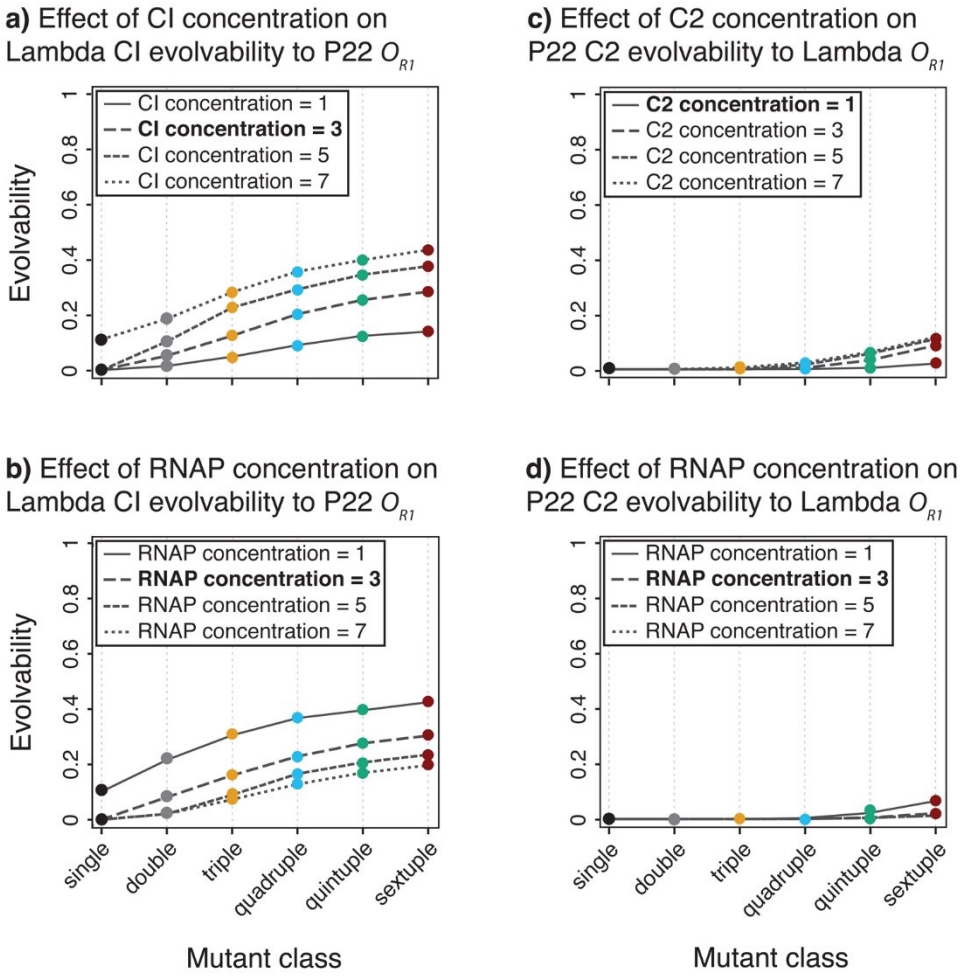

Figure 2.10. Evolvability depends on repressor and RNAP concentrations.

Points are evolvability values for a given mutant class. Colors indicate the mutant class. ***a)*** *Effect of CI concentration on Lambda CI evolvability;* ***b)*** *Effect of RNAP concentration on Lambda CI evolvability;* ***c)*** *Effect of C2 concentration on P22 C2 evolvability;* ***d)*** *Effect of RNAP concentration on P22 C2 evolvability. Each line connects evolvability measures across mutant classes for a given parameter value. Cooperativity values are given in kcal/mol. CI and RNAP concentration are given as x10^{-6}M. The parameter values for cooperativity and concentrations that were used in the model) are bolded.*

Repressor and RNAP concentrations, as well as binding cooperativity, influence robustness, tunability and evolvability to different degrees, though not always in a straightforward manner (Fig. 2.7a; Fig. 2.8, 2.9, 2.10). As such, the evolutionary

potential for rewiring depends on intra-cellular conditions that change with cellular physiology[104], and on the promoter architecture that can determine binding cooperativity. Experimental measurements of relative repressor concentrations revealed 3.8 to 5.5-fold higher intracellular Lambda CI levels (Fig. 2.4). Reassuringly, the difference in evolutionary potential between repressors was consistently identified across a range of repressor and RNAP concentrations, making the model results largely independent of uncertainty in these parameters (Fig. 2.11).

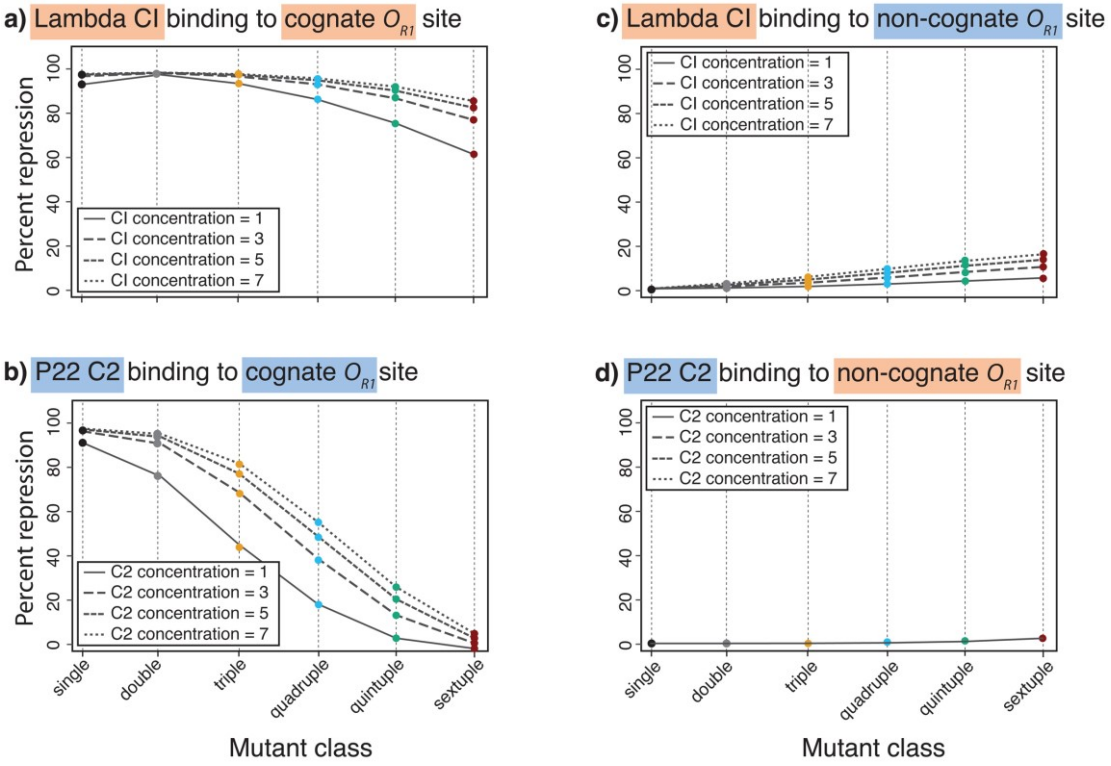

Figure 2.11. Various concentration values show the same differences in repression values between the two repressors.

Points are mean repression values from model simulations for all mutants with a given number of mutations (colors indicate the mutant class) for ***a)*** *Lambda cognate mutants;* ***b)*** *P22 cognate mutants;* ***c)*** *Lambda non-cognate mutants; and* ***d)*** *P22 non-cognate mutants. Each line connects mean repression across mutant classes for a given concentration value, which are given as x10^{-6}M. Standard errors of the mean are not shown, as they were smaller than the points indicating mean repression values.*

Biophysical determinants of evolutionary potential

We asked if it was possible to reconcile the differences in the evolutionary potential between Lambda CI and P22 C2 by swapping their model parameters. Specifically, we calculated robustness and tunability for one repressor after swapping either repressor concentration or cooperativity with the parameter values of the other repressor. For evolvability, we only swapped repressor concentration, since the absence of a cognate O_{R2} site prevented cooperative binding.

Swapping either repressor concentration or cooperativity between Lambda CI and P22 C2 decreased the differences in robustness and evolvability, but still left a disparity in robustness, tunability and evolvability of at least 50% (Fig. 2.7b). Therefore, intrinsic binding characteristics of repressors - the offset and the energy matrix - crucially determine their evolutionary potential, as previously found for the regulation of the *lac* promoter[105]. When we swapped the offset between the two repressors, we found that the effect was comparable to the effects of swapping either repressor concentration or cooperativity. Notably, swapping all three parameters did not lead to a full reconciliation between the two repressors (Fig. 2.7b), indicating that the energy matrices accounted for the remaining differences of at least 30% (except for robustness when swapping from P22 C2 to Lambda CI).

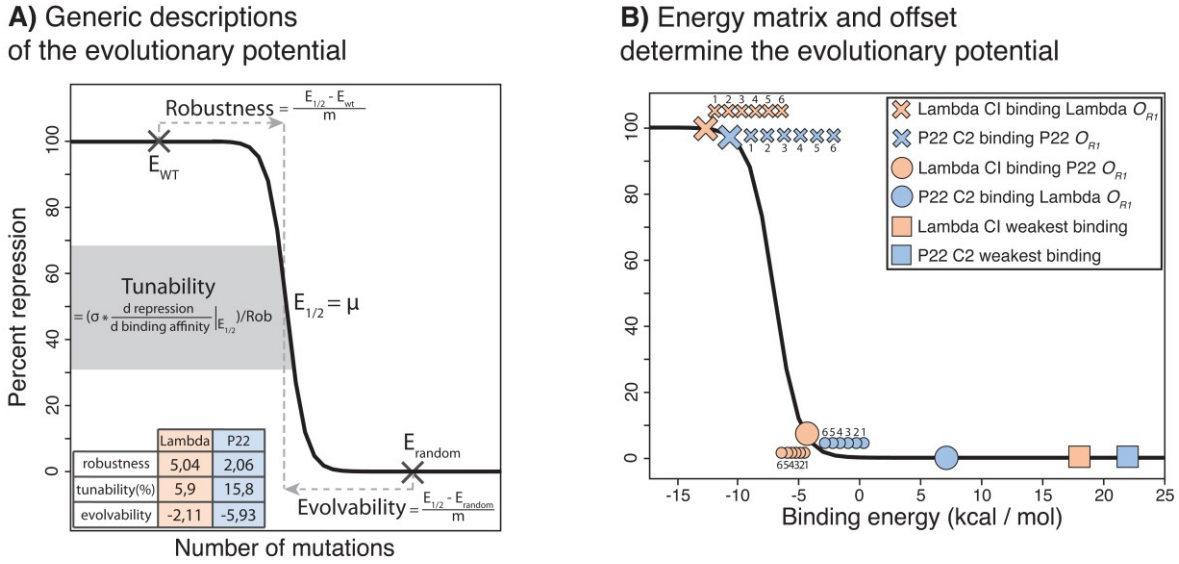

Figure 2.12. Biophysical determinants of the evolutionary potential.

a) *Generic definitions of robustness, tunability and evolvability that utilize only the offset and the energy matrix.* $Rob = \frac{E_{1/2} - E_{WT}}{m}$ *and* $Evo = \frac{E_{1/2} - E_{random}}{m} = Rob + \#mut$*, where* $E_{1/2}$ *is the binding energy at half repression (which equals the chemical potential, μ),* E_{random} *is the typical binding energy to a random sequence, m the average mutational effect size, and #mut the distance of the random sequence to the cognate operator in number of mutations (see Methods). Evolvability is negative as mutations towards* $E_{1/2}$ *improve binding.* $Tun = (\sigma * \frac{d\ repression}{d\ binding\ affinity}|_{E_{1/2}})/Rob$*, where σ is the standard deviation of the energy matrix and* $\frac{d\ repression}{d\ binding\ affinity}|_{E_{1/2}}$ *the slope of the sigmoid curve at* $E_{1/2}$*. The table shows the values for robustness, tunability and evolvability for the experimental systems (Fig. 2.1b). Here, we calculated evolvability for the non-cognate sites of Lambda CI and P22 C2.* ***b)*** *Locations of Lambda CI and P22 C2 binding to three categories of operators (*E_{WT}*,* $E_{non\text{-}cognate}$*,* E_{max}*) are indicated by large symbols on the sigmoidal curve relating binding energy and repression. Repressor concentrations are kept equal. Small symbols show mean energy values obtained through model simulations for different mutant classes (1 – single, 2 – double, etc) when mutating the cognate (crosses) or the non-cognate (circles) operators.*

To better understand the mechanism by which intrinsic binding characteristics of a repressor (offset and energy matrix) determine the differences in the evolutionary potential, we developed an intuitive and generic description of robustness, tunability and evolvability based on the sigmoidal curve relating repressor binding energy to repression (Fig. 2.12a). The formulas in Figure 2.12a describe the evolutionary potential in terms of the offset and the energy matrix, rather than using the full thermodynamic model. Robustness is the average number of mutational steps needed to lose 50% of repression. Evolvability is the average number of mutational steps necessary to gain 50% of repression starting from a given random sequence. Tunability is the ease of generating variation in gene expression levels, i.e. the variation in repression around the half-repression point, defined in relation to the distance between this point and the cognate operator (Fig. 2.12a).

Adopting these generic definitions results in simple analytical expressions (Fig. 2.12a), which show that robustness and evolvability are positively correlated through the number of mutations that separate the given random sequence from the cognate operator. This correlation holds true as long as: (i) the average mutational effect size (m) is relatively small and similar between repressors – which is a reasonable assumption in general because the scale of m is set by the energetics of hydrogen bonds (1-3 kcal/mol)[24], but also an assumption that is specifically testable for any particular set of TFs for which the energy matrices are known; and (ii) the energy matrix is a fixed property of a repressor, meaning that m stays constant when mutating towards a random non-cognate site. Tunability, on the other hand, is in a trade-off with robustness, although the dependence of tunability on the standard deviation of mutational effects suggests that this relationship can be adjusted to some extent.

Applying these generic definitions to the systems used in this study, we observe higher robustness and evolvability, but lower tunability for Lambda CI (Fig. 2.12a). To illustrate that these generic definitions are in accordance with the binding landscape obtained through model simulations, we used the simplest model setup where repressors bind only a single operator site and repressor concentrations are the same. We selected three operator sequences for each repressor - the cognate (E_{WT}), the non-cognate ($E_{non\text{-}cognate}$), and the weakest binding (E_{max}) sequence - computed their binding energies, and positioned them on the sigmoidal repression curve.

The consistently stronger binding of Lambda CI to all three types of operators (Fig. 2.12b) arises from its lower offset (-13.2 kcal/mol, compared to -12 kcal/mol for P22 C2) and smaller average mutational effect size (1.23kcal/mol, compared to 2.43kcal/mol for P22 C2). Positioning the mean binding energy of each mutant class (Fig. 2.2) on the sigmoidal curve (hence not using the full model but only the offset and the energy matrix) allowed accurate predictions of the experimental measurements, at least for cognate sites (Fig. 2.13). Therefore, the lower offset of Lambda CI places it further away from the slope of the repression curve (Fig. 2.12b), resulting in higher robustness, but lower tunability. Similarly, Lambda CI binds the

non-cognate operator, all of its mutants, and even the operator sequence with weakest possible binding more strongly (Fig. 2.12b), illustrating that, on average, Lambda CI binding a random sequence will be closer to the rise of the sigmoidal curve and hence, more evolvable.

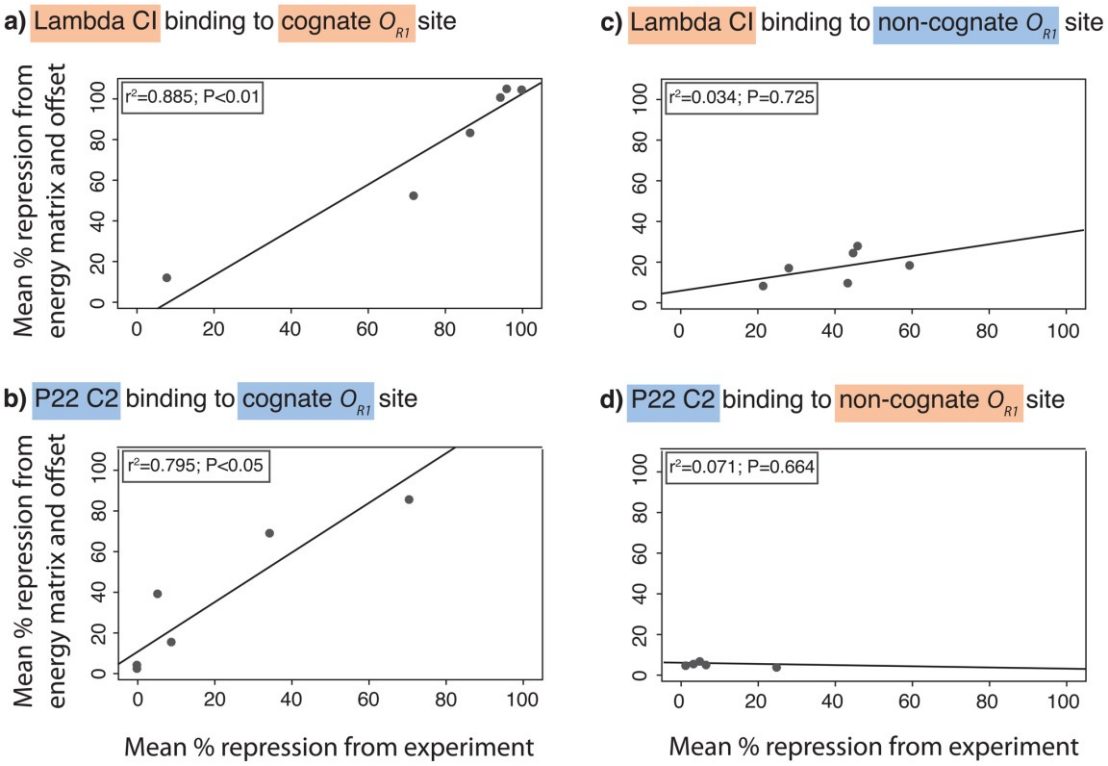

Figure 2.13. O_{R1} mutant fit between mean repression values obtained from the sigmoid curve (energy matrix, E_{seq} + offset, E_{WT}) and experimental data.

Mean repression values were calculated for each mutant class directly from mean energy values, by locating them on the sigmoidal repression curve (Gene expression $= \frac{1}{1+e^{E_{tot}-\mu}}$*, with* $E_{tot} = E_{WT} + E_{seq}$*), instead of using the full model (shown in Fig. 2.5). The correlation between these repression values and the experimentally measured means per mutant class was determined through a linear regression for:* ***a)*** *Lambda cognate mutants;* ***b)*** *P22 cognate mutants;* ***c)*** *Lambda non-cognate mutants; and* ***d)*** *P22 non-cognate mutants.*

Role of inter-operator epistasis

We investigated experimentally if promoter architecture — the existence of multiple operator sites — can affect the observed trade-off between robustness/evolvability and tunability. We first tested the effects of mutating four residues in the Lambda cognate O_{R2} (Table 2.4). The effects of mutations in O_{R2} on repression (Fig. 2.14a) were modest (75-100% repression), but less robust than mutations in O_{R1} (comparing Fig. 2.14a to Fig. 2.2b top panel), despite the supposedly weaker influence of O_{R2} on repression[2].

We tested for interactions between mutations in two operators (inter-operator epistasis) by creating a cognate library with mutations in both O_{R1} and O_{R2}. Because the trade-off between high robustness and low tunability was observed only in Lambda CI, we focused only on inter-operator epistasis in the cognate Lambda system. We randomly selected three neutral O_{R1} mutants, and combined each with eight randomly selected O_{R2} mutants (Table 2.1, 2.4). We observed a wider spectrum of repression values (40-80%), and hence higher tunability, among these mutants (Fig. 2.14b) compared to mutations in individual operators (Table 2.5). This meant that mutations in O_{R2} exacerbate the effects of phenotypically neutral O_{R1} mutations, indicating pervasive inter-operator epistasis (Table 2.6). Inter-operator epistasis arising from multiple mutations in both operators could not be captured by the thermodynamic model (Fig. 2.15), which is in contrast to a previous study where we introduced only a single point mutation into each operator[106]. However, the findings we report here are in line with studies showing that the presence of multiple operators can obstruct sequence-based predictions of gene expression[107].

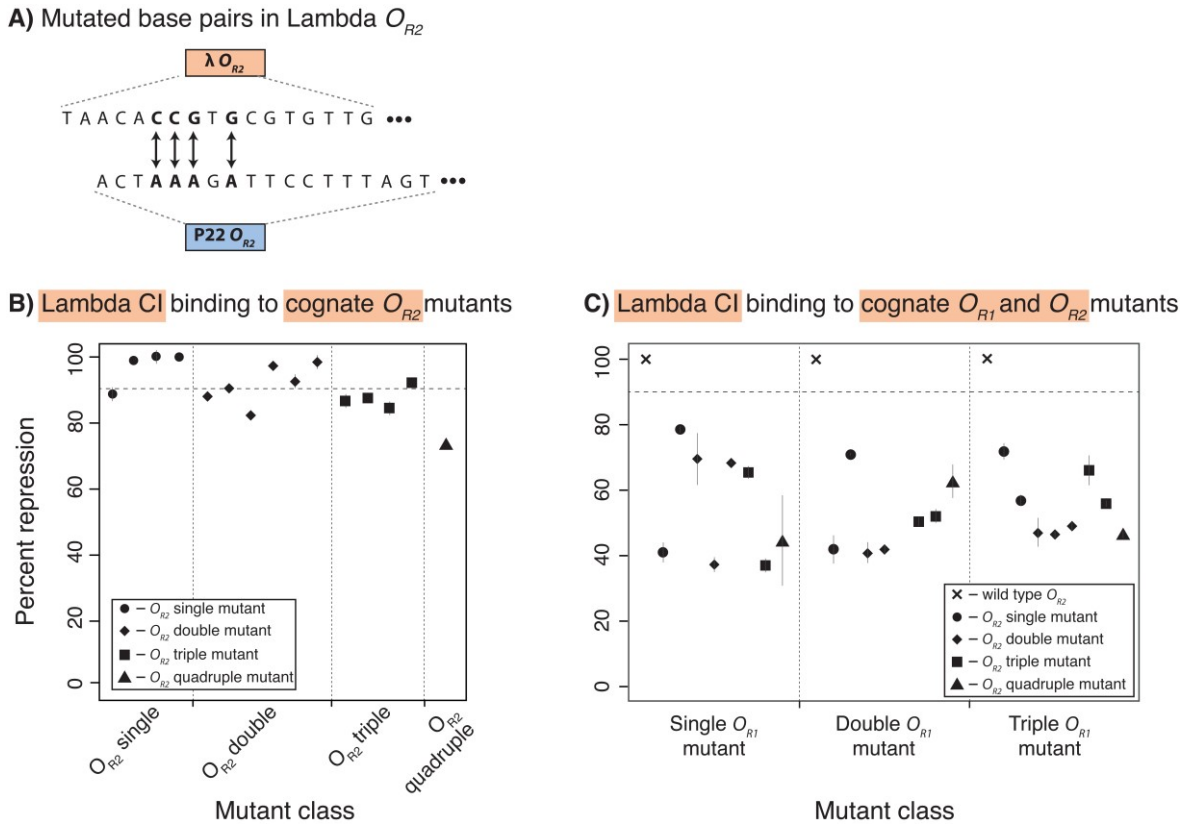

Figure 2.14. Inter-operator epistasis alleviates the trade-off between robustness and tunability.

***a**) Homology alignment of Lambda and P22 O_{R2}, showing mutated sites in bold. Arrows show base pairs that were exchanged between the two operators (Table 2.4). Loss of Lambda CI binding due to mutations in **b**) cognate O_{R2}; **c**) both cognate sites. Points are mean percent repression of three replicates, bars are standard errors of the mean. Plot symbols indicate O_{R2} mutant class. 'x' symbols correspond to the operator with the given O_{R1} mutation(s) and the wild type O_{R2} sequence (Fig. 2.3b). One O_{R1}-O_{R2} mutant gave no measurable expression in the absence of repressor and is not shown.*

Inter-operator epistasis alleviated the trade-off between robustness and tunability for Lambda CI in O_{R1}, likely by effectively modifying cooperative repressor binding. This role of inter-operator epistasis could be specific to operators that are functionally connected through cooperative binding, and might be different for redundant operators. Our results suggest that for cooperative binding, additional operators can facilitate network rewiring, as inter-operator epistasis helps generate expression level diversity, while maintaining robustness to the existing operators.

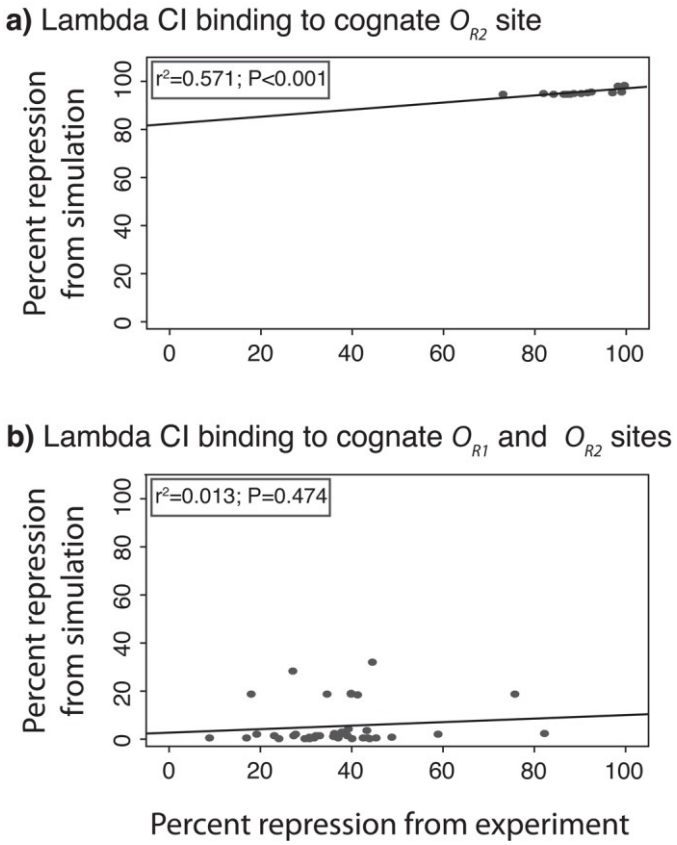

Figure 2.15. Thermodynamic model fit for the Lambda O_{R2} and the combined Lambda O_{R1} - O_{R2} mutant libraries.

Linear regression between experimental and model-derived percent repression of each experimentally tested mutant (from Fig. 2.14) is shown for Lambda CI binding to cognate ***a)*** *O_{R2} mutants;* ***b)*** *operators with mutations in both O_{R1} and O_{R2}. In order to evaluate the best possible performance of the model, here we show model estimates using parameter values that were fitted from the data (for repressor and RNAP concentrations, as well as offset and cooperativity).*

2.4 Discussion

The principles that govern gene regulatory evolution, which have been studied primarily from a global network perspective, remain poorly understood. Here, we identify the biophysical mechanisms that determine the evolutionary potential of

transcription factors for rewiring of regulatory network connections. Specifically, we provide an analytical expression (Fig. 2.12a) that, under reasonable assumptions, correlates robustness, tunability and evolvability (as defined in this study). Indeed, we experimentally observed these correlations for two closely related repressors: Lambda CI is more robust and at the same time more evolvable, while P22 C2 is more tunable. These differences in mutational effects likely arise from differences in specific DNA binding mechanisms[108]: while the binding specificity of Lambda CI is mostly based on direct contacts between operator bases and amino acid residues[97], the affinity of P22 C2 relies strongly on the local DNA conformation[98,109]. The nonlinear relationship between binding energy and repression, which is inherent to the thermodynamic model[110] (Fig. 2.3), captures the differences in robustness, tunability and evolvability, explaining how the intrinsic binding characteristics of a repressor determine its evolutionary potential for regulatory rewiring (Fig. 2.12a). The model does so by representing the evolutionary potential for each repressor through its total binding energy (offset E_{WT} plus energy matrix E_{seq}) and the average effect size of mutations (given by the energy matrix). Typically, energy matrices are used to determine and predict binding of TFs to a given DNA sequence[111]. However, our findings imply that the composition of the energy matrix crucially determines not only the current regulatory structure, but also the potential of the repressor to contribute to GRN evolution through making and breaking of individual connections. It is worth noting that while we only considered steady state expression levels, operator mutations could also affect expression dynamics, which might be subject to different constraints.

The *in vivo* positive correlation between robustness and evolvability is surprising, as molecular systems that are more persistent in the face of mutational pressure are generally assumed to be less likely to acquire novel functions[29]. Previous theoretical studies attempted to resolve this paradox by describing how robustness and evolvability 'emerge' as properties of existing networks[30,76,112,113], but so far, direct experimental approaches have been missing. We experimentally resolve this apparent paradox by showing that local mechanisms of TF-DNA binding intrinsically correlate robustness and evolvability in a positive manner. In fact, this positive correlation can be explained through an analytical expression that shows how robustness and evolvability are connected through the mutational distance between the cognate operator and a random DNA sequence (Fig. 2.12a). As such, a more promiscuous TF is simultaneously more robust and more evolvable, retaining cognate binding more easily while facilitating acquisition of novel operator sites. The positive correlation between robustness and evolvability can facilitate GRN evolution[82] by enabling a neutral network of genotypes, throughout which mutations have small phenotypic consequences[30,76]. Lambda CI is known to be promiscuous, showing nonspecific binding across the *E. coli* genome[58] and to non-cognate phage operators[114]. Thus, a Lambda CI-like TF has a higher potential to become a global regulator, whereas a P22 C2-like TF would be more suited as a local regulator, since its easy loss of binding could facilitate rewiring by reducing detrimental crosstalk[115]. However, the same biophysical mechanisms can impose a trade-off between evolvability and tunability, thus constraining the range of expression levels that can be achieved by a promiscuous TF at a single operator.

Given the key role that rewiring of local regulatory connections plays in changing GRN structure, the scarcity of direct experimental approaches studying the mechanisms of rewiring is striking. Our work provides a mechanistic link between the biophysics of TF-DNA binding and GRN evolution. Epistatic interactions, which emerge through the presence of multiple operators and alleviate the trade-off between tunability and robustness/evolvability, can prevent a straightforward prediction of how local rewiring properties determine global network evolution. Moreover, the binding landscape for regulatory rewiring we describe is based purely on biophysical characteristics that connect genotype (mutations) to phenotype (gene expression levels), which will be further shaped by selection forces acting on this landscape[91,92,116]. By integrating biophysical models with the existing molecular knowledge of regulatory elements, our work provides the first steps towards a quantitative mechanistic framework for understanding gene regulatory network evolution.

2.5 Tables

Table 2.1. Identity of O_{R1} mutants used in experiments.

Only the positions in the O_{R1} site that were mutated (as seen in Fig. 2.1d) are shown. The identities of introduced mutations are indicated by an 'X'. Mutants that showed no expression in the absence of repressor in the Lambda cognate background are marked in red. Orange shading indicates additional mutants made in the cognate Lambda background (see Methods section 'Construction of mutant O_{R1} libraries'). Grey shading indicates O_{R1} mutants that were used in the construction of the Lambda cognate library with mutations in both sites (each of the three O_{R1} mutants was combined with each of the eight O_{R2} mutants given in Table 2.4).

	O_{R1}						
Lambda P_R mutated positions	A	C	T	G	C	G	A
P22 P_R mutated positions	T	T	A	A	T	C	T
Single O_{R1} mutants	-	X	-	-	-	-	-
	-	-	X	-	-	-	-
	-	-	-	X	-	-	-
	-	-	-	-	X	-	-
	-	-	-	-	-	X	-
	-	-	-	-	-	-	X
Double O_{R1} mutants	-	X	X	-	-	-	-
	-	-	-	X	X	-	-
	-	-	-	X	-	X	-
	-	-	-	-	X	-	X
	-	X	-	-	X	-	-

	- - X - - X -
Triple O_{R1} mutants	- X X X - - -
	- - X X - X -
	- - X - - X X
	- X - - X - X
	- X X - - X -
	- - - X X X -
	- X - X X - -
Quadruple O_{R1} mutants	- X X - X X -
	- - - X X X X
	- X X X X - -
	- - X X - X X
	- X X X - X -
	- X - X X X -
Quintuple O_{R1} mutants	- X X X - X X
	- X - X X X X
	- X X X X X -
	X X X - X X -
	X X X X X - -
	X X - X X X -
Sextuple O_{R1} mutants	- X X X X X X
	X X X X X X -

Table 2.2. Normalized expression levels in the absence of repressor.

Mean and standard deviation for expression levels of three replicates for each cognate Lambda O_{R1} and P22 O_{R1} mutant. For easier comparison, each measurement was normalized by the Lambda P_R wild type fluorescence, as the wild type Lambda and P22 P_R promoters have different expression levels, and most P22 O_{R1} mutants increased fluorescence above the wild type level. Lambda O_{R1} mutants that had no expression in the absence of CI are not shown.

Lambda O_{R1} mutants		**P22 O_{R1} mutants**	
Mean wt-normalized expression	Standard deviation	Mean wt-normalized expression	Standard deviation

Single mutants			
0,852222	0,145704	0,383942	0,041988
0,877816	0,092266	0,429844	0,055555
0,647231	0,105471	0,408099	0,039533
0,70991	0,20474	0,271347	0,138469
0,874034	0,102582	0,447254	0,157922
		0,396608	0,144001
Double mutants			
0,805726	0,112751	0,416264	0,028769
0,940418	0,011987	0,318855	0,03031
0,913362	0,075526	0,723779	0,194892
0,44648	0,177222	0,375827	0,056349
0,500238	0,046582		
Triple mutants			
0,761709	0,170817	0,376119	0,016535
0,852082	0,136546	0,338433	0,032626
0,901182	0,108706	0,296501	0,044589
0,455281	0,061705	0,513411	0,092668
0,486204	0,049246	0,529408	0,145505
Quadruple mutants			
1,070794	0,117644	0,517823	0,033796
0,9803	0,096567	0,299942	0,05575
0,054822	0,003222	0,31519	0,038923
0,483373	0,063435	0,313736	0,105104
0,50058	0,014603		
Quintuple mutants			
0,052591	0,005129	0,34829	0,084974
0,962642	0,136703	0,287203	0,100048
0,52376	0,008103	0,510868	0,079894
0,503235	0,015616		
0,663547	0,047102		
Sextuple mutant			
0,536111	0,008606	0,289135	0,089228

Table 2.3. Parameter values used in the thermodynamic model.

All parameters were selected from the literature, except repressor concentrations, which were fitted to experimental data (see Methods).

Parameter	Parameter value
Lambda P_R promoter strength	50[118]
P22 P_R promoter strength	40[118]
Lambda CI dimer cooperativity	5 kcal/mol[12]
P22 C2 dimer cooperativity	3 kcal/mol[12]
Lambda CI concentration	3 µM (this study)
P22 C2 concentration	1 µM (this study)
Chemical potential due to non-specific binding	-7,4 kcal/mol[12]
RNAP concentration	3 µM[118]
Binding affinity of Lambda CI to wt O_{R1}	-13,2 kcal/mol[119]
Binding affinity of Lambda CI to wt O_{R2}	-11,7 kcal/mol[119]
Binding affinity of P22 C2 to wt O_{R1}	-12 kcal/mol[98]
Binding affinity of P22 C2 to wt O_{R2}	-10 kcal/mol[98]
Binding affinity of RNAP binding to P_R	-12,5 kcal/mol[63]

Table 2.4. Identity of Lambda O_{R2} mutants used in experiments.

Only the mutated positions in the O_{R2} site (as shown in Fig. 2.1d) are shown. The identities of introduced mutations are indicated by an 'X'. Grey shading indicates O_{R2} mutants that were used in the construction of the Lambda cognate library with mutations in both sites (each of the three O_{R1} mutants shown in grey in Table 1 was combined with each of the eight O_{R2} mutants). We only created O_{R2} mutants in the Lambda background.

	O_{R2}
Lambda P_R mutated positions	C C G G
P22 P_R mutated positions	A A A A
Single O_{R2} mutants	X - - -
	- X - -
	- - X -
	- - - X

Double O_{R2} mutants	X X - -
	X - X -
	X - - X
	- X X -
	- X - X
	- - X X
Triple O_{R2} mutants	X X X -
	X X - X
	X - X X
	- X X X
Quadruple O_{R2} mutants	X X X X

Table 2.5. Mutations in Lambda O_{R2} increase the variance in repression of O_{R1} mutants.

Variance between the three Lambda O_{R1} mutants (one single, one double, and one triple mutant, marked with grey shading in Table 2.1) was calculated on the wild type and each of eight O_{R2} mutant backgrounds. Red marks O_{R2} mutant backgrounds that significantly increase standard deviation between the three O_{R1} mutants, compared to the wild type O_{R2} background, calculated with an F-test for equality of variances. Identities of O_{R2} mutants correspond to Table 2.4.

O_{R2} background	wild type	single 1	single 2	double 1	double 2	double 3	triple 1	triple 2	quadru-ple
Std. deviation	10^{-14}	0,031	0,012	0,023	0,001	0,019	0,01	0,010	0,010

Table 2.6. Statistical significance of epistasis between mutations in Lambda O_{R1} and O_{R2}.

The left-hand column indicates the specific combination of O_{R1} and O_{R2} mutations (numbers of different mutants correspond to grey shaded mutants in Tables 2.1, 2.4). FDR-corrected t-tests were carried out to test if epistasis of each $O_{R1} - O_{R2}$ mutant was significantly different from the multiplicative expectation based on single mutant effects (shown in red).

$O_{R1} - O_{R2}$ mutant identity	**p-value**	**t-value (6 d.f.)**
Single O_{R1} mutant (#2) - Single O_{R2} mutant (#1)	2,06E-05	-272,715
Single O_{R1} mutant (#2) - Single O_{R2} mutant (#4)	1,67E-06	-1513,59
Single O_{R1} mutant (#2) - Double O_{R2} mutant (#1)	0,0045	-15,1928
Single O_{R1} mutant (#2) - Double O_{R2} mutant (#3)	1,01E-05	-435,482

Single O_{R1} mutant (#2) - Double O_{R2} mutant (#6)	5,26E-08	-14784,2
Single O_{R1} mutant (#2) - Triple O_{R2} mutant (#2)	1,95E-05	-289,947
Single O_{R1} mutant (#2) - Triple O_{R2} mutant (#3)	7E-06	-546,572
Single O_{R1} mutant (#2) - Quadruple O_{R2} mutant	0,023241	-6,44467
Double O_{R1} mutant (#2) - Single O_{R2} mutant (#1)	7,19E-05	-133,29
Double O_{R1} mutant (#2) - Single O_{R2} mutant (#4)	5,26E-08	-16929,5
Double O_{R1} mutant (#2) - Double O_{R2} mutant (#1)	2,38E-05	-245,52
Double O_{R1} mutant (#2) - Double O_{R2} mutant (#3)	1,97E-06	-1292,91
Double O_{R1} mutant (#2) - Triple O_{R2} mutant (#2)	2,1E-06	-1102,63
Double O_{R1} mutant (#2) - Triple O_{R2} mutant (#3)	1,29E-05	-370,905
Double O_{R1} mutant (#2) - Quadruple O_{R2} mutant	0,003547	-17,5286
Triple O_{R1} mutant (#7) - Single O_{R2} mutant (#1)	6,4E-05	-145,371
Triple O_{R1} mutant (#7) - Single O_{R2} mutant (#4)	2,88E-06	-894,356
Triple O_{R1} mutant (#7) - Double O_{R2} mutant (#1)	9,99E-05	-110,067
Triple O_{R1} mutant (#7) - Double O_{R2} mutant (#3)	1,63E-06	-1879,73
Triple O_{R1} mutant (#7) - Double O_{R2} mutant (#6)	1,63E-06	-1944,52
Triple O_{R1} mutant (#7) - Triple O_{R2} mutant (#2)	0,00042	-52,3063
Triple O_{R1} mutant (#7) - Triple O_{R2} mutant (#3)	1,67E-06	-1639,08
Triple O_{R1} mutant (#7) - Quadruple O_{R2} mutant	2,1E-06	-1122,88

2.6 Methods

Strains and plasmids

The experimental system is based on the 'genetic switches' of the bacteriophages Lambda and P22, which have similar regulatory architecture and substantial structural homology due to shared ancestry[95]; specifically we use the P_R promoter system. We constructed a template plasmid consisting of two parts that are separated by 500 random base pairs and a terminator sequence (represented by a hairpin structure in Fig. 2.1b): an inducible repressor gene on one strand and a regulatory region controlling a fluorescence marker on the other strand. Either Lambda CI or P22 C2 were placed after an inducible P_{TET} promoter. The fluorescent protein gene *venus-yfp*[120] was placed under the control of the P_R regulatory promoter region, containing an RNAP binding site as well as two operators, O_{R1} and O_{R2}, either from Lambda or P22. Specifically, for Lambda P_R we used the region from -60bp upstream of the transcriptional start site to +9bp downstream. To our knowledge the specific location of the transcriptional start site for P22 P_R has not been defined. Therefore, upstream of O_{R2} and downstream of O_{R1} we used the wild type P22 sequence that was of the same bp length as the analogous Lambda P_R regions. This meant that we used the wild

type P22 sequence from -65bp upstream up to the start codon of *cro*. O_{R1} more strongly binds the repressor and is in direct overlap with the RNAP binding site (-10). O_{R2} has a weaker affinity for the repressor, and assists in repression mainly through cooperative binding between two repressor dimers[121]. Downstream of the phage sequences both promoter regions contain the same ribosomal binding site in front of the reporter gene. These parts were cloned in all four combinations (cognate combinations: Lambda *cI* with Lambda P_R, and P22 *c2* with P22 P_R; non-cognate combinations: Lambda *cI* with P22 P_R, and P22 *c2* with Lambda P_R) into a low copy number plasmid (pZS*) containing a kanamycin resistance marker[99]. The TL17 terminator sequences followed the repressor genes, and the T1 terminator the *venus-yfp* (Fig. 2.1b). The plasmid libraries were then transformed into MG1655 derived *E. coli* cells (strain BW27785, CGSC#: 7881)[122].

Construction of mutant O_{R1} libraries

We created a library of mutants in O_{R1} by selecting six base pairs that were found to be most important for the binding of either of the two repressors[97,98], and that differed between Lambda and P22 O_{R1} sequences. This was done by aligning the O_{R1} sites from Lambda and P22 wild type operators (according to homology, not symmetry) and comparing the corresponding base pairs in the operator sites. The six base pairs that were most important for repressor binding and that differed between the two operators were substituted by the base pairs of the non-cognate O_{R1} in both directions: starting with wild type Lambda O_{R1} and mutating it to be more similar to P22 O_{R1}; as well as starting with wild type P22 O_{R1} and mutating it to be more similar to Lambda. We generated all six single mutants, four double, five triple, four quadruple, three quintuple, and the sextuple mutant. For mutating Lambda O_{R1} from cognate to non-cognate, ten additional mutants were constructed that did not contain mutations in base pairs overlapping the -10 binding region of RNAP: two double, two triple, two quadruple, three quintuple, and another sextuple mutant. For the quintuple and sextuple mutants an additional base pair was chosen, that was linked to high affinity binding of Lambda CI (Table 2.1). The additional double and triple mutants were also created for the P22 non-cognate library. O_{R1} operator libraries were constructed by synthesizing oligos of 73bp length (Sigma Aldrich), carrying wild type O_{R2} and mutated O_{R1} (Table 2.1), and cloning them into the experimental system plasmid backbone (Fig. 2.1b). Clones carrying correct mutants were confirmed through Sanger sequencing.

We also tried to construct promoter regions containing cognate O_{R1} and non-cognate O_{R2}. As both operators contain parts of the RNAP binding site, we did not obtain fluorescence expression in the absence of CI from these promoters even when we varied the spacing between the operators. This is possibly due to factors other than sequence-dependent binding energy playing a role in the regulatory context of these promoters[107].

Fluorescence assays

We measured fluorescence of all O_{R1} mutants (Lambda and P22 cognate and non-cognate systems), both in the presence and in the absence of the inducer aTc. Three biological replicates of each mutant of the library were grown at 37°C overnight in M9 media, supplemented with 0.1% casamino acids, 0.2% glucose, 30μg/ml kanamycin, and either without or with 15ng/ml aTc. Overnight cultures were diluted 1,000X, grown to OD_{600} of approximately 0.05, and their fluorescence measured in a Bio-Tek Synergy H1 plate reader. All replicate measurements were randomized across multiple 96-well plates. All measured mutants had fluorescence levels significantly above the detection limit of the plate reader, resulting in measurements at least 1.5-fold greater than the non-fluorescent control.

Fluorescence values were normalized by OD_{600} values (in RFU=Relative Fluorescence Units) and averaged over three replicates. Repression values were calculated as a normalized ratio between the measured fluorescence with and without the repressor:

$$Percent\ repression = \left(1 - \frac{RFU_{repressor}}{RFU_{no\ repressor}}\right) * 100.$$

Standard errors of the mean repression values were calculated using error propagation in order to account for the inherent variability in the fluorescence measurements. The fluorescence levels measured in the absence of repressor were comparable across all Lambda operator mutants, as well as all P22 operator mutants (Table 2.2). This means that the reported differences in percent repression arose mainly from changes in repressor binding, rather than alterations to the RNAP binding site. Moreover, our simulations showed that changes in RNAP concentration, which correlates with the strength of RNAP binding, do not change the qualitative pattern of binding for the two repressors. Interestingly, when compared to P22 wild type O_{R1}, all of the P22 cognate O_{R1} operator mutants showed increased expression levels in the absence of repressor. Lambda P_R is a stronger promoter than P22 P_R, and introducing mutations in the operator region of P22 P_R increased promoter strength by making it more similar to Lambda P_R.

Direct comparisons between the *in vivo* effects of operator mutations on gene expression level that we measured, and the previous published studies of the same operators[97,98] were hindered by the *in vitro* nature of previous studies. All previous studies of Lambda P_R and P22 P_R mutants relied on biochemical filter binding assays, which do not account for cooperativity between the two sites, and as such do not necessarily translate quantitatively into gene expression levels. As such, comparisons between published data and our data are possible only through a modeling framework, such as the one we utilize (see Materials and Methods section 'Thermodynamic model of repression at the P_R promoter').

For the experimental data, the evolutionary properties were calculated in the following way: robustness and tunability of the repressors were evaluated on the cognate operator mutants. Robustness for the experimental data was calculated as the percent of mutants for which >90% of the wild type repression was retained. Tunability was calculated as the standard deviation in repression levels for mutants

that exhibited between 10% and 90% of the wild type repression. On the cognate background, mutants that were repressed less than 10% were considered neither robust nor tunable. Evolvability was calculated as the portion of non-cognate mutants that were repressed to more than 10%.

Cellular concentrations of the two repressors were determined using Western blots. Lambda CI and P22 C2 were cloned with a His-Tag or an HA-Tag, respectively, at their carboxy-terminal end. Rat and rabbit primary antibodies (Roche and Thermo Fisher, respectively) in combination with Goat anti-rat and anti-rabbit secondary antibodies (Thermo Fisher) were used to detect them. Samples were processed once at full concentration and once at 2-fold dilution. The obtained bands from gel electrophoresis were normalized by a household gene and normalized concentrations between the two repressors were compared as ($\frac{concentration_{Lambda\ CI}}{concentration_{P22\ CI}}$). Lambda CI was present in excess over P22 C2: 3.8-fold for full concentration samples and 5.5-fold for diluted samples. We also tested variation in repressor levels by measuring fluorescence from the P_{TET} promoter on the same plasmid construct as used in the library measurements for 6 replicates either without or with 15ng/ml aTc and found only minor variability (without aTc: 3.6% CV, with aTc: 2% CV) that cannot explain the experimentally observed differences between the repressors.

Thermodynamic model of repression at the P_R promoter

The model is based on previously described thermodynamic approaches[51,57], which rely on several assumptions: (i) TF binding to DNA takes place at thermodynamic equilibrium; (ii) gene expression can be equated with the probability of binding of participating proteins (in our case RNAP and repressor); and (iii), the contribution of each base pair in the operator to binding is additive. The probability of a gene being expressed is derived by summing the Boltzmann weights over all promoter occupancy states where RNAP is bound. Boltzmann weights are given by $w_i=[N]* e^{(E_{tot}-\mu)}$, where E_{tot} is the energy of a certain configuration, N is the molecule concentration (in μM), and μ is the chemical potential. E_{tot}, the total binding energy, is composed of the offset (E_{WT}), which is the energy of binding to a reference (wild type) sequence; and the binding energy derived for a specific sequence from the energy matrix of the binding protein $E_{seq}=\sum_{i=1}^{l} \epsilon_i(a_i)$, where l is the length of the sequence, a_i the specific nucleotide at position i, and ϵ_i the energy contribution due to the energy matrix of the specific nucleotide a at position i. Total binding energy is therefore $E_{tot}= E_{WT}+ E_{seq}$. Binding energies and chemical potential are given in *kcal/mol*. In our model system, there are two operator sites (O_{R1} and O_{R2}) that can each be occupied by a repressor dimer, and binding to each operator site is affected by the strength of cooperative binding between them. The probability of the gene being expressed is then given by the sum of all states conducive to promoter expression (RNAP bound) normalized by the sum over all possible states:

$$Gene\ expression = \frac{1}{1 + \frac{K_p}{[RNAP]} * \frac{\left(1 + 2\frac{[R]}{K_R} + \left(\frac{[R]}{K_R}\right)^2 e^{\omega}\right)}{\left(1 + \frac{[R]}{K_R}\right)}}$$

, where $K_x = e^{(E_{tot,x} - \mu)}$ represents the effective equilibrium dissociation constant (relative to the genomic background) – which is the concentration for half-maximal occupation of the site - of, either RNAP (K_P) or the repressor (K_R). Please note that we account for concentration-specific effects separately and μ incorporates only non-specific background binding and other unspecific cellular effects. The probability of transcription factor (TF)–DNA binding is of the form[84]: $p_i = \frac{[TF_i]/K_i}{1 + [TF_i]/K_i}$. Based on Garland (2002), we can assume that K_x is individually tunable for each binding site. [R] is the concentration of repressor dimers, which is the effective concentration, as repressors only bind as dimers and, as we assume fast dimerization[123], this corresponds to half of the total monomer concentration in the cell. [RNAP] is the concentration of RNAP, and ω is the cooperativity energy value, describing the strength of interaction between two repressor dimers. All concentrations and dissociation constants are given in units of μM. The calculated gene expression value is a relative measure, with 1 indicating full expression and 0 no expression. Percent repression was then calculated using the formula:

$$Percent\ repression = \left(1 - \frac{gene\ expression_{repressor}}{gene\ expression_{no\ repressor}}\right) * 100.$$

In the 'main model', which is used throughout the study, RNAP competes with repressor binding at O_{R1}, and repressor binding to O_{R1} is increased by cooperative binding of a second dimer to O_{R2}. Therefore, the following scenarios are possible: (i) the promoter can be bound by neither protein; (ii) RNAP can be bound either alone or together with repressor at O_{R2}; or (iii) repressor bound to O_{R1} keeps RNAP from binding, either by binding on its own or cooperatively together with another repressor at O_{R2}. The corresponding formula was taken from Bintu et al.[57] (Case 4). We also considered an 'alternative model' where O_{R2} binding impedes RNAP binding as well (Bintu et al[57]; Case 6), but as the main model always gave a better fit to experimental data, we utilized only the main model throughout.

Energy values for binding to mutated sequences were calculated for RNAP and repressor binding using the respective energy matrices by adding up the individual relative contributions of each base pair and adding an offset. The offset is the energy of binding of the repressor to the wild type sequence, which was added because the energy matrix calculates only energy differences relative to wild type binding. Binding energy matrices were based on Sarai & Takeda (1989) for Lambda CI, on Hilchey et al. [98] for P22 C2 - which were both determined biochemically - and, for RNAP, on an ongoing work on RNAP binding to Lambda P_R within the group. Wild type binding affinities of Lambda CI to both operators (offset) were taken from Vilar[119]. Other model parameters were taken from the following sources: binding cooperativity and nonspecific binding energy were adopted from Hermsen et al.[12]; wild type binding

affinities for both operators were obtained from Hilchey et al.[98] for P22 repressor; and binding energy and concentration for RNAP were taken from Santillan & Mackey[63]. Promoter strength for both Lambda P_R and P22 P_R was based on previously published values for the Lambda P_L promoter[118], but we also found that the results were not sensitive to this parameter. Repressor dimer concentrations were the only parameters that were fitted to the data by means of a Monte Carlo algorithm. The algorithm used simulated annealing to find the optimal parameter values minimizing the squared difference between the predicted and observed percent repression between the data and the model. The fitted difference in concentration values between the two repressors is slightly lower than found experimentally (Fig. 2.4). We tested the model for concentration values from 0- to 7-fold difference, and always found the same trends in the evolutionary potential (Fig. 2.11). Note that standard experimental measures cannot provide effective TF concentrations (i.e. proteins that are free to bind at the target site), especially when two TFs are not equally promiscuous, as these measures cannot distinguish free and non-specifically bound proteins. Because of this, and because the overall differences in evolutionary potential did not depend on variations in repressor concentration parameters, we used repressor concentrations determined by the best model fit, and not those we experimentally measured. All parameter values used in the model are shown in Table 2.3.

In order to verify the fit of our model to the experimental data, linear regression was performed between the data obtained experimentally (see *Fluorescence assays*) and the prediction of repression values produced through the thermodynamic model. Matlab R2015a software was used to calculate the regression, R squared and P-values for the O_{R1} library (Fig. 2.5). The model accurately reproduced experimental observations in cognate mutants, but did not fit non-cognate mutant measurements (Fig. 2.5). The lack of fit to non-cognate mutants is not surprising, as thermodynamic models assume an independent contribution of each position, which does not hold when mutated far away from the wild type operator sequence[101,119]. Nevertheless, because the model provided a lower bound on the experimentally measured non-cognate repression levels (Fig. 2.5), we used it to explore parameters affecting repression at non-cognate sites as well.

Robustness

Robustness was calculated for repressors binding to cognate mutants only if they retained more than 20% repression. We counted the number of robust neighbors for each operator, where 'robust neighbor' refers to an operator sequence that is exactly one mutation away from the reference and exhibits more than 90% repression of the reference repression value. Specifically, starting from the wild type, each mutant (above the 20% repression threshold) was taken as a reference and repression of all other mutants that are exactly one mutation away was calculated. The relative count of robust neighbors was averaged for each reference operator and the mean was taken over each mutant class. This procedure was repeated with different values for cooperativity (1,3,5,7 kcal/mol), repressor concentration (1,3,5,7 μM) and RNAP concentration (1,3,5,7 μM). We tested if the results were sensitive to the percent repression thresholds by calculating robustness for 80% and 95% thresholds, and

found no qualitative differences. For comparison with the experimental data and the definition of robustness used there, we also calculated robustness as the percent of all mutants for which >90% of the wild type repression was retained.

Tunability

Tunability was determined for repressor binding to cognate mutants with repression values between 10% and 90%, as the standard deviation over those mutants for each mutant class. Tunability was calculated for different values of cooperativity (1,3,5,7 kcal/mol), repressor concentration (1,3,5,7 μM) and RNAP concentration (1,3,5,7 μM). We tested if the results were sensitive to the percent repression thresholds by calculating tunability for 5% and 20% lower, as well as 80% and 95% upper threshold bound, and found no qualitative differences.

Evolvability

Evolvability was calculated for repressor binding to non-cognate mutants exceeding a threshold of 10% repression. For each mutant class the number of mutants above the threshold was counted and averaged. This procedure was repeated with different values for cooperativity (1,3,5,7 kcal/mol), repressor concentration (1,3,5,7 μM) and RNAP concentration (1,3,5,7 μM). We tested if the results were sensitive to the percent repression thresholds by calculating evolvability for 5% and 20% thresholds, and found no qualitative differences.

Evolvability on random operators

The promoter region for the random sequence library was based on the *lac* operon[117], because the binding sites for RNAP and repressor do not overlap in this system, thereby avoiding unwanted modifications of RNAP binding by an introduction of a random operator. Binding affinities for RNAP were calculated for this system using the energy matrix from Kinney et al., 2010. For the operator sites, 1,000,000 random 17bp-long sequences for Lambda CI, and 18bp-long sequences for P22 C2 were created in Matlab R2015a. The 1bp difference in the length of the sites used for the two repressors corresponds to the actual length of their respective cognate operator sites. Binding affinities to these operators were calculated for Lambda and P22 repressors using their energy matrices.

Swapping model parameters of the two repressors and comparing evolutionary properties

We calculated robustness and tunability for Lambda CI after swapping the values for repressor concentration, cooperativity, and offset with the respective values for P22 C2. The values were calculated separately for each mutant class (number of mutations). We first swapped each parameter value individually, and then we swapped all three parameters with the values of P22 C2. For evolvability, only the values for repressor concentration and offset were swapped individually and simultaneously. The same simulations were done for P22 C2 with Lambda CI

parameters. For each evolutionary property, we used a linear regression to determine the R^2 value for the goodness of fit between the reference repressor with its wildtype parameter values, and the other repressor with the swapped parameter(s). Regression was carried out across the six mutant classes. The fact that swapping repressor concentrations did not reconcile the evolutionary potential of the two repressors provides further evidence that the experimentally observed differences in the evolutionary potential between the two repressors (Fig. 2.2) could not be attributed solely to the measured differences in their intracellular concentrations (Fig. 2.4).

Relationship between binding energy and repression

The total binding energy (E_{tot}) is related to gene expression through:

Gene expression $= \frac{1}{1+[R]e^{E_{tot}-\mu}}$, with $E_{tot} = E_{WT} + E_{seq}$

, where μ describes the chemical potential of a repressor. The relationship between binding energy and repression is sigmoidal, with the position of the curve for a given repressor determined by μ and repressor concentration (which we set to 1 as we do not want to consider concentration effects here). The same chemical potential and repressor concentration was used for Lambda CI and P22 C2 and taken from Hermsen et al.[12]. The positions of a certain operator sequence for a specific repressor on the curve are then given by the total binding energy, E_{tot}, with concentrations for the two repressors being the same. We wanted to develop generic definitions of robustness, tunability and evolvability as properties of only the energy matrix and E_{WT}. The average effect size of one mutation (m) is determined by taking the average of the energy matrix for a given repressor (grand mean over the non-zero entries of the energy matrix, calculated in our example for the six mutated positions) and the deviation in mutational effects (σ) is calculated as standard deviation over all non-zero entries of the energy matrix. Robustness can then be defined as $Rob = \frac{E_{1/2}-E_{WT}}{m}$ and evolvability as $Evo = \frac{E_{1/2}-E_{random}}{m}$,where $E_{1/2}$ is the binding energy at half repression (50%) and E_{random} is the typical binding energy to a random sequence, which will be equal to non-specific binding above a certain number of mutations[101] and is from that point on independent of the energy matrix. Derivation shows that evolvability and robustness are correlated by the number of average mutations between the cognate operator binding energy and the binding energy of a random sequence (*#mut*), as m determines the positioning of E_{random} relative to E_{WT}: $Evo = \frac{E_{1/2}-E_{random}}{m} = \frac{E_{1/2}-(E_{WT}+\#mut*m)}{m} = Rob + \#mut$. This correlation depends critically on two assumptions. First, we assume that the typical mutational effect size (m) is relatively small compared to the offset (E_{WT}) and comparable between different repressors. We base this assumption on the notion that TF-DNA binding is determined by the strength of hydrogen bonds, which range between 1-3kcal/mol[24]. The second assumption is that the energy matrix is an intrinsic property of a repressor, meaning that it doesn't change depending on the DNA sequence that the repressor is binding to. In other words, we assume that m is constant across all binding sites, cognate and non-cognate. Tunability can be defined around $E_{1/2}$ as $Tun = (\sigma * \frac{d\ repression}{d\ binding\ affinity}|_{E_{1/2}})/Rob$,where $\frac{d\ repression}{d\ binding\ affinity}|_{E_{1/2}}$

gives the slope of the sigmoid curve at $E_{1/2}$. Positions on the curve for both repressors were calculated for binding to cognate operators, non-cognate operators and the operator with weakest possible binding (according to the energy matrix). Moreover, mean energy values for each mutant class were calculated from model simulations for the cognate and non-cognate operators and placed on the curve. Their locations on the curve provide mean repression values that were then compared to the experimental data through linear regression (Fig. 2.13). Matlab R2015a software was used to calculate the regression, R squared and P-values. The fit was similar to the one obtained using the full model (Fig. 2.5).

Lambda cognate O_{R2} mutant library

O_{R2} mutant operators were synthesized analogously to O_{R1} mutants. Based on the assumption that energy matrices between the two closely related operators are likely to be very similar, mutated base pairs in O_{R2} were chosen in positions corresponding to the mutations in O_{R1}. However, the last two were discarded as possibly interfering with RNAP binding (-35 region), leaving four base pairs for mutation (Fig. 2.2b). Four single, six double, four triple and the quadruple mutant were constructed in the Lambda cognate system and measured as described previously. The fit between data and model was determined through linear regression (Fig. 2.15a).

Lambda cognate O_{R1} - O_{R2} mutant library

O_{R1}-O_{R2} mutant operators were synthesized analogously to O_{R1} mutants, but with one to three mutations in O_{R1} and one to four mutations in O_{R2}. One single, one double and one triple O_{R1} mutant, that showed no decrease in repression, were combined with each of eight randomly selected O_{R2} mutants (two single, three double, two triple, and the quadruple). O_{R1}-O_{R2} mutant operators were constructed in the Lambda cognate system, as P22 C2 had very low robustness and hence no trade-off, and measured as described previously. The fit between data and model was determined through linear regression (Fig. 2.15b).

Calculation of epistasis in O_{R1}-O_{R2} mutants

We measured epistasis in two ways. First, through its effect on the tunability of the system, where we considered that a given combination of O_{R1}-O_{R2} mutations is in epistasis when the presence of mutations in both operators significantly increased the variance in the observed gene expression levels, compared to the variance achieved by mutations in O_{R1} alone. We compared the variance independently for each mutant class (number of mutations). Second, we calculated epistasis between mutations in the two operators as a deviation from the multiplicative expectation of double mutant repression level based on single mutant effects:

$$epistasis = \frac{percent\ repression_{O_{R1}-O_{R2}}}{percent\ repression_{O_{R1}} * percent\ repression_{O_{R2}}},$$

and conducted FDR-corrected two-tailed t-tests for each of the double mutants, to determine if epistasis was significantly different from the null multiplicative expectation (Table 6).

2.7 Author contributions

C.I. (Claudia Igler), M.L. (Mato Lagator), G.T. (Gašper Tkačik), J.P.B. (Jonathan P. Bollback), C.C.G. (Călin C. Guet) conceived the study together. C.I. and M.L. designed and carried out the experiments and analyzed the data. C.I. wrote the code and ran the model. C.I. and M.L. wrote the initial draft of the manuscript and revised it together with G.T. J.P.B and C.C.G.

3. Global crosstalk between transcription factors can enhance specificity

3.1 Abstract

Crosstalk in transcriptional regulation, which occurs when a transcription factor interferes with regulation at a non-cognate promoter, can have substantial consequences on the targeted cellular program. While these consequences have been explored theoretically, little is known about the mechanisms that determine how crosstalk affects gene expression levels in cells. We utilized components of bacteriophages Lambda and P22 to create synthetic systems in which two repressors have varying degrees of crosstalk at a promoter. We found that the presence of a second repressor elevates repression levels beyond what can be achieved by either of the repressors alone, indicating that crosstalk can enhance binding of transcription factors at their cognate binding site. Using a thermodynamic model of gene regulation, we found that global crosstalk in the form of binding competition at many weak non-cognate sites across the genome could elevate binding to the cognate site, and do so in a concentration-dependent manner. Our findings, which indicate that crosstalk can increase effective specificity, provide the first experimental insights into the mechanisms that determine how crosstalk with a non-cognate repressor can impact gene expression and, hence, organismal function.

3.2 Introduction

Specificity of molecular interactions is an essential property of all living systems, arising from the need for critical reactions to occur between cognate substrates even in the presence of high concentrations of similar non-cognate molecules. The widespread presence of proofreading mechanisms dedicated to ensuring correct molecular pairing, observed in diverse processes such as the correct matching between tRNAs and amino acids[124], immune system recognition of antigens[125], protein-protein interactions[126,127] and ligand sensing[128], points to the biological need to reduce unwanted 'crosstalk' – a generic term describing interactions between non-cognate substrates.

In transcriptional regulation - the primary mechanism controlling gene expression - specificity is achieved through the binding of a transcription factor (TF) to its DNA-binding site (operator). Yet, TF-operator specificity is far from perfect: while every TF preferentially binds a single unique operator sequence (cognate site), it is also capable of binding a range of related, less specific sequences[101,129,130]. Given this flexibility in TF binding, the short length of operator sequences (ranging from 5 to 35 base pairs), and the comparatively large genome sizes, non-cognate TF binding is a common occurrence[41,84]. In other words, crosstalk between a TF and non-cognate operators is the rule, not the exception.

Crosstalk in transcriptional regulation can have severe consequences for organismal fitness[131,132], as binding of a TF to non-cognate DNA sequences can lead to misregulation: repression of an essential gene, or expression of a gene in the wrong amount, or at a wrong time or place. Yet, in spite of the importance of crosstalk in determining fitness, experimental work on transcriptional regulation has been almost exclusively devoted to understanding how to achieve reliable cognate TF binding[133]. Here, we take the first steps in understanding the mechanisms that determine how crosstalk impacts gene regulation, by asking how the presence of an additional non-cognate TF affects gene expression levels at a promoter.

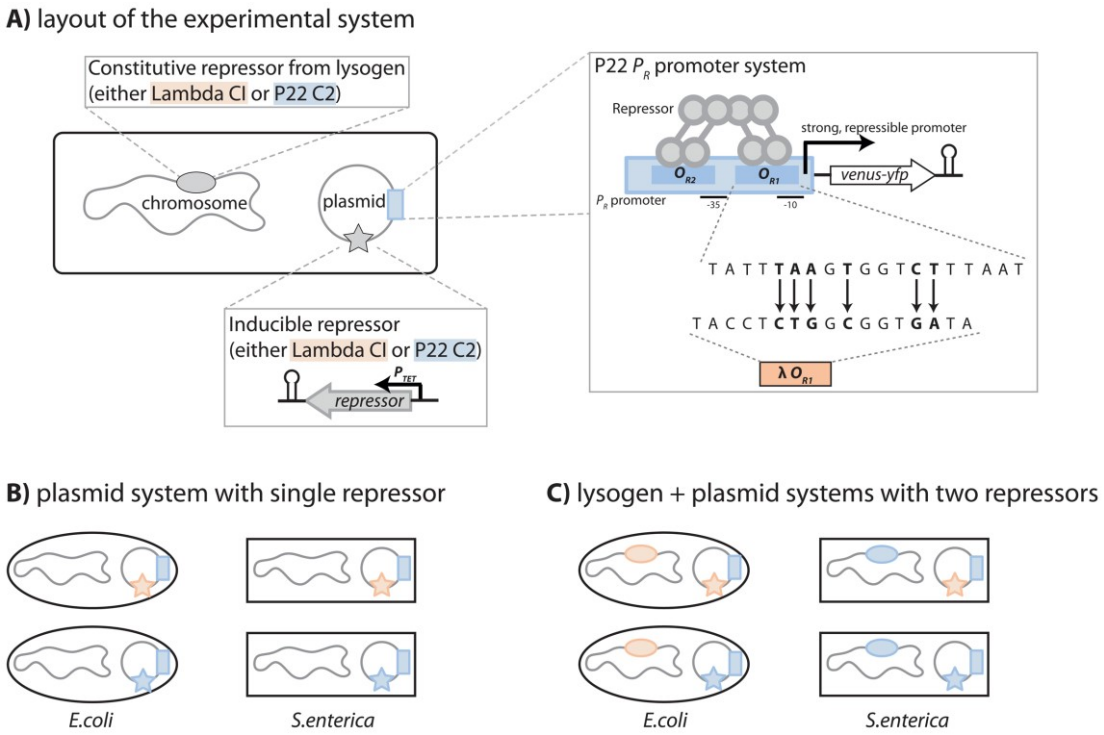

Figure 3.1. Experimental systems.

A) *Generic layout of the experimental system, showing the plasmid and the lysogen construct. The plasmid system contains either Lambda cI (orange) or P22 c2 (blue) repressor, under an inducible P_{TET} promoter. It also contains a yellow fluorescence marker under the control of the P22 P_R promoter, which consists of RNA polymerase binding site (marked with -10 and -35) and two operators, O_{R1} and O_{R2}. We mutated 6 positions in the P22 O_{R1} operator in various combinations (marked in bold) to be more similar to Lambda O_{R1} operator (Table 3.1), hence creating a mutant library exhibiting a range of binding to both repressors (blue rectangle represents this O_{R1} mutant library).* ***B)*** *Lysogen-free systems. Two plasmid systems, one with Lambda cI (orange star) and the other with the P22 c2 (blue star), were introduced into two hosts, E. coli and S. enterica.* ***C)*** *Lysogen systems. Lambda lysogen (orange oval) was introduced into E. coli, and P22 lysogen (blue oval) into S. enterica. Then, both plasmid systems (one with Lambda cI and the other with P22 c2) were introduced into both lysogen strains.*

3.3 Results

Single repressor systems

To explore the effects of crosstalk on local gene regulation, we utilized components of two canonical molecular systems – repressors and operators involved in the 'genetic switches' of bacteriophages Lambda and P22[96,121], which share common ancestry[95]. Specifically, we placed either Lambda *cI* or P22 *c2* repressors, under an inducible P_{TET} promoter, on a very low copy number plasmid[99]. The same plasmid carried the strong P22 P_R promoter, which contained P22 C2 cognate operators O_{R1} and O_{R2}, and controlled the expression of a yellow fluorescence protein *venus-yfp* (Fig. 3.1A). Binding of a repressor to O_{R1} is sufficient for repression, while P22 C2 binding to O_{R2} primarily serves to stabilize the binding to O_{R1} through cooperative binding between P22 C2 dimers[96,121]. Lambda CI and P22 C2 dimers do not show any appreciable cooperativity between them[134].

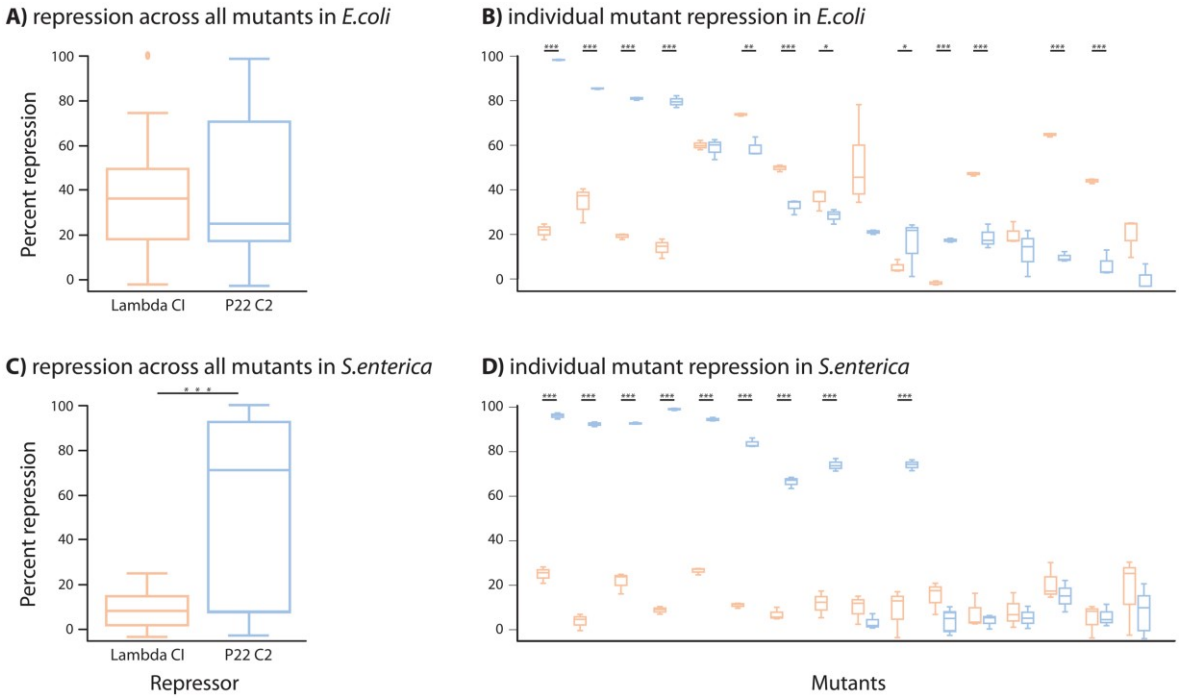

Figure 3.2. Repression in lysogen-free strains.

*Percent repression achieved by plasmid-borne Lambda CI (orange frames) or P22 (blue frames) in the lysogen-free systems (Fig.1B) is shown as box plots for **A)** all mutants in E. coli; **B)** individual mutants in E. coli; **C)** all mutants in S. enterica; **D)** individual mutants in S. enterica. Percent repression was calculated as* $\left(1 - \frac{RFU_{presence\ of\ inducer}}{RFU_{absence\ of\ inducer}}\right) * 100$. *One, two, and three stars indicate p-values of less than 0.05, 0.001, and 0.0001, respectively. Full statistical tests are shown in Table 3.2. Mutants are ordered according to Table 3.1.*

Drawing from our previous work[135], we introduced mutations into O_{R1} that enabled differential binding of Lambda CI and P22 C2 (Table 3.1). Then we used the two plasmid systems, one with Lambda *cI* and the other with P22 *c2*, to measure repression of the P22 P_R promoter mutants in two host species – *Escherichia coli* and *Salmonella enterica* (Fig. 3.1B). By measuring the difference in fluorescence in the absence and in the presence of either repressor, we observed that O_{R1} mutants exhibited a wide range of repression levels for each of the two repressors, albeit with a greater range in *E. coli* than in *S. enterica* (Fig. 3.2).

Non-cognate repressors increase repression more than cognate ones

In order to investigate how crosstalk – binding of two repressors at one operator – impacts overall repression, we introduced bacteriophage lysogens into their natural host species (Fig. 3.1A): Lambda lysogen was introduced into *E. coli*, and P22 lysogen into *S. enterica*. Lambda and P22 lysogens naturally produce CI and C2 repressors, respectively, resulting in strains that always have one of the repressors present, albeit at a relatively low concentration compared to plasmid expression. We then inserted the plasmid mutant libraries into these strains, giving rise to: *E. coli* strains that had either two copies of Lambda *cI* (*concentration effect*), or a copy of Lambda *cI* and a copy of P22 *c2* (*crosstalk*); and *S. enterica* strains with either a copy of P22 *c2* and a copy of Lambda *cI* (*crosstalk*), or two copies of P22 *c2* (*concentration effect*) (Fig. 3.1C).

We measured the difference in fluorescence before and after induction of the plasmid-borne repressor. In single repressor systems, such measurements gave the *total percent repression* of each O_{R1} mutant by a given repressor (Fig. 3.2), as it was the only repressor present in the system. In contrast, such measurements in lysogen strains gave the *additional repression* achieved by the induction of either the second repressor (*crosstalk*) or the second copy of the same repressor (*concentration effect*) - meaning the repression achieved on top of the one resulting from the lysogen-borne repressor, which is always expressed in the cell.

We observed that crosstalk – induction of the second repressor – resulted in significantly higher additional repression in both *E. coli* ($F_{1,47}$=55.45, P<0.0001) and *S. enterica* ($F_{1,47}$=92.23, P<0.0001), compared to inducing the second copy of the same repressor (Fig. 3.3). This meant that additional repression achieved by two repressors was not simply due to an increase in the overall concentration of the repressors in the system, but rather due to an interaction of the two repressors that resulted in higher overall binding at O_{R1}. This difference is best observed in *S. enterica*, where inducing Lambda CI alone has very low repression across all mutants (Fig. 2D), while inducing it when P22 C2 is already present achieves large additional repression in most mutants (Fig. 3.3D). Indeed, in 13/16 mutants additional repression (P22 C2 together with Lambda CI) is at least 10% greater than the percent repression achieved by Lambda CI alone, while the reverse is never the case. Our data indicates that, in both hosts, crosstalk enhances local repression.

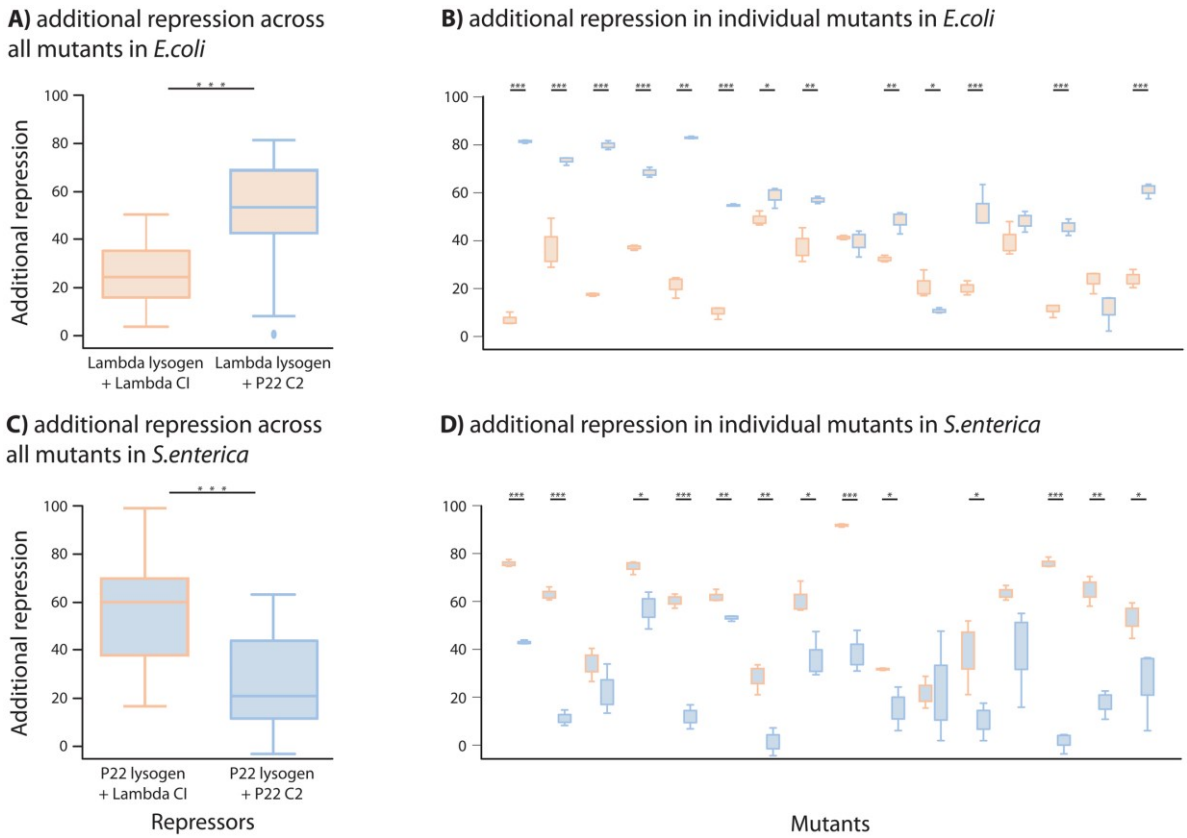

Figure 3.3. Repression in lysogen strains.

Additional repression achieved by inducing the plasmid-borne Lambda CI (orange frame) or P22 (blue frame) in the lysogen systems (Fig. 3.1C) is shown as box plots for ***A)*** *all mutants in E. coli;* ***B)*** *individual mutants in E. coli;* ***C)*** *all mutants in S. enterica;* ***D)*** *individual mutants in S. enterica. Lysogen strains are always expressing either Lambda CI (orange fill) or P22 C2 (blue fill), meaning that inducing a repressor from the plasmid adds either the second repressor or more of the same repressor that is already present in the cells. Additional repression was calculated as* $\left(1 - \frac{RFU_{presence\ of\ inducer}}{RFU_{absence\ of\ inducer}}\right) * 100$. *One, two, and three stars indicate P-values of less than 0.05, 0.001, and 0.0001, respectively. Full statistical tests are shown in Table 3.3. Mutants are ordered according to Table 3.1.*

Global crosstalk can enhance specificity

Why would the presence of a second repressor enhance overall repression? We already excluded the possibility that the *crosstalk-enhanced repression* arises from a concentration effect (Fig. 3.3). The observed results could also have arisen if there was a bias in how mutations in O_{R1} affected organismal fitness (measured as growth rates), but we found no evidence this was the case (Fig. 3.4, 3.5). Similarly, we found no evidence that the effects of O_{R1} mutations on RNA polymerase binding were biased, as expression levels measured in the absence of any repressors did not correlate with repression levels (Fig. 3.6, 3.7). As there are two operators in our system, enhanced binding could also be due to cooperativity between P22 C2 and Lambda CI, but the two repressors were shown not to be able to interact with each other measurably[134].

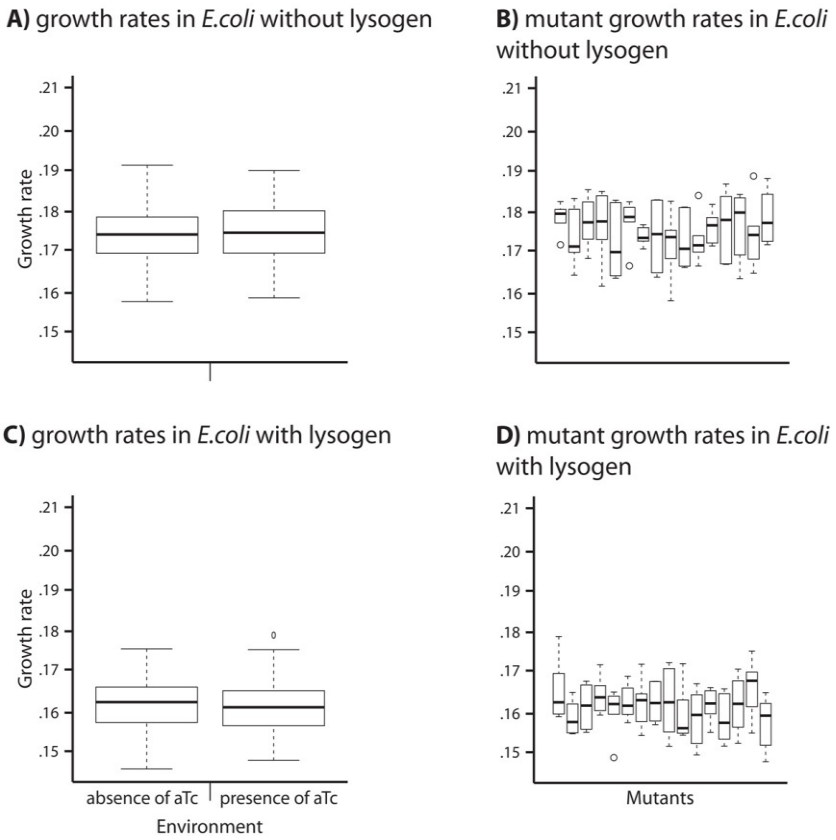

Figure 3.4. Effects of O_{R1} mutations on growth rates in *E. coli*.

We measured growth rates of three replicate populations in the absence and in the presence of the inducer aTc. ***A)*** *Box plots comparing growth rates in the absence and in the presence of the inducer across all O_{R1} mutants in the lysogen-free strains.* ***B)*** *Box plots comparing individual differences in growth rates between mutants in the lysogen-free strains.* ***C)*** *and* ***D)*** *are same as above, but for lysogen E. coli strains.*

Crosstalk-enhanced repression can arise if the concentration of the cognate repressor at a local promoter is higher when both repressors are present in the system. One way this could occur is that the two repressors share and compete for binding at an appreciable portion of their non-cognate sites. As was recognized previously, many transcriptional regulators are bound non-specifically to random DNA sequences along the bacterial genome[41]. However, there will be many sequences (referred to as non-cognate) that are more closely related to a TF's target sequence and hence will be bound more strongly, trapping the TF through non-cognate binding. A higher portion of those non-cognate sequences will be shared between related TFs, as we found for P22 C2 and Lambda CI, which share 7-8% of their non-cognate sites. In such cases, binding competition for the shared non-cognate sites can either increase the dissociation rate of the bound repressor through a process called facilitated dissociation[14,136], or sterically hinder binding of the other repressor[12]. Regardless of the specific mechanism, we will call the competition for binding at non-specific DNA sites global crosstalk. Global crosstalk can lead to increased turnover and hence increased free concentration of the cognate repressor when the second repressor is present, if the second repressor (i) is present at higher concentrations; or (ii) binds more strongly to the shared non-cognate sites. As Lambda CI has stronger overall non-

cognate binding[135], its presence could lead to an increase in the free concentration of P22 C2 and, as P22 C2 binding is additionally aided by cooperativity, to greater binding at the cognate operators. Besides, modeling suggests that increasing concentrations of non-cognate proteins bound randomly to the DNA increase binding at the cognate promoter due to confinement of the cognate TF in the vicinity of its target site[45]. However, it is questionable if the low levels of repressor from a low copy plasmid (and even lower levels from the lysogen) could result in the necessary amount of random non-cognate binding to produce a significant effect[45]. It is rather more likely in this case that the effect comes from a shared pool of non-cognate binding sites.

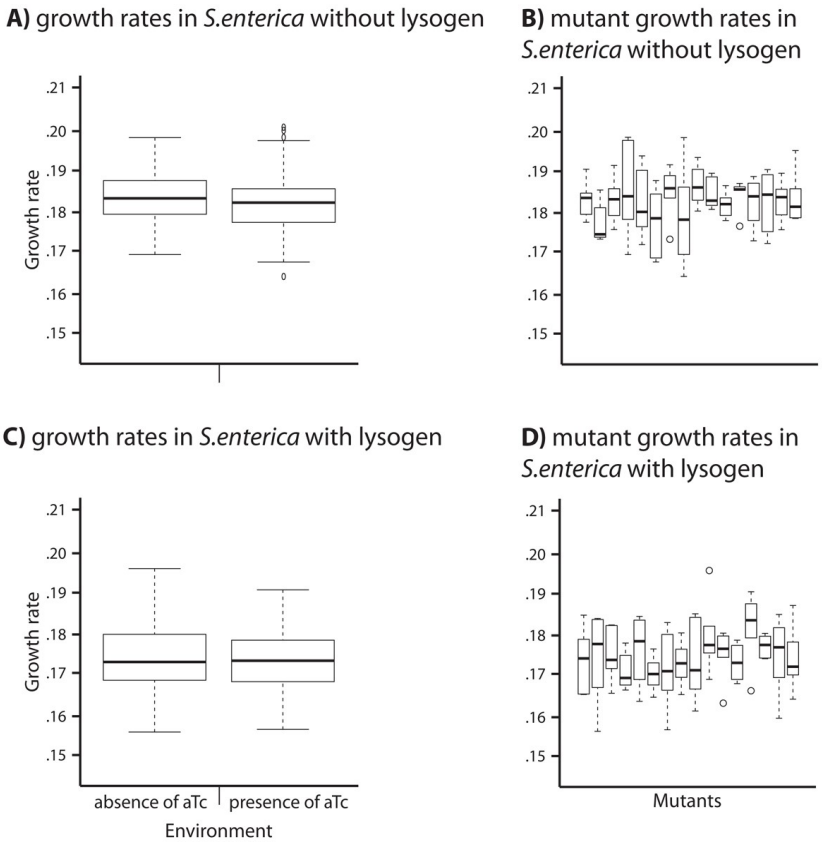

Figure 3.5. Effects of O_{R1} mutations on growth rates in *S. enterica*.

We measured growth rates of three replicate populations in the absence and in the presence of the inducer aTc. ***A)*** *Box plots comparing growth rates in the absence and in the presence of the inducer across all O_{R1} mutants in the lysogen-free strains.* ***B)*** *Box plots comparing individual differences in growth rates between mutants in the lysogen-free strains.* ***C)*** *and* ***D)*** *are same as above, but for lysogen S. enterica strains.*

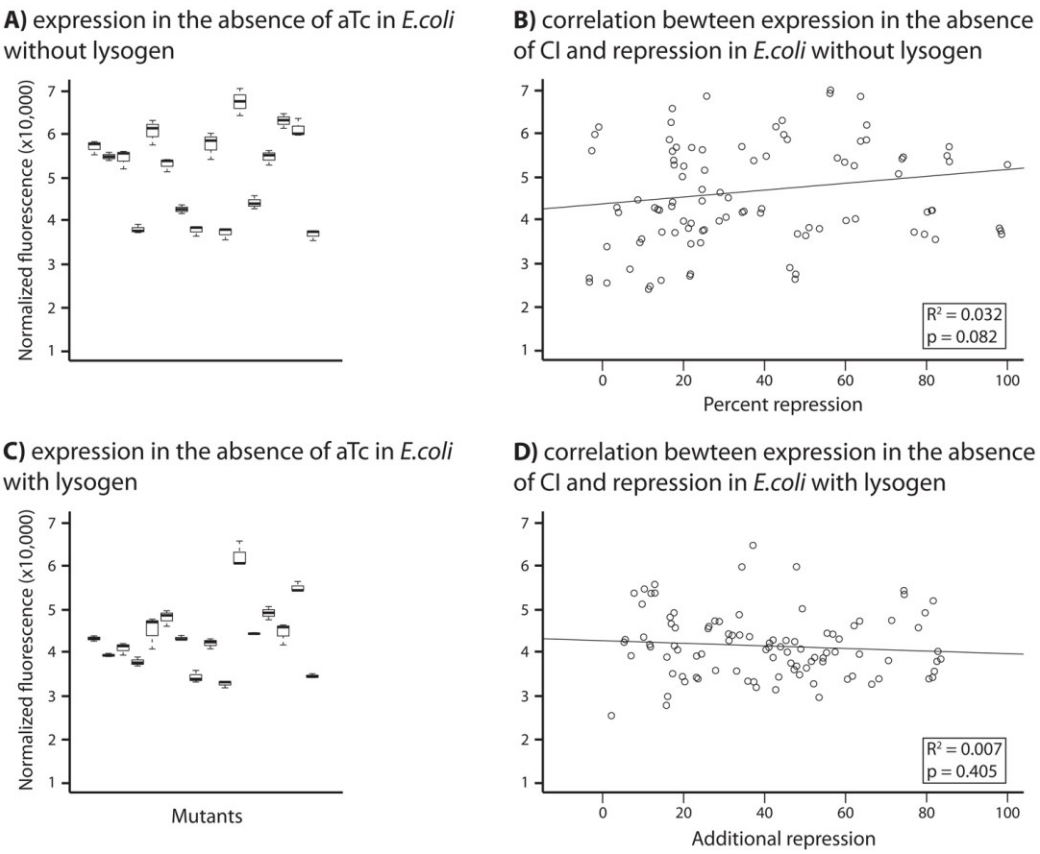

Figure 3.6. Effect of O_{R1} mutations on RNA polymerase binding in *E. coli*.

A) *Expression levels in lysogen-free E. coli strains in the absence of the inducer aTc for three replicates of each mutant. Expression levels under these repressor-free conditions indicate how O_{R1} mutations impact RNA polymerase binding.* ***B)*** *As O_{R1} mutant identity had a significant effect on RNA polymerase binding ($F_{31,62} = 192.9$; $P < 0.0001$), we checked if these differences impacted measured repression levels, and found no significant correlation.* ***C)*** *Expression levels in lysogen E. coli strains in the absence of the inducer aTc for three replicates of each mutant.* ***B)*** *As O_{R1} mutant identity had a significant effect on inducer-free expression levels ($F_{31,62} = 35.96$; $P < 0.0001$), we checked if these differences impacted measured repression levels, and found no significant correlation.*

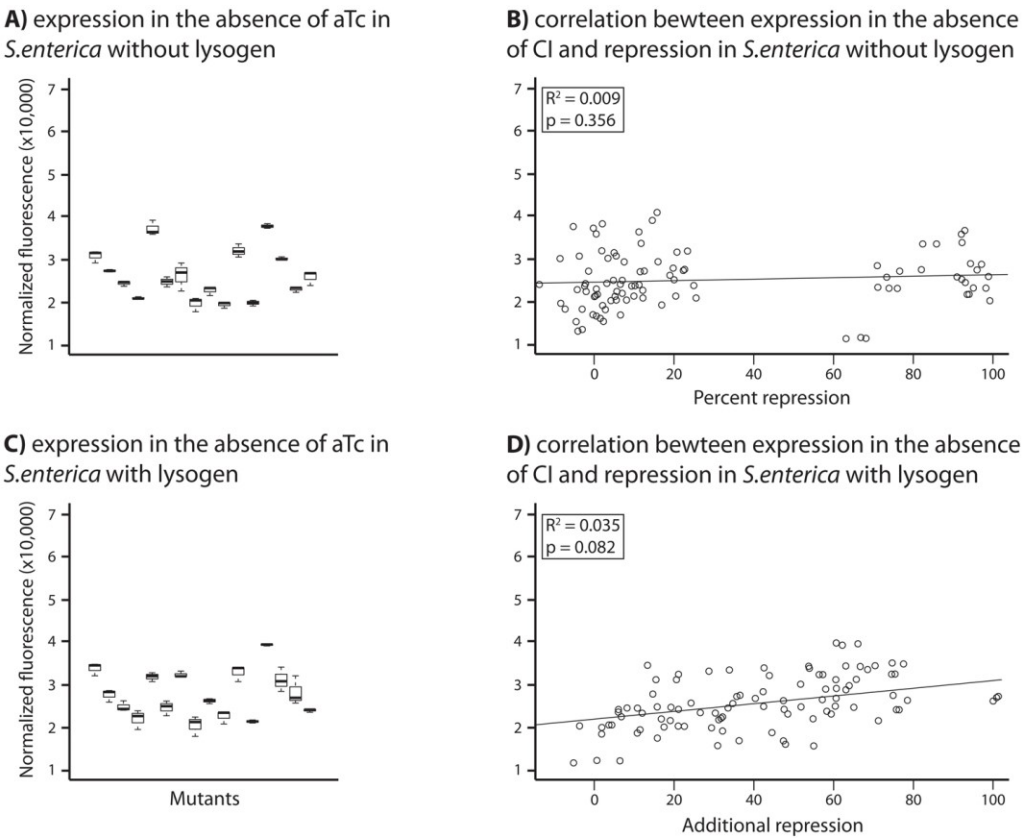

Figure 3.7. Effect of O_{R1} mutations on RNA polymerase binding in *S. enterica*.

A) *Expression levels in lysogen-free S. enterica strains in the absence of the inducer aTc for three replicates of each mutant. Expression levels under these repressor-free conditions indicate how O_{R1} mutations impact RNA polymerase binding.* ***B)*** *As O_{R1} mutant identity had a significant effect on RNA polymerase binding ($F_{31,62} = 45.57$; $P < 0.0001$), we checked if these differences impacted measured repression levels, and found no significant correlation.* ***C)*** *Expression levels in lysogen S. enterica strains in the absence of the inducer aTc for three replicates of each mutant.* ***B)*** *As O_{R1} mutant identity had a significant effect on inducer-free expression levels ($F_{31,62} = 30.00$; $P < 0.0001$), we checked if these differences impacted measured repression levels, and found no significant correlation.*

To test if global crosstalk at non-specific binding sites might play a role in the observed enhanced repression, we adapted the model we previously used to study the binding landscape of Lambda CI and P22 C2[135]. This model, which is based on the thermodynamic properties of protein-DNA binding[51,57], relates the binding energy between a repressor and an operator to the repression level of the promoter, by assuming that the duration of the TF x retention on the operator (determined by the equilibrium dissociation constant, $K_{seq,x}$) is the rate-determining step in transcriptional initiation[12]. To keep the model and its interpretations simple, we considered the scenario where one of the repressors (Lambda CI) binds the local operator weakly with a constant binding energy, and we calculated repression levels across a range of P22 C2 binding energies. Then, we compared repression levels across this range when: (i) non-cognate sites are not shared between the two repressors, meaning that $K_{seq,x}$ of each repressor is independent of the other repressor; (ii) non-cognate sites are shared

between the two repressors, leading to a dependency of P22 C2 $K_{seq,P22}$ on Lambda CI concentration and non-cognate binding energy; and (iii) only P22 C2 is present, but at double the concentration, to compensate for the absence of Lambda CI in terms of total TF concentration.

The model showed that global crosstalk can play a critical role in repression at the target operator (Fig. 3.8). Presence of global competition at shared non-cognate sites between two repressors can lead to increased repression at the local promoter – by changing operator affinity in dependence of the non-cognate repressor concentration and non-cognate binding energy – going beyond what can be achieved by doubling the concentration of a single repressor (Fig. 3.8A). In fact, a comparison with the model that did not account for non-specific binding dependency showed that at least a 10-fold excess of Lambda CI would be required to produce the same increase in repression (Fig. 3.9), suggesting that the experimental results are not simply a consequence of a difference in concentrations between the two repressors. When repressor concentrations are fixed, local repression was increased when the non-cognate repressor binds the globally shared non-cognate sites more strongly than the other (Fig. 3.8B) – as was suggested to be the case for Lambda CI and P22 C2[135].

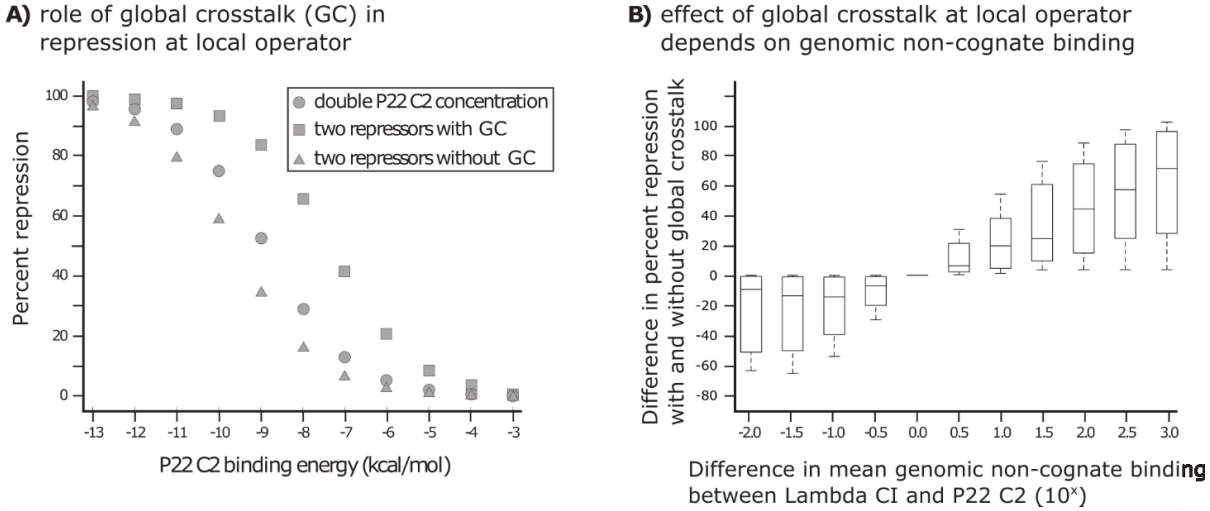

Figure 3.8. Global crosstalk impacts repression at a local operator.

A) *Using a thermodynamic model of gene regulation, we simulated binding at the O_{R1} site when: (i) only P22 C2 is present at double the concentration (circles); (ii) two repressors are present and global crosstalk takes place across all their shared non-cognate sites (squares); and (iii) two repressors are present but no global crosstalk occurs (triangles). Lambda CI binding energy is kept constant at -9 kcal/mol, which is a low binding energy in the range of mean genomic non-cognate binding.* ***B)*** *For global crosstalk to increase repression at a local operator, mean genomic non-cognate binding of the repressor with weaker binding at the local operator (Lambda CI in the case of our simulations) must be stronger than that of P22 C2. Mean genomic non-cognate binding represents the average binding energy of a repressor across all of its non-cognate sites in the genome. For each specific 'difference in mean genomic non-*

cognate binding between Lambda CI and P22 C2', the box plot is based on a range of P22 C2 cognate binding energies from -13 to -3, as shown in ***A)****.*

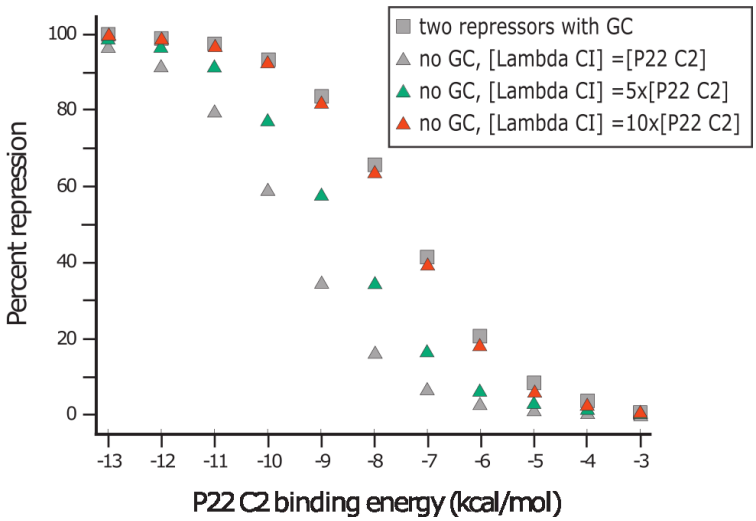

Figure 3.9. Dependency of cognate repression on Lambda CI concentration.

We used the thermodynamic model to calculate repression across all O_{R1} mutants for a range of P22 C2 binding energies (as seen in Fig. 3.4A). We compare repression achieved in the presence of two repressors when the model allows for global crosstalk (squares) and when it does not (triangles). When the model does not allow for global crosstalk, we modified Lambda CI concentration by increasing it 5- (green) and 10-fold (red) compared to P22 C2 concentration, and found that the effect on repression achieved by global crosstalk requires at least a 10-fold increase in Lambda CI concentration. Note that a 10-fold increase in Lambda CI concentration might not be biologically realistic.

3.4 Discussion

In this study, we experimentally investigated the effects of crosstalk on gene expression regulation and found that the presence of two related repressors leads to elevated repression at a local promoter. This means that having two different TFs in the system leads to stronger overall binding at the cognate operator compared to having only the cognate TF at higher concentrations (equal to the concentration of having both TFs). Previous modeling suggests that this could be due to confinement of the cognate TF near its target site if a large part of the DNA is bound by non-cognate TFs[45], which is however unlikely in our system. By contrast, using a thermodynamic model of gene regulation, we find that global crosstalk can give rise to elevated binding at a local operator at reasonable repressor levels: binding competition between two TFs at their shared non-cognate operators can either enhance the removal of the bound TF from, or hinder its binding to these sites, leading to an increase in the free concentration of TF in the cell and hence an increase in binding to stronger affinity (i.e. cognate) operators. This implies that global crosstalk can increase

the apparent specificity of TF binding. Apparent specificity here describes an 'effective' property, which arises from context effects (e.g. binding of other TFs), as opposed to 'bare' specificity, which is giving through binding affinities measured in isolation (e.g. *in vitro*). The action of the non-cognate repressor can be compared to the role of an 'inducer', which is similar to sigma factors 'inducing' RNAP off non-specific sites as has been described previously[52].

Global crosstalk will increase binding at a given operator when several conditions are met. First, the TFs must have the potential for non-cognate binding, which is true for all known TFs due to their tolerance for operator mutations[133]. Second, the two TFs must share non-cognate binding sites, meaning that they participate in global crosstalk with each other. This condition is often met, especially between related TFs that have a substantial overlap in the binding sites they can bind[137,138]. For such TFs, shared non-cognate binding sites at a genomic scale are common due to the short length of operators. Third, the non-cognate TF must have stronger non-specific binding, meaning greater overall propensity for binding DNA, as is the case for Lambda CI compared to P22 C2[98,119]. Finally, global crosstalk will increase binding only at higher affinity operators, which are likely dissimilar even between related TFs, a difference which is enhanced with cooperativity in cognate binding, whereas shared weak operators will be predominantly bound by the more promiscuous, non-cognate TF. As such, introduction of a global regulator, which is by definition promiscuous, ought to increase specificity of more locally-acting TFs. Put together, global crosstalk leads to higher binding at the few more specific operators, and to a reduction in binding to the many shared weak sites. It is also worth noting that there might be other indirect means of how non-cognate TF binding can affect cognate binding. For example, Lambda CI could increase P22 C2 operator occupancy by introducing large-scale modifications to the chromosomal structure – a conceivable scenario given the propensity of Lambda CI to induce DNA looping[139].

Ordinarily, discussions of crosstalk in transcriptional regulation focus on how non-cognate binding of one TF interferes with cognate binding of other TFs in the cell[37,115]. We show that non-cognate interactions can also dramatically increase specificity at a local cognate operator. While this effect can increase robustness in the system, it could also affect organismal function negatively by interfering with the transcriptional program, depending on the targeted promoter output. Although random molecular crowding can also affect binding specificity at sufficiently high concentrations of non-cognate proteins[45], crosstalk-enhanced specificity is especially likely when homologous TFs are present in the cell, as their shared ancestry often results in similar DNA-binding specificities[140,141] even at low TF levels. Therefore, TFs that are horizontally transferred between related species, TFs that diverged following a duplication event, and decision-making bacteriophage TFs during co-infection are particularly likely to participate in crosstalk. Indeed, divergence in TF binding preference and operator sequence often follows gene duplication[93,142,143] and horizontal gene transfer events[79,144], indicating overall detrimental fitness effects of crosstalk. Similarly, many bacteriophages develop 'immunity' to prevent related bacteriophages from successfully infecting the host cell and interfering with their

lifestyle decision making[114,145]. Our findings provide an insight into how selection acts on homologous TFs and their binding sites, by identifying a novel mechanism through which crosstalk can affect organismal fitness.

3.5 Tables

Table 3.1. Identity of O_{R1} mutants used in experiments.

Only the positions in the O_{R1} site that were mutated (as seen in Fig. 3.1A) are shown. The identities of introduced mutations are indicated by an 'X'. Mutants are arranged in order of decreasing percent repression by P22 C2 in the lysogen-free *E. coli* strain (corresponding to blue mutants in Fig. 3.2B).

	O_{R1}					
P22 P_R mutated positions	T	T	A	A	T	C
Lambda P_R positions	A	C	T	G	C	G
mutant 1	-	X	-	-	-	-
mutant 2	-	-	-	-	-	X
mutant 3	-	-	-	X	-	-
mutant 4	-	-	-	-	X	-
mutant 5	-	-	X	-	-	-
mutant 6	-	X	X	-	-	-
mutant 7	-	-	-	X	X	-
mutant 8	-	X	X	X	-	-
mutant 9	-	X	X	-	-	X
mutant 10	-	-	X	X	-	X
mutant 11	X	X	X	X	X	X
mutant 12	-	-	-	X	-	X
mutant 13	-	X	X	X	-	X
mutant 14	-	X	X	X	X	X
mutant 15	X	X	X	X	X	-
mutant 16	-	X	X	X	X	-

Table 3.2. T-tests comparing percent repression in lysogen-free strains.

FDR-corrected two-tailed t-tests were carried out to test if, for each mutant, total percent repression achieved by Lambda CI and P22 C2 was significantly different. The tests were performed independently for each host, *E. coli* and *S. enterica.* P-values were evaluated with 4 degrees of freedom. Significance is shown in Fig. 3.2. Mutants are ordered according to Table 3.1.

Mutant	t-test in *E. coli*	P value in *E. coli*	t-test in *S. enterica*	P value in *S. enterica*
1	45,5	1,4 E-06	40,3	2,3 E-06
2	13,5	1,7 E-04	51,1	8,8 E-07
3	90,5	9,0 E-08	33,3	4,8 E-06
4	27,3	1,1 E-05	111,8	3,8 E-08
5	0,5	0,6	88,1	10,0 E-08
6	7,3	1,9 E-03	68,1	2,8 E-07
7	9,7	6,3 E-04	35,5	3,8 E-06
8	2,9	0,05	21,3	2,9 E-05
9	2,4	0,08	1,9	0,1
10	4,0	0,02	13,2	1,9 E-04
11	11,1	3,8 E-04	0,9	0,4
12	33,5	4,7 E-06	2,4	0,1
13	1,7	0,2	1,0	0,4
14	10,5	4,7 E-04	0,2	0,8
15	44,5	1,5 E-06	1,1	0,3
16	1,4	0,2	0,6	0,6

Table 3.3. T-tests comparing percent repression in lysogen strains.

FDR-corrected two-tailed t-tests were carried out to test if, for each mutant, additional repression achieved through plasmid induction of Lambda CI or P22 C2 was significantly different. The tests were performed independently for each host, *E. coli* and *S. enterica.* P-values were evaluated with 4 degrees of freedom. Significance is shown in Fig. 3.3. Mutants are ordered according to Table 3.1.

Mutant	t-test in *E. coli*	P value in *E. coli*	t-test in *S. enterica*	P value in *S. enterica*
1	56,6	5,8 E-07	41,3	2,1 E-06
2	28,2	9,4 E-06	17,6	6,1 E-05
3	67,5	2,9 E-07	1,9	0,1
4	28,5	9,0 E-06	4,5	0,01
5	7,1	2,1 E-03	25,6	1,4 E-05
6	34,1	4,4 E-06	6,6	2,0 E-03
7	3,8	0,02	6,7	2,5 E-03
8	5,6	4,9 E-03	4,4	0,01

9	0,7	0,5	15,2	1,1 E-04
10	7,1	2,0 E-03	3,4	0,03
11	3,7	0,02	0,1	0,9
12	6,8	2,5 E-03	3,8	0,02
13	2,7	0,05	2,4	0,07
14	16,1	8,6 E-05	31,8	5,8 E-06
15	2,1	0,1	11,5	3,3 E-04
16	15,9	9,2 E-05	3,0	0,04

Table 3.4. Parameter values used in the thermodynamic model.

All parameters were selected from the literature, except repressor concentrations, which were taken from previous simulations[135].

Parameter	Parameter value
P22 P_R promoter strength	40[118]
P22 C2 dimer cooperativity	3 kcal/mol[12]
Lambda CI concentration	1 µM
P22 C2 concentration	1 µM[135]
Non-specific binding P22 C2 (when independent)	-7.4 kcal/mol[12]
RNAP concentration	3 µM[118]
Binding affinity of Lambda CI	-9 kcal/mol
Binding affinity of P22 C2 to wt O_{R1}	-12 kcal/mol[98]
Binding affinity of P22 C2 to wt O_{R2}	-10 kcal/mol[98]
Binding affinity of RNAP binding to P_R	-12,5 kcal/mol[63]

3.6 Methods

Strains and plasmids

The experimental system is based on the 'genetic switch' of the bacteriophage P22, and more specifically on the P_R promoter system. We constructed two template plasmids consisting of two parts that are separated by 500 random base pairs and a terminator sequence: an inducible repressor gene (either Lambda *cI* or P22 *c2*) under an inducible P_{TET} promoter followed by TL17 terminator sequences on one strand; and the P22 P_R promoter (containing the RNA polymerase binding site as well as two operators, O_{R1} and O_{R2}) controlling the expression of a *venus-yfp*[120] fluorescence marker on the other strand (Fig. 3.1A). Binding to O_{R1} leads to repression. Binding to O_{R2} assists in repression mainly through cooperative binding between two repressor

dimers[121], which in our experiments is possible only between P22 C2 dimers, as Lambda CI does not bind the P22 O_{R2} site and there are no cooperative interactions between Lambda CI and P22 C2 dimers[134]. Downstream of the promoter is a ribosomal binding site in front of the reporter gene. These parts were cloned into a low copy number plasmid (pZS*) containing a kanamycin resistance marker[99], resulting in two template plasmids, one with Lambda *cI* and the other with P22 *c2* (Fig. 3.1A). The plasmid libraries were then transformed into: MG1655 derived *E. coli* cells (strain BW27785, CGSC#: 7881) [44] (i) without; and (ii) with Lambda lysogen; *S. enterica* strains (iii) without (strain LT2); and (iv) with P22 lysogen (strain TH2680) (Fig.1B,C). The P22 lysogen strain was obtained from Marc Erhardt and introduced as described in Benson & Goldman (1992)[146]. The Lambda lysogen was obtained from Dominik Refardt and introduced into BW27784 by mixing serially diluted Lambda chlor lysate with 0.1 ml of an overnight culture (LB) in 3 ml of phage soft agar and spread on phage plates. Lysogens were used to bring a second repressor into the cell because expression of both repressors from plasmids resulted in toxicity and strongly reduced growth.

Construction of crosstalk libraries

Informed by our previous work on the effects of mutations in P22 O_{R1} on the binding of Lambda CI and P22 C2[135], we selected 16 P22 O_{R1} mutants that have a minimal effect on RNA polymerase binding, while exhibiting a gradient in both Lambda CI and P22 C2 binding (Fig. 3.2). These mutants contained between 1 and 6 point mutations in O_{R1} (Table 3.1). Mutants were constructed by synthesizing oligonucleotides of 73bp length (Sigma Aldrich) carrying wild type P22 O_{R2} and mutated O_{R1}, and cloning them into the two experimental template plasmids. Clones carrying correct mutants were confirmed through Sanger sequencing.

Fluorescence assays and growth rate measurements

We measured fluorescence of all mutants (in all four experimental systems shown in Fig. 3.1), both in the presence and in the absence of the inducer aTc. Three biological replicates of each mutant of the library were grown at 37°C overnight in M9 media, supplemented with 0.1% casamino acids, 0.2% glucose, 30μg/ml kanamycin, and either without or with 8ng/ml of the P_{TET} inducer aTc. Overnight cultures were diluted 1,000X, grown to OD_{600} of approximately 0.1, and their fluorescence measured in a Bio-Tek Synergy H1 plate reader. We also continued growing these cultures and measured their OD_{600} every 20 minutes, in order to obtain growth curves for all mutants in all experimental systems. All replicate measurements were randomized across multiple 96-well plates.

All measured mutants (including lysogens) had fluorescence levels significantly above the detection limit of the plate reader, resulting in measurements at least 1.5 fold greater than the non-fluorescent control. Fluorescence values were normalized by OD_{600} values (in RFU=Relative Fluorescence Units) and averaged over three replicates. Repression values were calculated as a normalized ratio between the measured fluorescence with and without the repressor:

$$Percent/additional\ repression = \left(1 - \frac{RFU_{presence\ of\ inducer}}{RFU_{absence\ of\ inducer}}\right) * 100.$$

Total percent repression was calculated for the strains without the lysogen, as no repressors were present in the absence of the inducer. In contrast, in lysogenic strains a repressor was always present (Lambda CI in the *E. coli* lysogen, and P22 C2 in the *S. enterica* lysogen), so adding the inducer added either the second repressor or more of the same repressor, leading to additional repression measures. Standard errors of the mean percent/additional repression were calculated using error propagation in order to account for the inherent variability in the fluorescence measurements.

We obtained growth rates for three replicates of each mutant in every experimental system, both in the absence and in the presence of the inducer aTc, by calculating the highest slope of six consecutive $\log_{10}(OD_{600})$ measurements using a sliding windows approach.

Statistical analyses

In lysogen-free strains, we tested if there was an overall difference in percent repression between mutants carrying Lambda CI and those carrying P22 C2 using ANOVA (aov function in R statistical software version 3.5.0), with repressor identity as a fixed variable, percent repression as the response, and replicates nested within mutant identity as the error structure. Then, we performed a series of FDR-corrected two-tailed t-tests asking if there was a significant difference between having Lambda CI or P22 C2 for each mutant individually. We performed both the ANOVA and t-tests for mutants measured in *E. coli* and *S. enterica* independently. Then we performed the same tests for lysogen strains using additional repression as the response variable, asking if there was an effect of having one or two repressors across all mutants (ANOVA) and for each mutant individually (FDR-corrected t-tests). Note that we did not perform direct statistical comparisons between lysogen-free and lysogen strains, because their output measurements (percent repression vs. additional repression) constitute fundamentally different variables.

To analyze if growth rates were constant across the strains and between environments (without and with the inducer, aTc), we performed two sets of tests. First, we used ANOVA (absence or presence of aTc as the fixed variable; percent/additional repression as the response variable; and replicate nested within mutant identity as the error structure) to test if the growth rates depended on the presence of the inducer, and we tested this across all mutants in a given host. Then, we tested if growth rates differed between mutants in a given host, using ANOVA with mutant identity as the fixed variable and replicate as the error.

We were also interested in whether expression levels in the absence of the inducer were consistent between mutants. In lysogen-free strains, differences in inducer-free expression levels would arise if the mutations introduced into O_{R1} had an effect on RNA polymerase binding. In lysogen strains, these differences would arise if mutations

affect either repressor or RNA polymerase binding. To test for these effects, we used ANOVA with expression levels in the absence of aTc as the response variable, mutant identity as the fixed variable, and replicate as error. We performed these tests independently for all four strains (*E. coli* without and with lysogen, and *S. enterica* without and with lysogen), and found that there were significant differences in RNA polymerase binding between mutants (Fig. 3.6,3.7). These differences could impact our findings if the expression levels in the absence of the inducer affected repression levels when the inducer was present. To test if this was the case, using a linear regression we tested for a correlation between expression levels in the absence of the inducer and percent/additional repression in the presence of aTc and found no correlation (Fig. 3.6,3.7). These results indicate that, while mutants differ with respect to their inducer-free expression levels, these differences did not significantly influence percent/additional repression, and hence did not play a major role in our findings.

Thermodynamic model of gene regulation

The model is based on previously described thermodynamic approaches[51,57], which rely on several assumptions: (i) TF binding to DNA takes place at thermodynamic equilibrium; (ii) gene expression can be equated with the probability of binding of participating proteins (in our case RNAP and repressor); and (iii), the contribution of each base pair in the operator to binding is additive. The probability of a gene being expressed is derived by summing over all states where RNAP is bound relative to all possible binding states. The dissociation constant of any transcription factor x from an operator sequence is calculated in the following way[12]:

$$K_{seq,x} = ns_x \cdot e^{-E_{seq,x}}$$

, where ns_x is the non-specific binding constant of repressor x, $K_{seq,x}$ the dissociation constant, and $E_{seq,x}$ the binding energy of repressor x for a sequence *seq* (parameters are given in Table 3.4). Note that we account for concentration-specific effects separately and ns_x incorporates only non-specific background binding and other non-specific cellular effects. Repression was calculated either in the presence of both repressors, or only with P22 C2 at double the concentration in the following way:

$$Gene\ expression = \frac{1}{1 + \frac{K_p}{[RNAP]} * \frac{\left(1 + \frac{[R_{Lambda}]}{K_{Lambda}} + 2\frac{[R_{P22}]}{K_{P22}} + \left(\frac{[R_{P22}]}{K_{P22}}\right)^2 e^{\omega}\right)}{\left(1 + \frac{[R_{P22}]}{K_{P22}}\right)}}$$

, where K_x represents the effective equilibrium dissociation constant (relative to the genomic background) as calculated above for a specific sequence – which is the concentration for half-maximal occupation of the site - of, either RNAP (K_P) or one of the repressors (K_{Lambda}, K_{P22}). [R] is the concentration of repressor dimers, which is the effective concentration, as repressors only bind as dimers and, as we assume fast dimerization[123], this corresponds to half of the total monomer concentration in the cell. [RNAP] is the concentration of RNAP, and ω is the cooperativity energy value, describing the strength of interaction between two P22 C2 repressor dimers. All concentrations, non-specific binding constants and dissociation constants are given in units of µM (Table 3.4). The calculated gene expression value is a relative measure,

with 1 indicating full expression and 0 no expression. Percent repression was then calculated using the formula:

$$Percent\ repression = \left(1 - \frac{gene\ expression_{repressor}}{gene\ expression_{no\ repressor}}\right) * 100.$$

We compared repression with both repressors when non-cognate binding was independent of the second repressor, and when the non-cognate binding constant of P22 C2 $ns_{P22\ C2}$ in the calculation of K_x was replaced with the factor:

$$ns_{P22\ C2} \cdot \frac{ns_{P22\ C2} \cdot c_{P22\ C2}}{ns_{Lambda\ CI} \cdot c_{Lambda\ CI}}$$

The factor results from the hypothesis that non-specific binding of P22 C2 decreases proportionally with Lambda CI binding at shared global non-specific sites. Hence, P22 C2 non-specific binding is inversely proportional to Lambda CI concentration and its non-specific binding constant - which both increase global crosstalk - but the modification is relative to the amount of non-specific binding of P22 C2 itself, hence relative to P22 C2 concentration and its non-specific binding constant. Here we do not differentiate between shared and separate non-cognate binding sites but introduced a constant that modulates the overall binding as an average effect of binding competition. At first, P22 C2 binding energy was varied from -13 to -3 kcal/mol, Lambda CI was kept at -9 kcal/mol and concentrations were the same ($1*10^{-6}$M). Non-cognate binding was 10^7M for P22 C2 and 10^8M for Lambda CI. For bar plots over a range of relative non-cognate values between P22 C2 and Lambda CI, non-cognate binding was 10^7 for P22 C2 and for Lambda CI varied from 10^5M to 10^{10}M; and we subtracted repression values of both repressors with and without modified non-specific P22 C2 binding. For the comparison with concentration effects of Lambda CI, P22 C2 concentration was kept at $1*10^{-6}$M for Lambda CI was varied between $1*10^{-6}$M and $10*10^{-6}$M. We could not use the model to directly verify these findings for our system, as already the data fit for Lambda CI repression in the absence of a second repressor was very poor, likely because energy matrices are not valid when moving too far away from the wild type operator sequence[101].

Note that we model non-cognate binding competition specifically as a change in the binding constant of the repressor instead of assuming that competition in itself would lead to variation in occupancy of binding sites. As the latter possibility would effectively result in a change in effective repressor concentration, those two scenarios lead to very similar results in an equilibrium system.

Calculating the shared non-cognate sites of Lambda CI and P22 C2

We tested for shared non-cognate binding sites between Lambda CI and P22 C2 across the Salmonella genome, as well as across 1,000,000 random DNA sequences. For the Salmonella genome we used a sliding window approach to calculate the energy penalty at every genome position for Lambda CI or P22 C2. From energy penalties calculated for genome positions or to the random sequences, we obtained the binding affinities of either repressor through:

$$K_{seq,x} = ns_x \cdot e^{-E_{seq,x}}$$

$$q_{seq,x} = \frac{c_x}{K_{seq,x}}.$$

Then we took the sequences that were above the non-cognate binding threshold (10^{-7}M) and calculated the % of P22 C2 non-cognate sites that were shared with Lambda CI non-cognate sites. For random sequences we found 7,6% shared, whereas for the Salmonella genome it was 7,1% shared relative to the total number of P22 C2 'more specific' non-cognate sites.

3.7 Author contributions

C.I. (Claudia Igler), M.L. (Mato Lagator), G.T. (Gašper Tkačik), C.C.G. (Călin C. Guet) conceived the study together. C.I. and M.L. designed and carried out the experiments and analyzed the data. C.I. wrote the code and ran the model. C.I. and M.L. wrote the current draft of the manuscript.

4. The evolution of phage immunity regions

4.1 Crosstalk between phage repressors

Recent years have shown that phages are omnipresent entities, which manipulate the behavior of their bacterial hosts, thereby profoundly affecting most biological organisms as well as the biosphere. Their study has yielded profound insights into molecular mechanisms of gene regulation, horizontal gene transfer and drivers of host co-evolution[225]. An area of research that has been neglected, especially with regard to the underlying molecular bases, are interactions between phages themselves. Phages, however, are likely to encounter other phages during their search for a new host. This issue is especially relevant for temperate phages as they can choose to turn their host cells into lysogens, where the phage genome (prophage) is either integrated into the host genome or maintained extrachromosomally, replicating together with the host cell without producing phage progeny[226]. During their stable propagation over extended periods of time, prophages are likely to encounter heterologous phages infecting their host cell, as evidenced by the large number of polylysogens found in nature[227,228]. In rare cases the integration of several phage genomes into the same chromosome after superinfection was even witnessed for phage λ, even though the paradigm deems it immune to such superinfection[229]. The induction of polylysogens can lead to within-host competition, which is generally detrimental for the productivity of at least one of the phages involved[230]. Although the molecular mechanisms remained unclear, within-host competition showed a significant impact on phage fitness[230].

Another indicator for the importance of phage superinfection is the amount of different mechanisms temperate phages encode to prevent other phages from successfully infecting an already lysogenized host[231]. The general mechanism that confers immunity against other phage infections by the prophage residing in the cell is conveyed by the phage repressor responsible for maintaining the lysogenic cycle. This lysogenic repressor can interfere with a newly infecting phage by binding to the regulatory regions responsible for deciding the phage lifestyle (lytic or lysogenic), thereby aborting the incoming infection[232]. Phages interacting in this manner - meaning that their repressors and regulatory binding sites are compatible – are called homoimmune and belong to the same immunity class. The evolution of these superinfection immunity classes can be driven by invasion of ultra-virulent mutants – phages whose binding sites are immune to repressor binding of the residing prophage – and co-evolution of repressors to those binding sites[232]. As the immunity modules in different phage species likely evolved through extensive horizontal gene transfer[145], it is unclear how many members one immunity group contains, but a classification of 100 wild bacteriophages revealed 20 different immunity classes with one consisting of 48 members[233]. As evidenced by the example of cross immunity between the phages λ and VT2-Sa, this classification is likely based on similarities in operator sequence recognition and repressor binding motifs, even though the sequences can diverge quite substantially from each other[114]. So far, immunity classes have been based on phenotypic classification: whether a second phage can infect a lysogenized host cell or not. However, the intermediate steps in the immunity diversification

process likely involve binding crosstalk between phages, meaning that the repressor of one phage can bind to the operator of another phage, even though the operator sequence has diverged from its own (preferred) target sequence. This kind of crosstalk – which can also be asymmetric between the phages - and its impact on phage fitness remain unknown.

In order to study the impact of different strengths of crosstalk between phages from different immunity classes, we used phage P22 and introduced mutations into the P22 O_{R1} operator, which would make it more similar to the λ O_{R1} operator[135]. The O_{R1} operator is responsible for repression of P_R, the promoter driving the expression of lytic genes. The repression strengths of the mutated O_{R1} operators by either the λ CI repressor or the P22 C2 repressor or both together are known from *in vivo* measurements of isolated P_R systems on plasmids (see Chapter 2&3). (The presence of both repressors in *S. enterica* cells for those measurements was achieved by introducing a P22 C2 lysogen and a plasmid expressing the λ CI repressor, which is very similar to the experiments with phages described below.) In this manner, phage fitness can be linked to underlying molecular mechanisms.

As a first step we introduced the mutations into P22 O_{R1} through recombination and subsequent selection according to plaque morphology on a lawn of *S. enterica* cells. Even single mutations resulted in an observable difference in plaque morphology (Fig. 4.1). This is in agreement with our repression measurements in isolated systems, where we found a significant reduction in P22 C2 binding already for single mutants (Fig. 2.2). Increasing the number of mutations up to six yielded a very similar plaque morphology as that of single mutants with the clearing in the center of the plaque becoming even more pronounced, indicating a decreased likelihood of lysogenization. Wild type plaques showed the typical, turbid phenotype produced by temperate phages, surrounded by a darker ring, which consists of a combination between older bacterial colonies that are being lysed by phages, and younger small lysogenic colonies[234]. For mutant phages, plaques displayed a clear center, surrounded by a ring of dense colonies and at the edge another, thinner ring of clear space (bull's eye morphology). The loss of turbidity in the center stems from a reduced likelihood to form lysogens as the P22 O_{R1} mutants reduce repression of the lytic genes by P22 C2. During later infection cycles the number of phages (and hence the multiplicity of infection) increases, restoring the ability to produce lysogens[234]. Additionally, during later infection cycles bacterial colonies will already approach stationary phase and changes in host cell physiology, which generally decrease the competence for successful lytic infection[235], ultimately result in the darker outer ring, consisting largely of lysogenic bacterial colonies which ran out of nutrients and remained almost invisibly small.

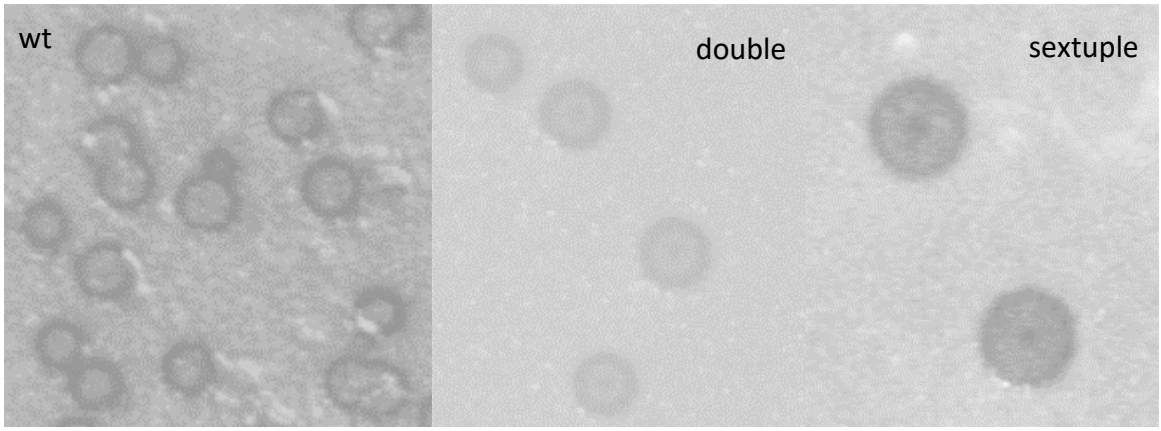

Figure 4.1. Plaque morphologies of wild type P22 phages (left) compared to a single O_{R1} mutant (middle) and the sextuple O_{R1} mutant (right).

Wild type phages produce turbid plaques, while mutant phages produce bull's eye formations.

As a next step we want to investigate the influence of crosstalk between the immunity region of P22 phages and λ CI. Therefore, plasmids containing an aTc-inducible λ CI repressor are introduced into host cells before infection with mutant P22 phages. In this manner, we avoid adding the complexity of infection by a second phage species, which can result in many other types of interference - from the competition for host resources to superinfection exclusion systems[230,236,237]. Assuming that λ CI only affects P22 fitness through binding at the mutated P22 O_{R1} operators, there are three possible outcomes:

i) Addition of λ CI does not influence the lifestyle decision of P22 phages at all, as the binding affinities (and/ or concentrations) of P22 C2 and Cro are higher, dominating the regulatory decision.

ii) Binding of λ CI increases repression of P22 O_{R1}, leading either to an increased frequency of lysogenic development, or – for strong enough binding – to abortion of the infection akin to superinfection immunity.

iii) Competition of λ CI for P22 O_{R1} binding interferes with P22 C2 binding and leads to a reduced frequency of lysogenic development. Additionally, this interference could also affect binding of Cro to O_{R1}, which is necessary for progression of the lytic cycle as it represses the lysogenic regulator CII.

Preliminary experiments with λ CI expressed in the host cells used for plating did not yield a visible difference in plaque morphology compared to the host cells without λ CI, for the triple mutant (Fig. 4.2), but suggested a decrease in the diameter of the clear center for the sextuple mutant (Fig. 4.3). However, plaque assays only allow the observation of quite substantial lifestyle changes, hence more subtle and direct measurements of burst size, latent period and lysogenization frequency would be required. First experiments with the sextuple mutant on host cells in the presence and absence of λ CI suggest that the latent period stays the same and that changes in burst size might not be significant (Fig. 4.4). The variation in burst size for the sextuple

mutant without λ CI is extremely high however and might be reduced in the presence of λ CI (Fig. 4.4).

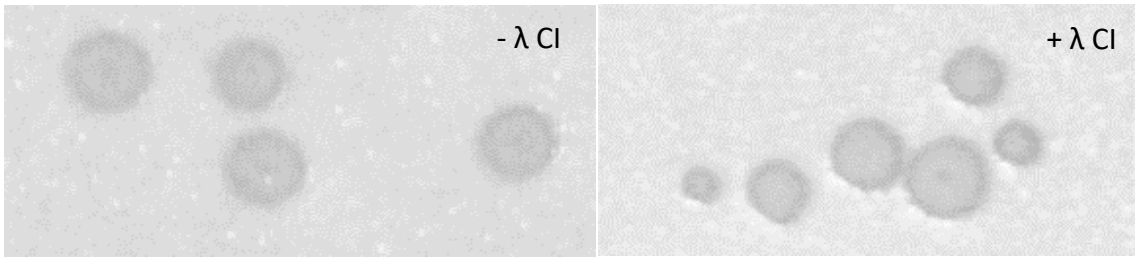

Figure 4.2. Comparison of plaques formed by a P22 triple O_{R1} mutant in the absence (left) or presence (right) of λ CI.

Bull's eye formation can be seen in plaques regardless of λ CI expression in the host cells.

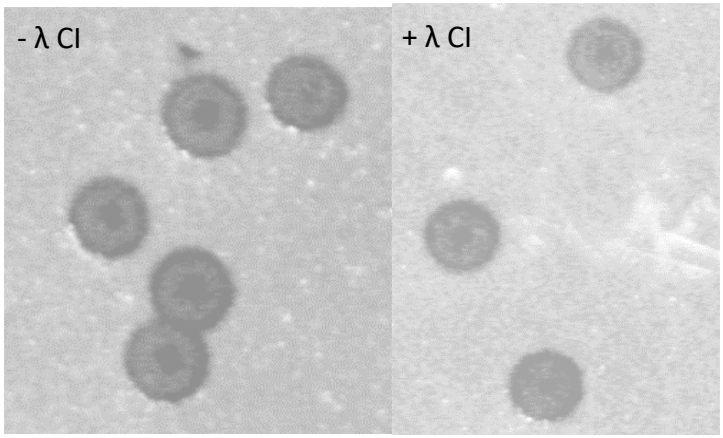

Figure 4.3. Comparison of plaques formed by the P22 sextuple O_{R1} mutant in the absence (left) or presence (right) of λ CI.

Bull's eye formation can be seen in plaques regardless of λ CI expression in the host cells, although the diameter of the clear zone seems to be decreased in the presence of λ CI.

After learning the impact of λ CI on the P22 phage lifestyle decision, we can speculate on the outcome when λ and P22 phages are competing for one host cell. Assuming that crosstalk between phage repressors affects phage fitness and assuming that it is asymmetrical (i.e. λ CI can bind to P22 O_{R1}, but P22 C2 cannot bind to λ O_{R1}), we can imagine several scenarios encountered by the two phages in one host cell:

1. Both phages are infecting the cell at the same time at MOI=1: Binding of λ CI to P22 O_{R1} will result in titration of λ CI, increasing the probability that phage λ will follow its lytic pathway. Depending on how λ CI binding interferes with the lifestyle decision of P22 (see i-iii), P22 might choose the lysogenic or the lytic lifestyle. The former will not lead to any P22 phage progeny as its lytic cycle will be repressed but the host cell is lysed by phage λ. The latter will lead to competition with phage λ for host resources

to produce virions, which in itself is disadvantageous, but the interference through λ CI likely provides another hindrance for P22 in within-host competition.

2. Both phages are infecting the cell at the same time at MOI>1: Lysogeny will be favored in both phages and both will integrate into the host genome as they use different attachment sites for integration.

3. P22 is present as a prophage and λ is infecting the cell at MOI=1: λ will most likely follow lytic development and lyse the cell. Except if λ CI strongly interferes with P22 C2 repression at P22 O_{R1}, P22 will remain lysogenic and will be destroyed together with the host cell. If P22 is induced, it will again compete with λ for the host resources.

4. P22 is present as a prophage and λ is infecting the cell at MOI>1: λ will likely integrate into the host chromosome, which does not interfere with the P22 prophage and the cell will become a polylysogen.

5. λ is present as a prophage and P22 is infecting the cell: This scenario corresponds to i-iii above, except if λ CI is titrated to significant amounts, reaching the threshold for induction. This is however unlikely, as λ CI negatively regulates its own expression, meaning that titration will only lead to higher λ CI production, which maintains the prophage state[201].

Overall, crosstalk will be neutral (if lysogeny is chosen) or beneficial for the crosstalking phage (in our case λ), but mostly detrimental for the phage that is interfered with (here P22) – which will likely lead to strong selection against crosstalk.

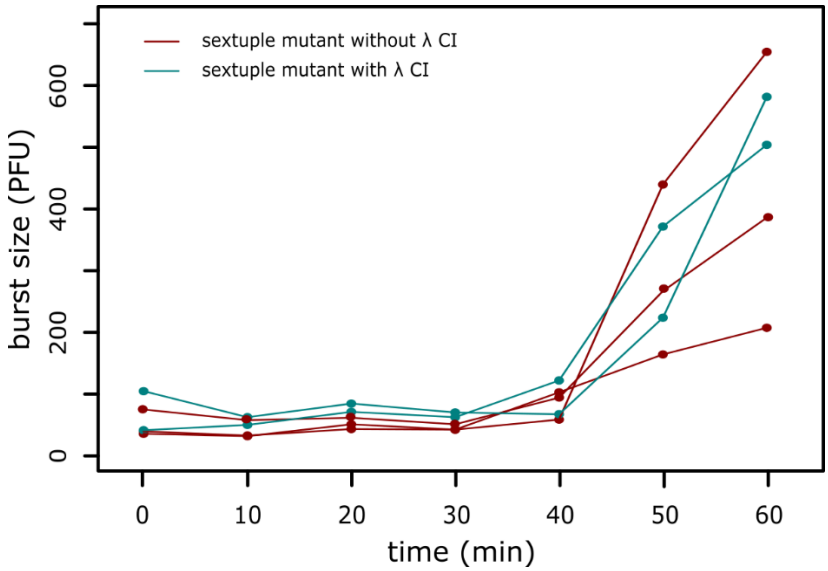

Figure 4.4. Burst size and latent period of the sextuple mutant in the presence (blue) or absence (red) of λ CI.

Time is given in minutes on the x-axis and the burst size in PFU on the y-axis. The latent period was roughly 40 minutes for all samples but the rise and thereby potentially burst size showed a large variation, especially for samples in the absence of λ CI. Red curves show 3 replicates of the phage sextuple mutant in the absence of λ CI and blue show 2 replicates of the same mutant in the presence of λ CI.

4.2 Difference in phage P_R promoter strengths

Although the genetic switch regions – or immunity modules – of lambdoid phages show substantial structural similarity, there are noticeable differences in the details of regulation[114]. The λ and P22 genetic switches seem to share the same molecular mechanisms of regulation, and yet we found substantial differences in how the two repressors λ CI and P22 C2 respond to mutations in their binding sites. This raises the question why there is no stronger selection for P22 C2 to resist mutations at O_{R1}? We found that the λ P_R promoter is substantially stronger than that of P22 (Fig. 4.5), suggesting that repression by P22 C2 does not need to be as efficient to keep the level of expression from P22 P_R low.

Most of the mutations that we introduced into either λ O_{R1} or P22 O_{R1} resulted not only in binding affinity changes for repressors but also for RNAP, as evidenced by the variation in fluorescence expression in the absence of repressor (Table 2.2). Hence, changes in O_{R1} operator sites due to divergence of immunity regions will likely lead to changes in P_R promoter strength as well. We tested this hypothesis by measuring fluorescence expression from six different phage P_R promoters from a low copy plasmid and found substantial variation between them (Fig. 4.5). As the CII and Q proteins, which are the master regulators of the lysogenic and lytic pathway after infection, are both expressed from the P_R promoter, this should not affect the balance between them. It can however affect the concentration of the lysogenic repressor (e.g. λ CI or P22 C2) needed to repress P_R, changing the selection pressures working on different parts of the switch. This could also be partly an explanation for the observed variety in molecular regulation mechanisms between immunity regions of different phages[114].

The differences in P_R promoter strengths can also indicate differences in the lifestyles of these phages. Shiga-toxin expressing phages like 933W were found to have a much higher rate of spontaneous induction frequency than phages like λ, P22 and 434 that do not express toxins[238]. Accordingly, we found that 933W has the strongest P_R promoter, which together with its low transcription rate of CI during lysogeny[238] explains the high incidence of spontaneous induction. λ and 434 on the other hand show high lysogen stability, stemming from high lysogenic CI transcription rates[238] coupled with less strong P_R promoters. Additionally, the reduced production of CI transcripts in 434 as compared to λ (by about one-third) seems to be balanced by a similarly reduced P_R strength (Fig. 4.5) to maintain lysogen stability. Hence, P_R promoter strength and lysogenic repressor expression might co-evolve to produce lifestyles that benefit phages in a specific environments.

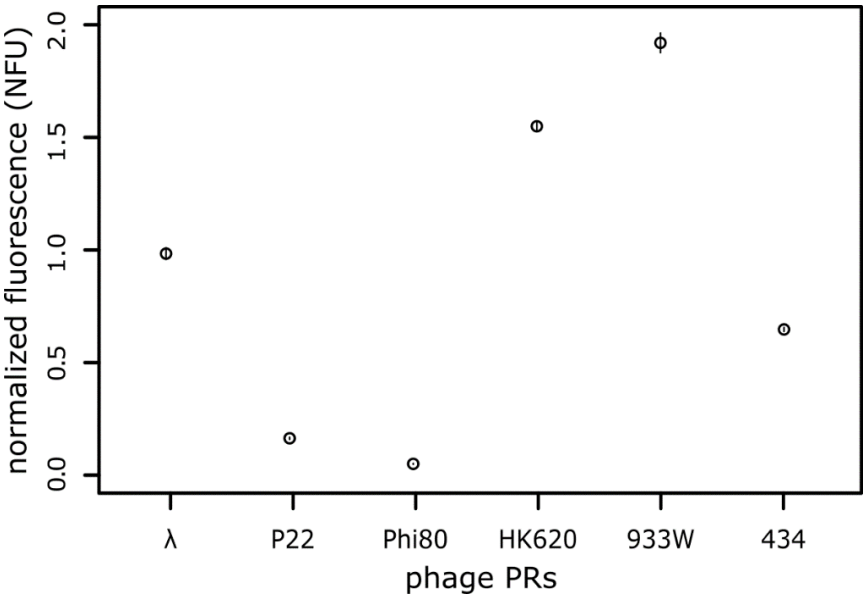

Figure 4.5. Comparison of phage P_R strengths for different phages.

The strength of different P_R promoters was measured through fluorescence expression and normalized by λ P_R strength (NFU).

4.3 Methods

Creation of mutant phages

A derivative of the *S. enterica* LT2 strain was lysogenized with P22 phage from a high titer lysate by creating a dilution series of the lysate and plating them on a lawn of stationary *S. enterica* cells using LB plates. Bacterial lawns were created by mixing 3ml LB soft agar with 100µl of cells grown in M9 medium (1x M9 salts (12.8 gl^{-1}, $Na_2HPO_4.7H_2O$, 3 gl^{-1} KH_2PO_4, 0.5 gl^{-1} NaCl, 1 gl^{-1} NH_4Cl), 2 mM MgSO4, 0.1 mM $CaCl_2$) at 37 °C. Turbid plaques were picked and confirmed by streaking on mint green plates (10 g Bacto tryptone, 5 g Bacto yeast extract, 5 g NaCl, 2.5 g d-glucose, 12 g Ameresco agar) for phage lysis. Mutations in P22 O_{R1} were created by λ Red recombineering[239] using short single-stranded DNA oligos (73bp)[219]. Plaques of mutant phages were picked according to morphology, as they displayed distinct bull's eye morphologies (see text). Mutants were confirmed by sequencing the entire P_{RM}-O_R-P_R region. All plates were incubated at 37°C overnight.

Lysate preparation

Phage lysates were prepared by plate lysis. Specifically, individual phage plaques were picked with a sterile pipette tip, resuspended in 3 ml of phage soft agar together with 100 µ l of overnight bacterial culture and plated on top of LB plates. The plates were then incubated at 37 °C overnight. The soft agar was scraped with a sterile microscope glass slide, resuspended in 10 ml of SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 200 mM Tris-Cl (pH 7.5)) with a few drops of chloroform to kill the residual bacteria. The lysates were then centrifuged to remove the leftover agar, sterilized by filtration (0.2 µ m) and stored at 4 °C.

Estimating burst size and latent period[240]

S. enterica cells from overnight cultures were diluted 1:100 into M9 medium and grown for 4h. Lysate was added to reach 10^6 pfu/ml and phages were adsorbed for 20 minutes at 37°C. After that time, the mixture was diluted 100-fold into fresh M9 medium and incubated at 37°C with vigorous shaking. Samples were taken every 10 minutes and several dilutions were plated on *S. enterica* lawns. One sample at the beginning of the experiment was filtered and plated to determine the number of unabsorbed bacteria.

Measuring strength of different P_R promoters

We constructed a template plasmid carrying a fluorescence marker gene *venus-yfp*[120] under the control of the P_R promoter region, containing an RNAP binding site as well as two operators, O_{R1} and O_{R2}. The P_R region was taken either from Lambda, P22, phi80, HK620, 933W or 434. Specifically, for Lambda P_R we used the region from -60bp upstream of the transcriptional start site to +9bp downstream. To our knowledge the specific location of the transcriptional start site has not been defined for the other phage P_R promoters. Therefore, upstream of O_{R2} and downstream of O_{R1} we used the wild type phage sequences that were of the same bp length as the analogous Lambda P_R regions.

We measured fluorescence of all phage P_R promoters using a Bio-Tek Synergy H1 platereader. Three biological replicates of each phage P_R promoter were grown at 37°C overnight in M9 media, supplemented with 0.1% casamino acids, 0.2% glucose and 30µg/ml kanamycin. Overnight cultures were diluted 1,000X, grown to OD_{600} of approximately 0.1 for fluorescence measurements. All replicate measurements were randomized across multiple 96-well plates. All measurements showed fluorescence levels significantly above the detection limit of the plate reader. Fluorescence values were divided by OD_{600} values (in RFU=Relative Fluorescence Units) and averaged over three replicates. All mean fluorescence measurements were normalized by the mean fluorescence value of Lambda P_R.

4.4 Author contributions

C.I. (Claudia Igler), C.C.G. (Călin C. Guet) conceived the study together. C.I. designed and carried out the experiments and analyzed the data. M.P. (Maroš Pleška) assisted in the experimental design. C.I. wrote the current draft of the manuscript and revised it together with S.A. (Steve Abedon).

5. Non-specific TF binding inhibits cellular growth

5.1 Abstract

Correct functioning of cellular programs depends on the appropriate expression of genes, which is largely determined through binding of transcription factors (TFs) to specific DNA targets. Non-specific binding of TFs however, might significantly interfere with cellular programs. Although it has been suggested that many TFs might be bound non-specifically to DNA most of the time, the overall impact of non-specific binding on the cell remains unclear. We show experimentally that gratuitous expression of phage repressors in *Escherichia coli* and *Salmonella enterica* significantly reduces fitness under certain conditions. The growth reduction is dependent upon a (i) repressor's ability to bind DNA, (ii) its ability to form cooperative interactions and (iii) its concentration relative to the concentration of DNA. We find that non-specific binding due to promiscuous or non-native TFs can be detrimental to the cell and might put a constraint on regulatory design.

5.2 Introduction

Although it has been acknowledged for decades that cellular gene expression is controlled by specific binding of transcription factors (TFs) to DNA sites[73,202], the suggestion that non-specific binding is a decisive factor in gene regulation[52] has received less attention. Indeed, many regulatory proteins might be bound to DNA non-specifically most of the time[52,58,129]. This might provide the TF with an advantage, speeding up the target search due to 1D-sliding on DNA (facilitated diffusion)[44,203] or protection of TFs from degradation[204]. However, non-specific binding could also disturb the cellular program, as modeling indicates that DNA occupancy at non-specific sites can interfere with binding of TFs at their target sites, potentially imposing global constraints on the regulatory architectures of cells[115]. Hence, non-specific TF binding seems to be an abundant mechanism with possibly far-reaching consequences on gene regulation, and yet the overall ramifications of non-specific TF-binding on cellular fitness remain unclear as they have not been investigated experimentally so far.

5.3 Results

Experimental setup

Here, we test the effects of non-specific TF binding employing gratuitous expression of DNA-binding proteins, i.e. TFs without a cognate binding site, on the host cell. Therefore, we cloned a phage repressor gene (*λ cI* or *P22 c2*) under the control of an aTc-inducible promoter (P_{tet}) on a low copy number plasmid (Fig. 5.1A). In their natural systems, both phage repressors bind to one operator as a dimer, but they can also bind cooperatively to adjacent operators or form long- and short-distance loops involving two to four dimers[96,121,156,205,206] (Fig. 5.1A). The plasmids were transformed

into *E. coli* and *S. enterica* cells, which are each the native host for one of the phages - λ and P22, respectively - but do not contain operator sites for either repressor. As we wanted to capture the overall impact of non-specific binding on cellular fitness, we used growth as a global determinant for the effects of repressor expression.

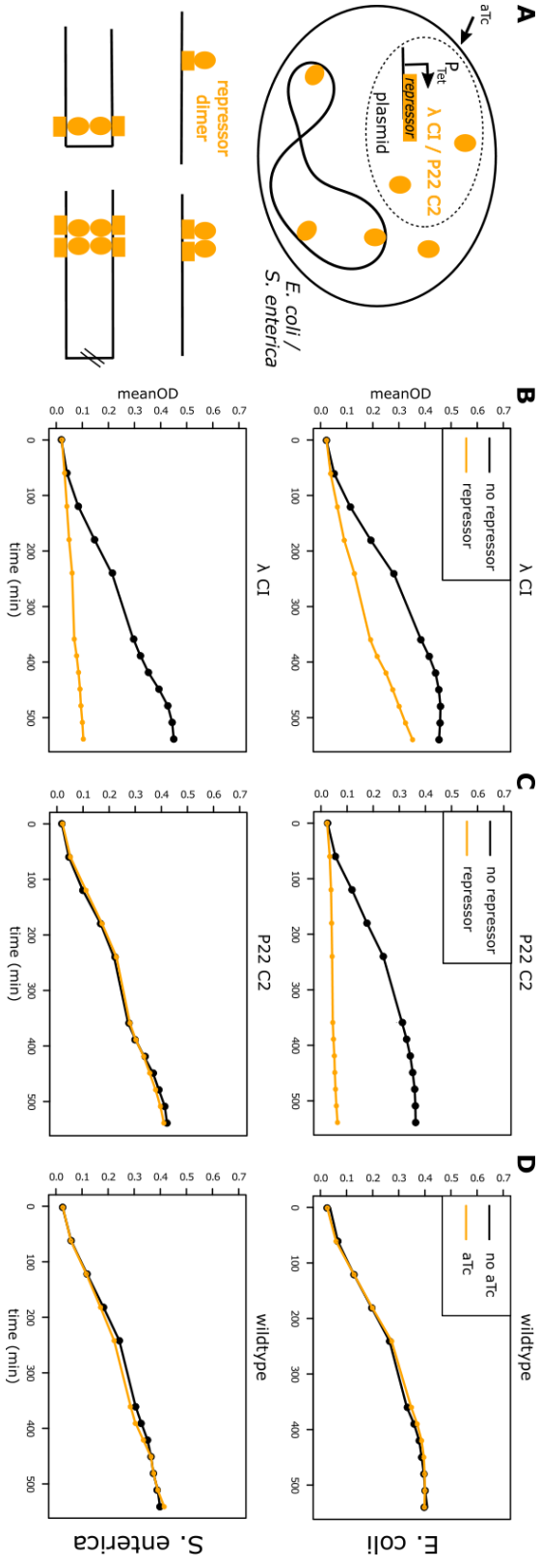

Figure 5.1. Growth defects in the presence of repressors in minimal media.

A) *A repressor (either λ CI or P22 C2) was put under the control of P_{tet} on a low-copy number plasmid and introduced into E. coli or S. enterica cells (upper). Both repressors bind as dimers and form cooperative interactions and loops (lower). Curves show mean OD_{600} for E. coli (upper) or S. enterica (lower) cells grown in minimal medium with glucose in the presence (yellow) or absence (black) of* ***(B)*** *λ CI* ***(C)*** *P22 C2, or just the chemical inducer aTc in cells without plasmid* ***(D)****; error bars show standard deviation. The x-axis shows time in minutes.*

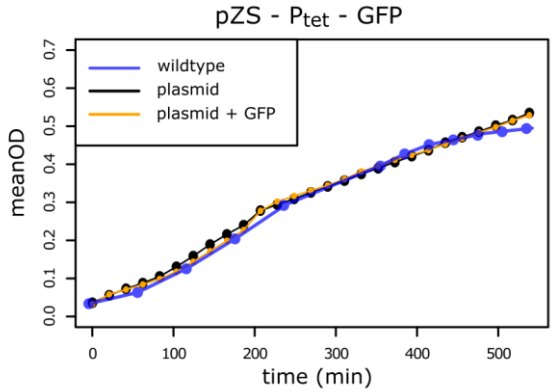

Figure 5.2. Expression of a fluorescence marker from P_{tet} on the pZS plasmid used for repressor expression.

Mean OD_{600} is shown for E. coli cells (growth curves are similar in S. enterica, not shown) grown in minimal media with glucose in the absence (blue) and presence of the plasmid (black), as well as when a fluorescence gene under the control of P_{tet} is expressed from the plasmid (yellow). Error bars give standard deviation. The x-axis shows time in minutes.

Growth effects in different media and induction treatments

For cells grown in minimal media with glucose over 10h the presence of λ CI resulted in a strong reduction of growth in *E. coli* and *S. enterica* when compared to growth in the absence of λ CI, with a stronger effect found in *S. enterica* (Fig. 5.1B, Table 5.1). P22 C2 on the other hand, showed no effect in *S. enterica*, while stopping growth completely when expressed in *E. coli* (Fig. 5.1C, Table 5.1). We found no growth changes due to the addition of aTc (Fig. 5.1D) or the presence of a plasmid alone; nor due to expression of a fluorescence marker from the same plasmid construct (Fig. 5.2).

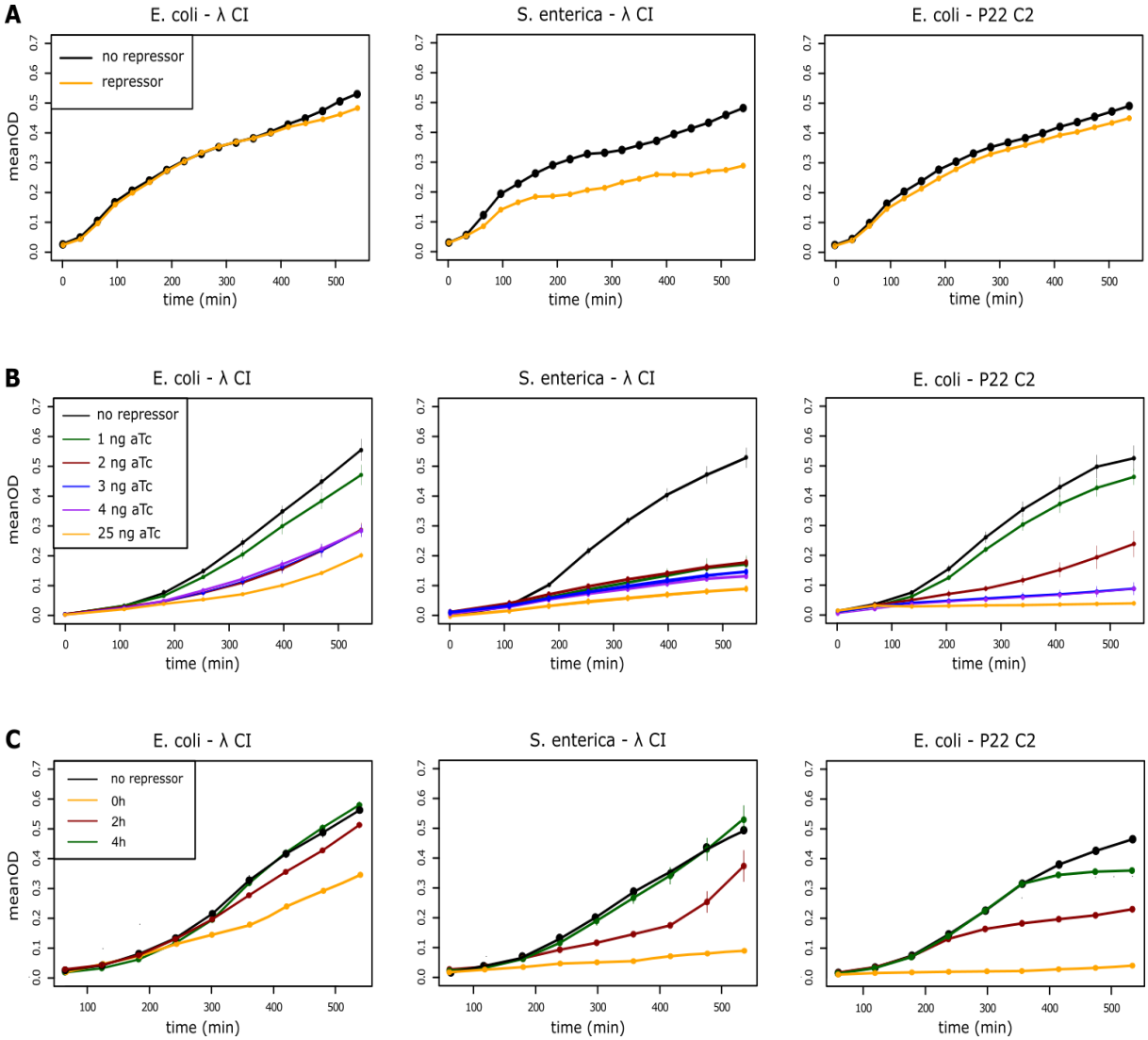

Figure 5.3. Conditional impact of repressors on cellular growth.

Curves show mean OD_{600} for E. coli or S. enterica cells grown in the presence (indicated color) or absence (black) of λ CI or P22 C2; error bars show standard deviation. The x-axis shows time in minutes. ***A)*** *Cells were grown in rich media (LB) at full induction.* ***B)*** *Inducer concentrations for repressor expression were varied from 1-25ng.* ***C)*** *Induction time points of repressor expression were varied from lag phase (0h) to early- and mid-exponential phase (2h and 4h).*

We further explored the conditions for repressor-mediated growth changes by varying the environmental conditions in which the cells were grown. In rich media, the growth defect was abolished completely in *E. coli* for both repressors, and substantially reduced with λ CI expressed in *S. enterica* (Fig. 5.3A, Table 5.2). Minimal media supplemented with Casamino acids (and glycerol or glucose) resulted in intermediate growth reductions between rich and poor media (Fig. 5.4, Table 5.3). P22 C2 did not have an effect on growth in *S. enterica* in any of the conditions (Table 5.1, 5.2, 5.3) and will not be discussed in further experiments.

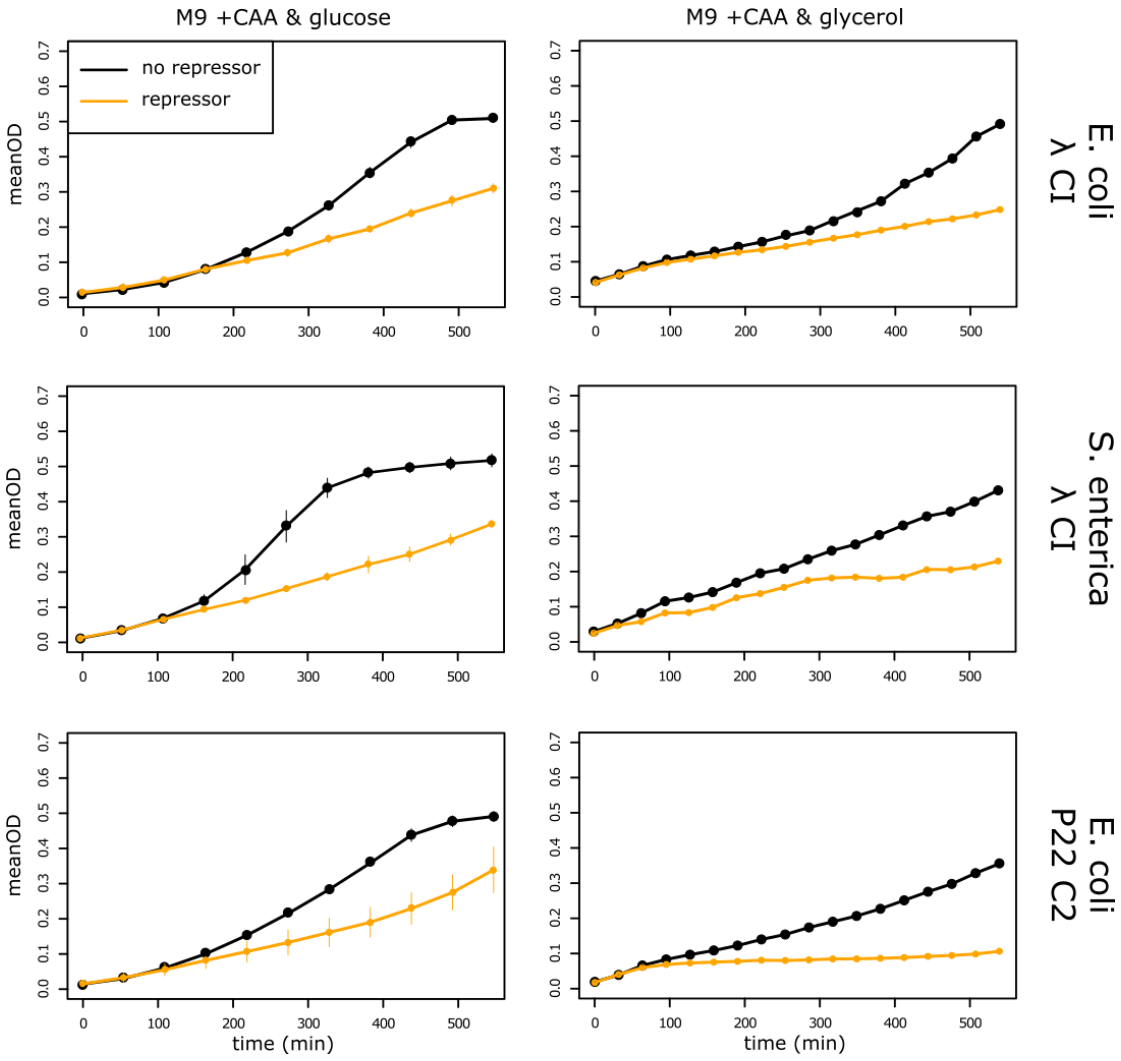

Figure 5.4. Expression of repressors in M9 minimal media supplemented with Casamino acids (CAA) and glucose (left) or glycerol (right).

Curves show mean OD_{600} for E. coli or S. enterica cells grown in the presence (yellow) or absence (black) of λ CI or P22 C2; error bars show standard deviation. The x-axis gives time in minutes.

In addition to nutrient availability, we tested the dependence of the growth defect on repressor concentration and induction timing. In *E. coli*, decreasing repressor concentrations showed a gradual recovery of normal growth (Fig. 5.3B, Table 5.4, 5.5), whereas even low expression of λ CI in *S. enterica* resulted in significant growth reductions (Fig. 5.3B, Table 5.6). We also found that the induction time point is a significant determinant - surprisingly however only relieving λ CI-induced growth defects. Whereas λ CI induction in early- and mid-exponential growth (as opposed to induction during the lag phase) gradually abolished the growth defect in *E. coli* and *S. enterica* (Fig. 5.3C, Table 5.7), this was not the case for P22 C2, where growth was always halted 1-2h after repressor induction (Fig. 5.3C, Table 5.7). Hence, gratuitous expression of repressors resulted in significant growth defects under various conditions.

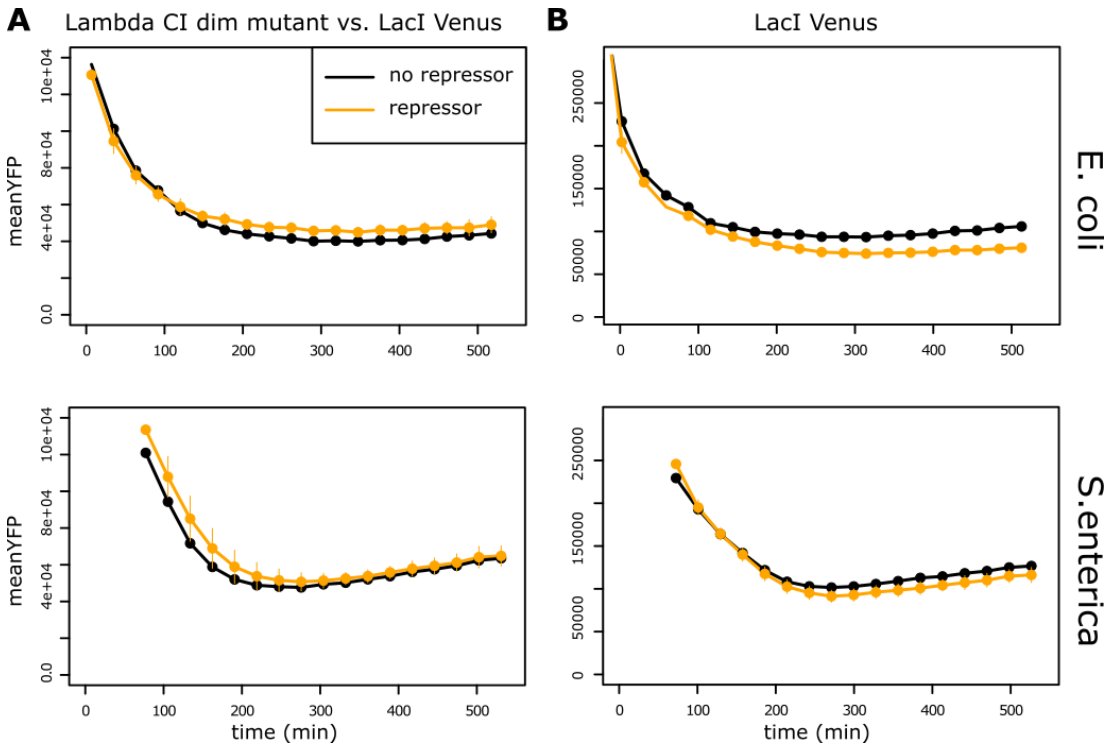

Figure 5.5. Controls for the competition assays.

*Curves show mean fluorescence for E. coli (upper) or S. enterica (lower) cells grown in minimal media with glucose in the presence (yellow) or absence (black) of a λ CI dimerization mutant **(A)** and LacI-Venus or LacI-Venus alone **(B)**; error bars show standard deviation. Cells containing the λ CI dimerization mutant were mixed 1:1 with cells containing LacI and a constitutively expressed Venus marker. Fluorescence (indicative of LacI-plasmid cells) was measured over several hours (x-axis shows time in minutes).*

Fitness effects of gratuitous repressor expression

We tested if the phage repressor-mediated growth effects also reduce cellular fitness using direct competition with cells that expressed a non-toxic TF (LacI) from the same plasmid construct (Fig. 5.5). Plasmids containing LacI were additionally labeled with a constitutive *Venus* marker, which also showed no fitness effect. 1:1 mixtures of cells with phage repressor plasmids and cells with LacI plasmids were grown in minimal media. The fluorescence production in the cell mixtures was compared between the absence (no fitness effect of the plasmids) and presence of repressor induction (growth reduction in phage repressor-expressing cells). In accordance with previous experiments, the expression of repressors led to a significant increase in LacI-expressing cells, except for P22 C2 in *S. enterica* (Fig. 5.6, Table 5.8).

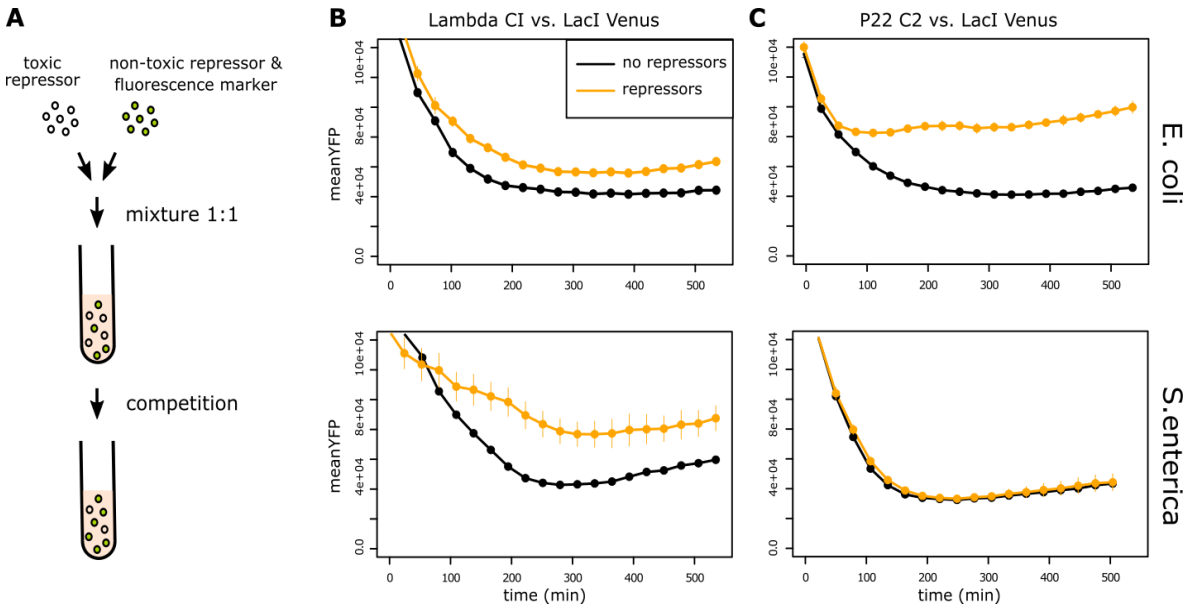

Figure 5.6. Competition assays.

(A) *Cells containing a toxic repressor (λ CI or P22 C2) were mixed 1:1 with cells containing a non-toxic repressor (LacI) and a constitutively expressed Venus marker. Cells were competed over 20h and fluorescence was taken as a measure of the relative change in LacI containing cells versus phage repressor containing cells. Curves show mean fluorescence for E. coli (upper) or S. enterica (lower) cells grown in minimal media with glucose in the presence (yellow) or absence (black) of λ CI* ***(A)*** *or P22 C2* ***(B)*** *and LacI; error bars show standard deviation. Fluorescence (indicative of LacI-plasmid cells) was measured over several hours (x-axis shows time in minutes). Selection coefficients (for calculation see Methods) after 20h were 0.16 (E. coli) and 0.35 (S. enterica) in* ***(A)*** *and 0.63 (E. coli) and 0.09 (S. enterica) in* ***(B)****.*

Distributed, non-specific binding as a basis for growth effects

Accordingly, gratuitous expression of phage repressors can be exceedingly detrimental to cellular growth, but what is the cause of the defect? By definition, repressors are DNA-binding proteins and could interfere with the cellular program through non-specific DNA binding. λ CI only binds DNA in dimeric form[207], so we tested expression of a mutant that cannot form dimers[134], as well as of a mutant defective in DNA binding, and found that the growth and fitness effect of repressor expression was almost completely abolished in *E. coli* as well as in *S. enterica* (Fig. 5.7A, 5.8, Table 5.8, 5.9). (As neither of these mutants has been characterized for P22 C2 so far, we only performed this experiment with λ CI). We found similar results for a λ CI mutant defective in cooperativity between repressor dimers (Fig. 5.7B, Table 5.9). To rule out that repressor misfolding or aggregation was responsible for the growth defect, we over-expressed a chaperone gene (*tig*) together with the repressors, which however did not diminish the detrimental effect (Fig. 5.9, Table 5.10). Hence, the ability to bind DNA – especially in a cooperative manner – seems to be crucial for repressor-mediated growth effects. In agreement with previous studies on λ CI and other TFs[44,58,203,208], this indicates substantial non-specific binding, which is supported through repressor oligomerization[209,210].

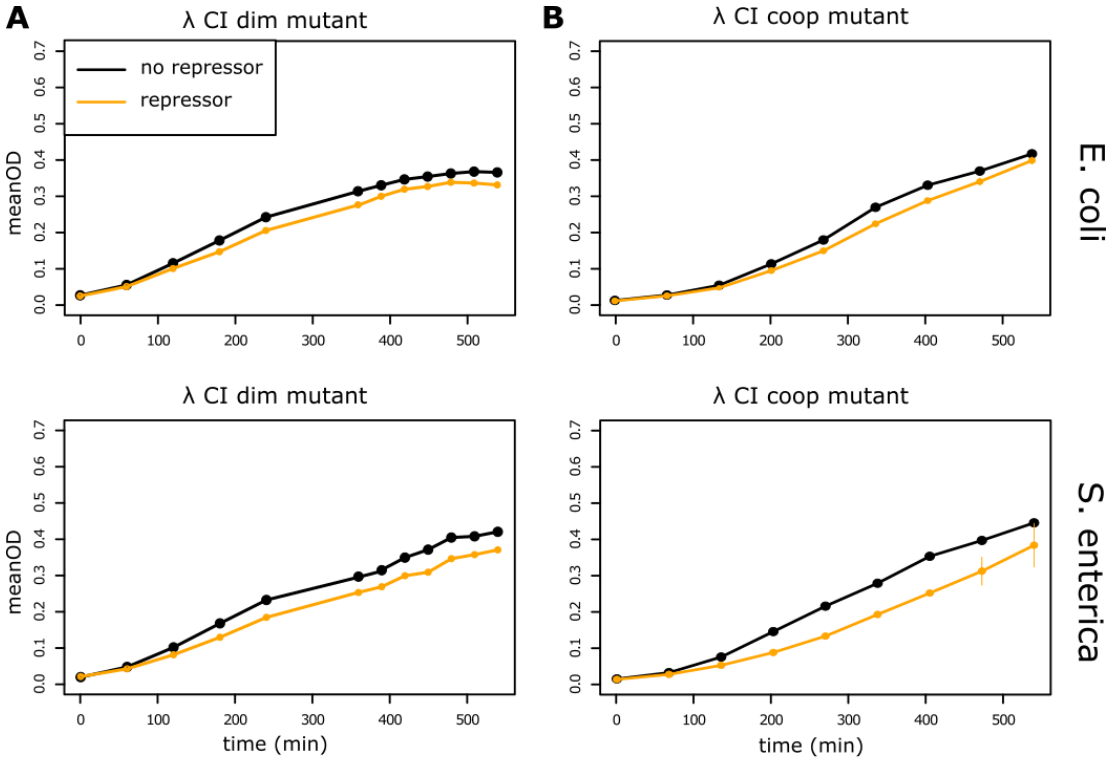

Figure 5.7. λ CI mutants support non-specific binding of wildtype λ CI.

*Curves show mean OD_{600} for E. coli (upper) or S. enterica (lower) cells grown in minimal medium with glucose in the presence (yellow) or absence (black) of a λ CI dimerization mutant **(A)**, or a λ CI cooperativity mutant **(B)**; error bars show standard deviation. The x-axis shows time in minutes.*

Non-specific binding might reduce growth by binding to and interfering with a few essential bacterial genes, so we performed ChIP sequencing to search for target genes of λ CI non-specific binding in *E. coli* and *S. enterica*, using HA-tags, which did not change the growth patterns caused by the repressors (Fig. 5.10). However, the data did not reveal strong peaks for any genomic site or obvious essential gene targets, but indicated weak binding at numerous sites all over both chromosomes (Fig. 5.11A). Further, non-specific binding was preferentially found within genes, not intergenic regions (Fig. 5.11B). A simple thermodynamic model of λ CI binding strength across the genome showed a surprising degree of correlation with the number of reads from ChIP sequencing (Fig. 5.12), considering that those models generally perform poorly for low affinity sites[101].

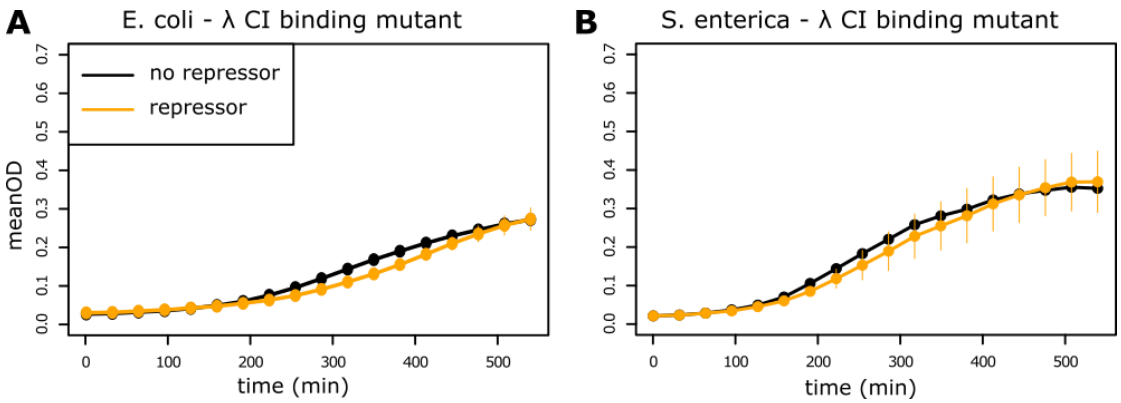

Figure 5.8. Expression of a binding mutant of λ CI abolishes growth defect completely.

*Curves show mean OD_{600} for **A)** E. coli or **B)** S. enterica cells grown in minimal media with glucose in the presence (yellow) or absence (black) of a λ CI binding mutant; error bars show standard deviation. The x-axis gives time in minutes.*

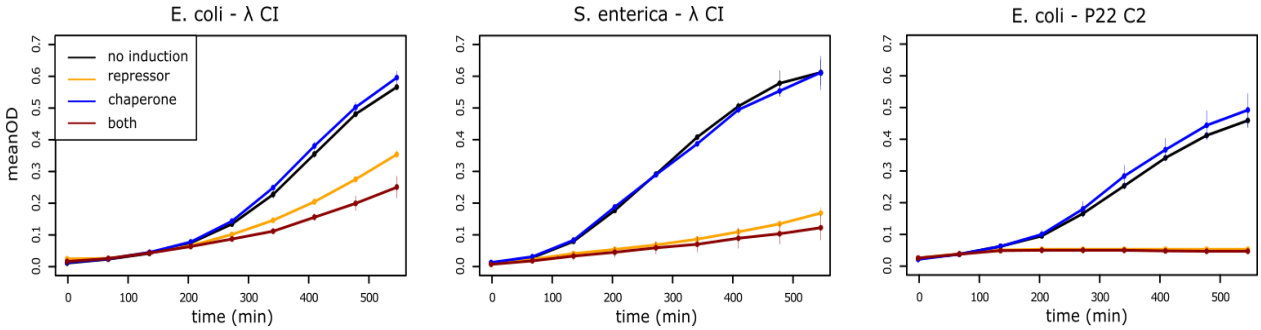

Figure 5.9. Expression of λ CI or P22 C2 together with a chaperone gene in *E. coli* or *S. enterica*.

Additional expression of a chaperone (Tig) that prevents formation of aggregates does not affect mean OD_{600} in minimal media with glucose (blue vs black), but does also not alleviate repressor-mediated growth defects (red vs yellow). Error bars give standard deviation. The x-axis shows time in minutes.

Non-specific binding leads to arrest of cell division

Distributed non-specific repressor binding is in agreement with the observation that increasing concentrations of repressor lead to a gradual increase in the severity of the growth defect. Additionally, the dependence on media and induction timing indicate that DNA concentration - or rather the concentration ratio between repressor and DNA - might also play a role. If cell doubling time is slower than the time needed for DNA replication and cell division (~60min. in *E. coli*[50] and ~50min. in *S. enterica*[211] - which is close to our observed doubling time in minimal media: ~63min. and 58min. respectively), each cell contains only one chromosome. At faster growth, replication cycles are overlapping and daughter cells inherit 2-8 origins at birth, together with partially replicated chromosomes[50]. Hence, the richer the medium and the faster the growth, the more DNA will be available to titrate potentially detrimental non-specific binding proteins. Similarly, cells that are induced during the lag or early-exponential

phase (after 1-2 doublings) will only have one chromosome as they did not inherit partially replicated chromosomes from their mothers and grandmothers yet. We tested this theory by introducing a high copy number plasmid carrying four native λ CI binding sites into *E. coli* cells containing P_{tet} - λ CI. Although expression of λ CI still slowed down growth, growth was significantly faster than for cells not carrying the binding sites (Fig. 5.13). Hence, titration of λ CI alleviates the growth defect, likely even more so if additional chromosomal DNA is present as it provides more potential binding sites than the ~200 from the high copy number plasmid.

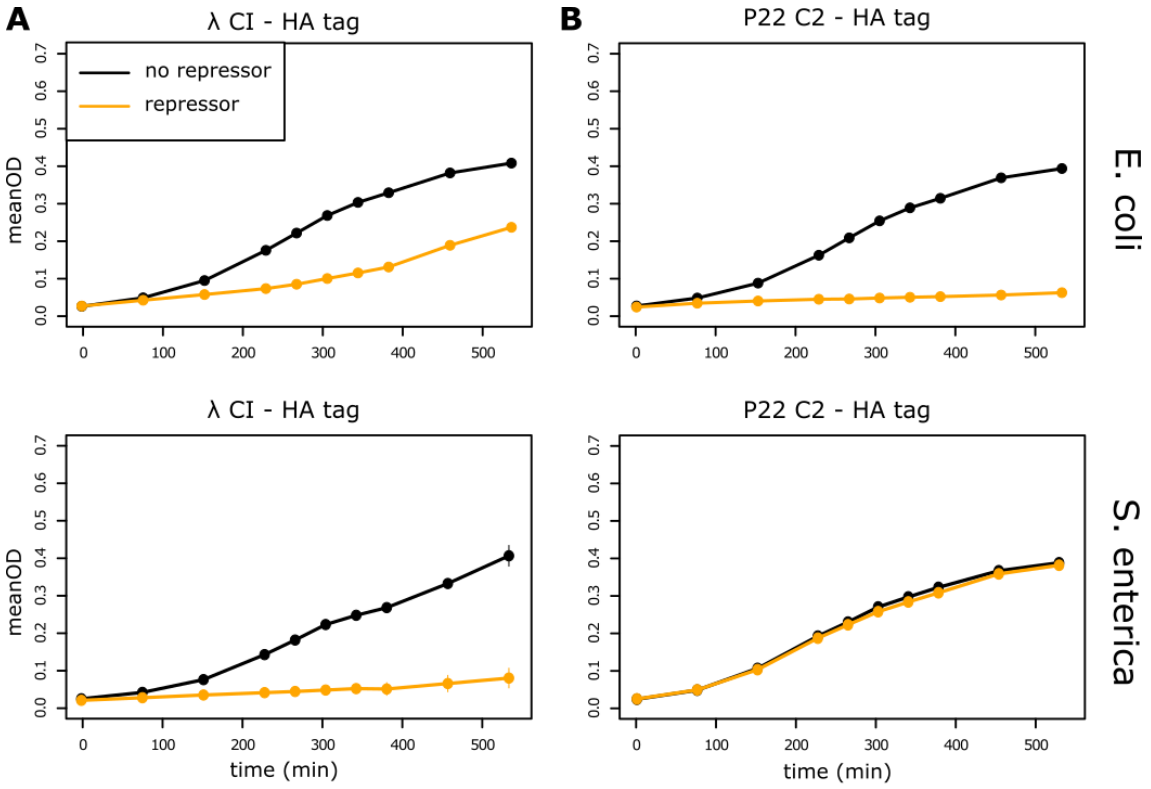

Figure 5.10. HA tags on λ CI and P22 C2 produce the same growth patterns as wildtype repressors in *E. coli* and *S. enterica*.

*Curves show mean OD_{600} for E. coli (upper) or S. enterica (lower) cells grown in minimal medium with glucose in the presence (yellow) or absence (black) of a λ CI with an HA tag **(A)**, or P22 C2 with an HA tag **(B)**; error bars show standard deviation. The x-axis shows time in minutes.*

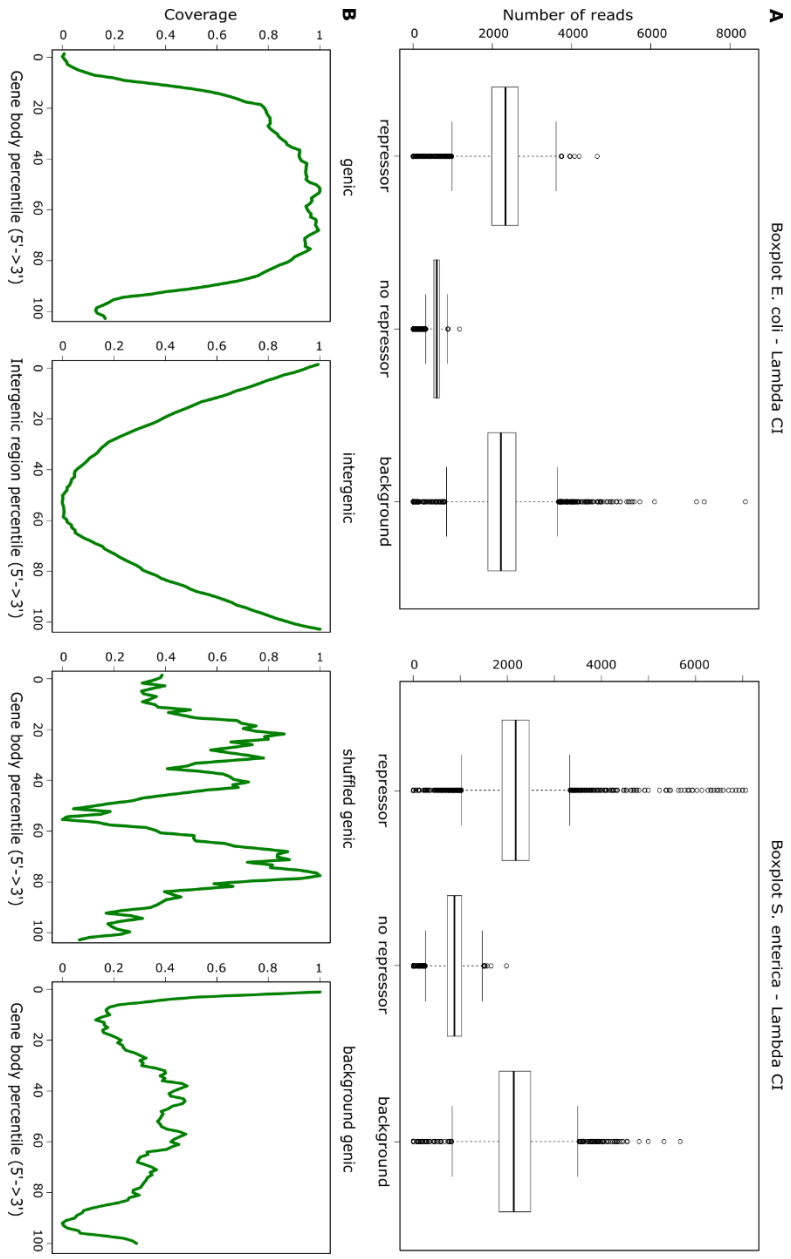

Figure 5.11. Non-specific binding of λ CI in *E. coli* or *S. enterica*.

A) *Boxplots showing the number of ChIP sequencing reads for 1000bp windows across the genome for cells grown in minimal media with glucose in the presence or absence of λ CI and for a background control.* ***B)*** *Binding footprints in genic or intergenic regions show enrichment within genes for λ CI binding in E. coli (footprints in S. enterica are not shown but look similar). Genic and intergenic regions (going from the 5' to the 3' end) were scaled to make them comparable and coverage is given for each percentile over all genic or intergenic regions. A control where reads in genic regions were shuffled as well as the footprint on the background control are shown for comparison.*

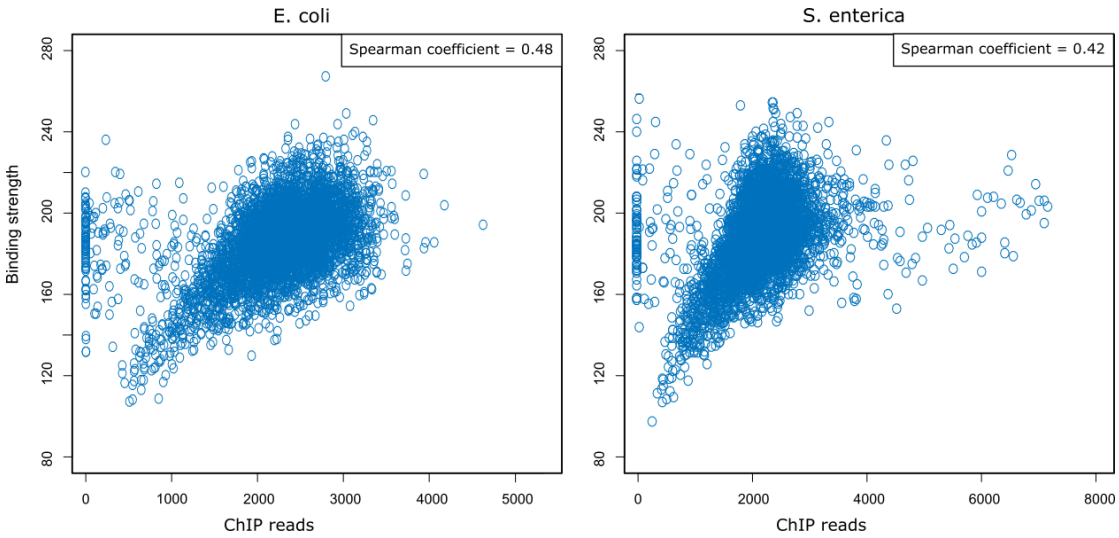

Figure 5.12. Fit between binding strength predictions of a thermodynamic model and ChIP sequencing reads.

*Plots show the correlation between binding strength predicted by a simple thermodynamic model using the energy matrix for λ CI binding and ChIP sequencing reads across 1000bp windows along the **A)** E. coli or **B)** S. enterica genome. The calculated Spearman coefficients for these correlations are shown.*

We used fluorescence microscopy of *E. coli* cells containing P_{tet} - λ CI to investigate if repressor expression was interfering with DNA replication or cell division. First, we imaged cells expressing a SeqA-gfp fusion protein, which is an indicator of replication fork progression. Even though most of the imaged cells formed long filaments, the fluorescent dots suggest ongoing replication but no division in those filaments (Fig. 5.14A, 5.15). Most of the filamentous cells did not show substantial activation of the stress response (Fig. 5.14B, 5.16), which would lead to self-cleavage of the repressor molecules[212] (relieving the stress), but also temporally to inhibition of cell division[213]. Cell division is dependent on FtsZ ring formation at the cell midpoint, which can be inhibited by the stress response but also by nucleoid occlusion, meaning that rings cannot form as long as the chromosome is located at the cell midpoint[214]. By using a Fis-gfp fusion protein, which is an unspecific DNA-binder, we found that in many filaments at least one chromosome was located close to the cell center (Fig. 5.14C, 5.17). This could also explain why some filaments manage to start dividing again at one end after growing to a substantial length, as FtsZ ring formation starts occurring at quarter points in filamentous cells[214]. Therefore, it seems that cell replication is not substantially affected, but that chromosomes are retained at the cell midpoint, thereby inhibiting cell division.

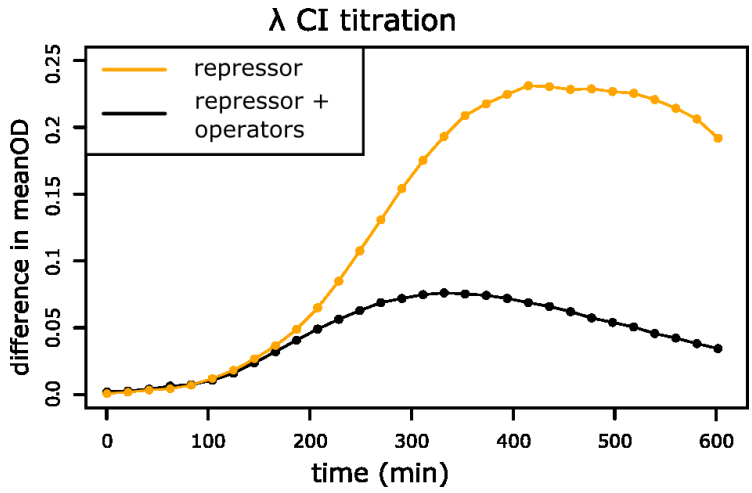

Figure 5.13. Titration of λ CI by target binding sites on a plasmid.

Curves show differences in mean OD_{600} for E. coli cells grown in minimal media with glucose containing λ CI and an empty high copy number plasmid (yellow) or λ CI together with 4 native λ CI binding sites on a high copy number plasmid (black). Differences in growth were calculated as the difference in growth between conditions with atc (repressor induction) and without atc. Error bars show standard deviation; the x-axis gives time in minutes.

5.4 Discussion

Overall, we find that non-specific binding of regulatory proteins can put a significant cost on the host cell, arresting growth and inhibiting cell division. The severity of the cost depends on the repressor's ability to bind DNA non-specifically – which seems to be enhanced through binding cooperativity - as well as its concentration relative to the DNA concentration within the cell. Slow cell growth compounds the effect as cells contain less DNA but accumulate more proteins than at fast growth[40]. Additionally, stress tolerance could be higher at fast growth in rich media. It does not seem likely however that media-specific genes are targeted through non-specific binding, as ChIP sequencing revealed distributed, weak binding by λ CI. Rather, inhibition of cell division seems at least partially to be due to nucleoid occlusion, as chromosomes tend to accumulate at the cell center. Clearance of the division site is impeded if sister chromosomes fail to be completely segregated[215], which could be caused by structural interference due to looping between non-specifically bound repressor molecules, explaining the dependence of the growth defect on cooperativity.

Our findings suggest that the potentially detrimental effect of DNA-binding proteins could be a universal problem that cells are facing due to cellular crowding[45], horizontal gene transfer[216], or mutations that affect the binding specificity of a TF. λ CI, which is a more promiscuous TF[135], reduces growth in *E. coli* and *S. enterica*, indicating a global constraint on promiscuous binders within cells, as implicated by previous modeling[115]. P22 C2, however, was shown to be a rather specific binder[135], which agrees with the observation that it was only detrimental in *E. coli*, but not in its natural host. This might also explain why later induction does not relieve the growth defect as P22 C2 is not titrated away by additional, partially replicated chromosomes if they do not contain

its ‘more specific’ non-specific sequences. Non-specific binding can exert a significant fitness cost on bacterial cells and might put a constraint on regulatory architectures, for example constraining the number of DNA-binding proteins and the influx of foreign genes. As the repressors, which we used, originated from temperate phages, interference with host cell growth might also limit the potential host range of temperate phages with regard to successfully establishing lysogeny. We found however no growth defect for two different phage repressors (Fig. 5.18) or for λ CI and P22 C2 in a different host strain (Fig. 5.19), which further indicates that the effect is dependent on the specific strain-TF combination.

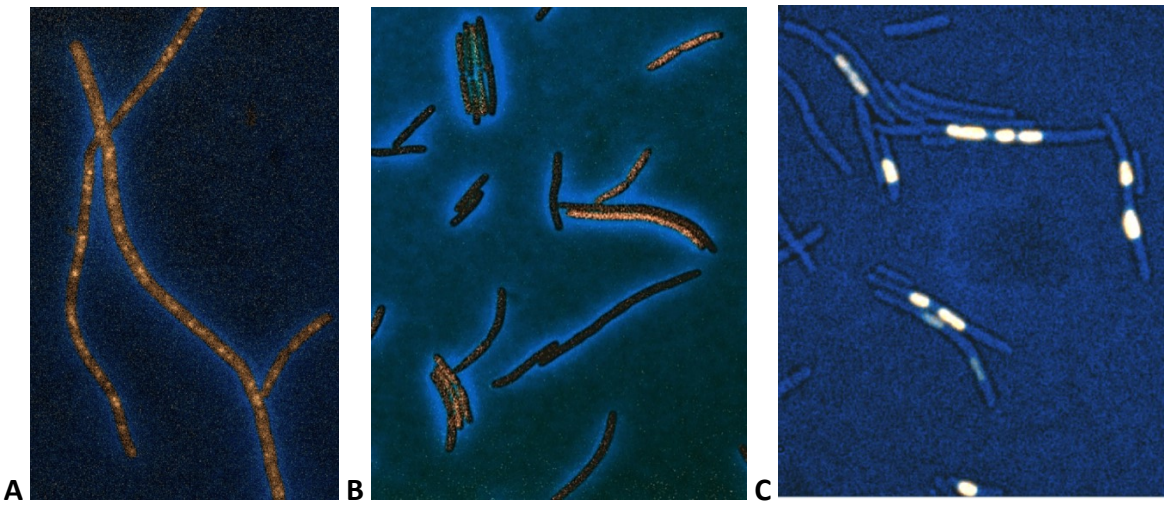

Figure 5.14. Microscope images of *E. coli* cells grown in minimal media with glucose and induced for λ CI.

*Fluorescence indicates **(A)** replication progression (SeqA-gfp fusion), **(B)** induction of the stress response (P_{sulA}-yfp reporter), and **(C)** chromosome positioning (Fis-gfp fusion).*

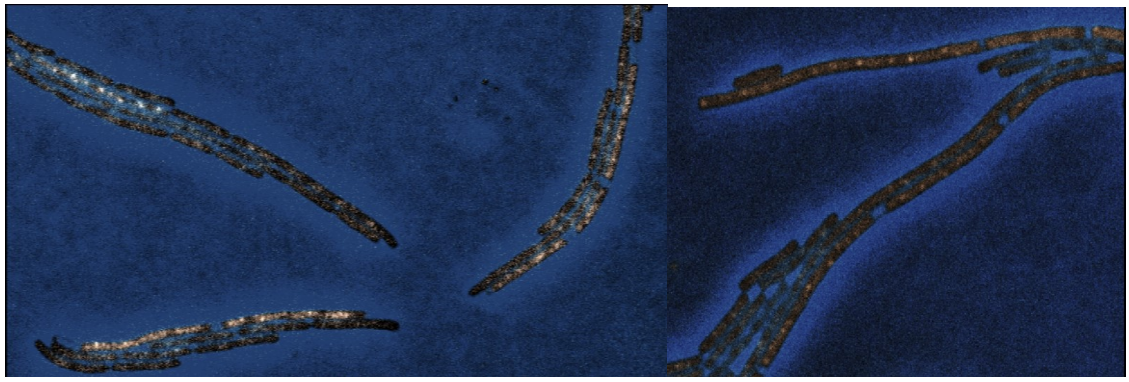

Figure 5.15. Microscope images of cell replication in *E. coli* cells grown in minimal media with glucose and induced for λ CI.

Fluorescence indicates replication progression (SeqA-gfp fusion).

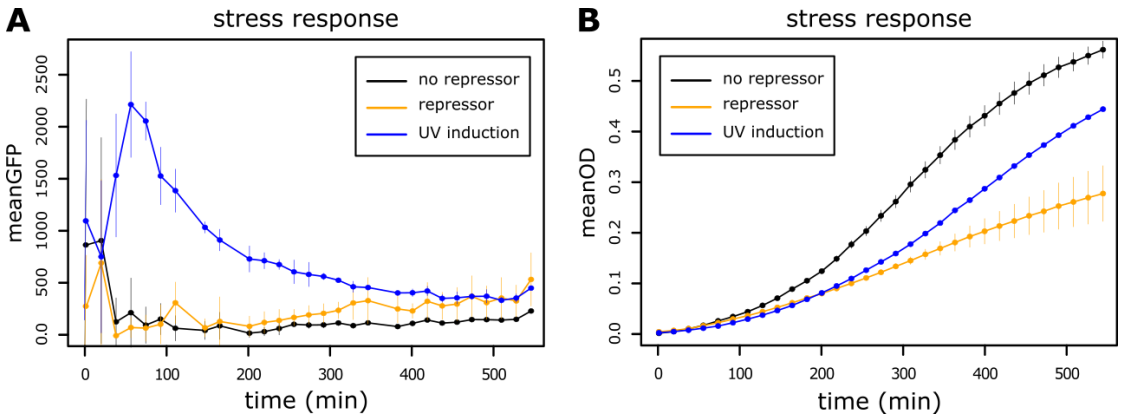

Figure 5.16. Stress response induction in *E. coli* with λ CI.

E. coli cells containing an stress response reporter gene (P_{sulA}-yfp) and λ CI were grown in minimal media with glucose. Cells were either not induced (black), or induced for different stressors: with aTc, producing λ CI (yellow), or with UV exposure for 30sec (blue). UV induction leads to a spike in the stress response and an initial decrease in growth, whereas λ CI does not induce the stress response but leads to an increasing growth defect.

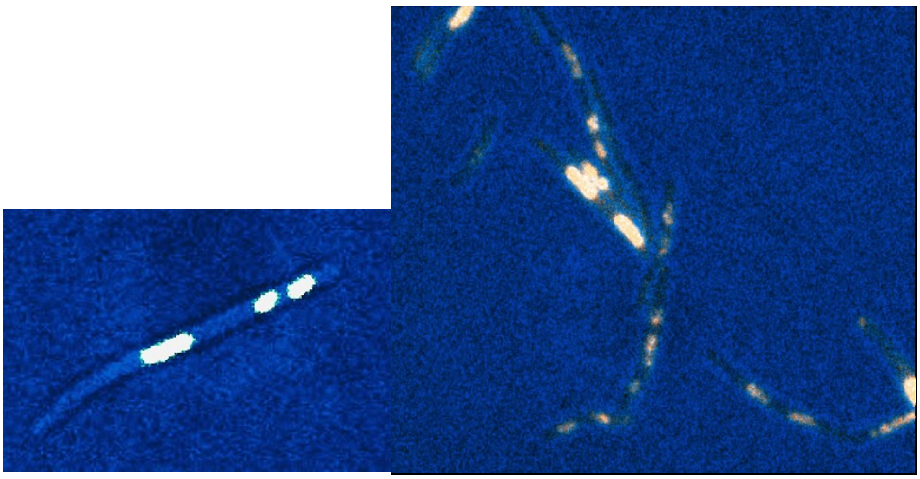

Figure 5.17. Microscope images of cell division in *E. coli* cells grown in minimal media with glucose and induced for λ CI.

Fluorescence indicates chromosome positioning (Fis-gfp fusion).

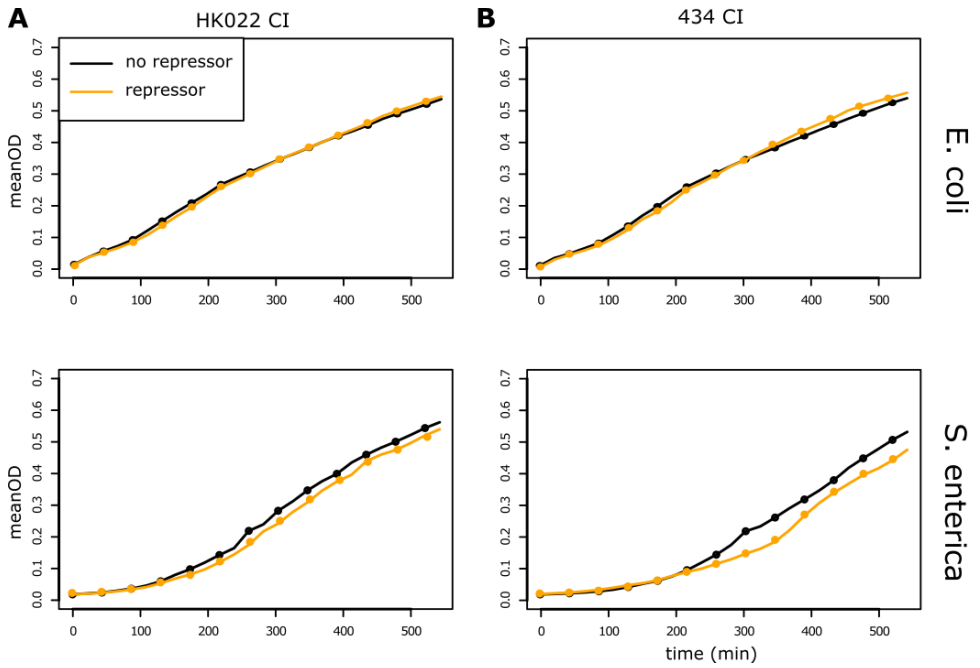

Figure 5.18. No growth defect due to phage repressors from 434 and HK022.

*Curves show mean OD_{600} for E. coli (upper) or S. enterica (lower) cells grown in minimal medium with glucose in the presence (yellow) or absence (black) of a HK022 CI **(A)**, or 434 CI **(B)**; error bars show standard deviation. The x-axis shows time in minutes.*

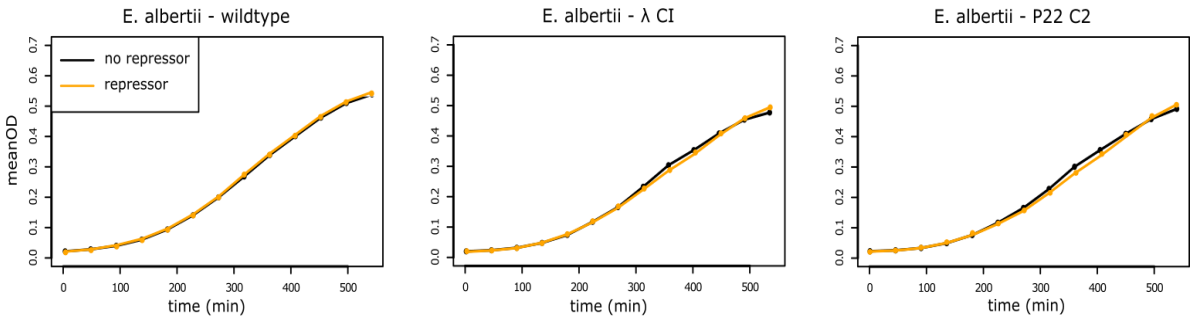

Figure 5.19. No growth defect of λ CI or P22 C2 in *E. albertii.*

Curves show mean OD_{600} for E. albertii cells grown in minimal medium with glucose in the presence (yellow) or absence (black) of the chemical inducer aTc in cells without plasmid (left), λ CI (middle), or P22 C2 (right); error bars show standard deviation. The x-axis shows time in minutes.

5.5 Tables

Table 5.1. Statistical significance of growth changes in minimal media with glucose due to the presence of λ CI or P22 C2 in *E. coli* and *S. enterica*.

FDR-corrected t-test were carried out at four time points to test if growth in the presence of a repressor was significantly different from growth in its absence; p-values and t-values are shown (red values indicate no significant difference).

Time (min)	***E. coli* - λ CI**	***E. coli* - P22 C2**	***S. enterica* - λ CI**	***S. enterica* -P22 C2**
120	p=0,047; f=0,05	p=0,020; f=0,02	p=0,053; f=0,05	p=0,021; f=0,02
240	p=0,002; f=0,00	p=0,000; f=0,00	p=0,002; f=0,00	p=0,047; f=0,05
360	p=0,000; f=0,00	p=0,000; f=0,00	p=0,000; f=0,00	p=0,953; f=0,95
480	p=0,000; f=0,00	p=0,000; f=0,00	p=0,000; f=0,00	p=0,334; f=0,33

Table 5.2. Statistical significance of growth changes in rich media (LB) due to the presence of λ CI or P22 C2 in *E. coli* or *S. enterica*.

FDR-corrected t-test were carried out at four time points to test if growth in the presence of repressor was significantly different from growth in its absence; p-values and t-values are shown (red values indicate no significant difference).

Time (min)	***E. coli* - λ CI**	***E. coli* - P22 C2**	***S. enterica* - λ CI**	***S. enterica* - P22 C2**
120	p=0,491; f=1,19	p=0,064; f=3,20	p=0,020; f=4,44	p=0,871; f=-0,27
240	p=0,967; f=-0,04	p=0,020; f=4,44	p=0,000; f=27,46	p=0,152; f=2,37
360	p=0,767; f=0,50	p=0,020; f=4,51	p=0,000; f=15,75	p=0,606; f=0,92
480	p=0,020; f=4,54	p=0,006; f=6,94	p=0,000; f=29,17	p=0,755; f=0,59

Table 5.3. Statistical significance of growth changes in minimal media supplemented with Casaminoacids and glycerol due to the presence of λ CI or P22 C2 in *E. coli* and *S. enterica*.

FDR-corrected t-test were carried out at four time points to test if growth in the presence of repressor was significantly different from growth in its absence; p-values and t-values are shown (red values indicate no significant difference).

Time (min)	***E. coli* - λ CI**	***E. coli* - P22 C2**	***S. enterica* - λ CI**	***S. enterica* - P22 C2**
120	p=0,182; f=1,75	p=0,008; f=5,68	p=0,045; f=3,27	p=0,144; f=2,00
240	p=0,096; f=2,42	p=0,002; f=9,70	p=0,004; f=7,25	p=0,040; f=3,44
360	p=0,024; f=4,18	p=0,001; f=12,22	p=0,024; f=4,15	p=0,083; f=2,59
480	p=0,006; f=6,20	p=0,001; f=15,39	p=0,003; f=8,13	p=0,968; f=-0,06

Table 5.4. Statistical significance of growth changes due to different concentrations of λ CI in *E. coli*.

FDR-corrected t-test were carried out at four time points to test if growth in minimal media with glucose in the presence of λ CI was significantly different from growth in its absence; p-values and t-values are shown (red values indicate no significant difference).

Time (min)	***E. coli* - λ CI 1ng atc**	***E. coli* - λ CI 2ng atc**	***E. coli* - λ CI 3ng atc**	***E. coli* - λ CI 4ng atc**

120	p=0,755; f=0,40	p=0,100; f=-2,61	p=0,027; f=-4,70	p=0,042; f=-3,81
240	p=0,017; f=-5,74	p=0,001; f=-25,63	p=0,003; f=-11,20	p=0,006; f=-9,00
360	p=0,169; f=-2,00	p=0,001; f=-26,96	p=0,001; f=-27,67	p=0,005; f=-9,34
480	p=0,218; f=-1,71	p=0,002; f=-14,09	p=0,002; f=-14,70	p=0,009; f=-7,23

Table 5.5. Statistical significance of growth changes due to different concentrations of P22 C2 in *E. coli*.

FDR-corrected t-test were carried out at four time points to test if growth in minimal media with glucose in the presence of P22 C2 was significantly different from growth in its absence; p-values and t-values are shown (red values indicate no significant difference).

Time (min)	***E. coli* - P22 C2 1ng atc**	***E. coli* - P22 C2 2ng atc**	***E. coli* - P22 C2 3ng atc**	***E. coli* - P22 C2 4ng atc**
120	p=0,078; f=-2,91	p=0,472; f=-0,87	p=0,011; f=-6,79	p=0,013; f=-6,06
240	p=0,042; f=-3,81	p=0,011; f=-6,62	p=0,001; f=-23,95	p=0,000; f=-28,53
360	p=0,119; f=-2,34	p=0,001; f=-15,94	p=0,001; f=-20,48	p=0,000; f=-32,55
480	p=0,128; f=-2,20	p=0,003; f=-10,98	p=0,001; f=-20,38	p=0,000; f=-32,67

Table 5.6. Statistical significance of growth changes due to different concentrations of λ CI in *S. enterica*.

FDR-corrected t-test were carried out at four time points to test if growth in minimal media with glucose in the presence of λ CI was significantly different from growth in its absence; p-values and t-values are shown (red values indicate no significant difference).

Time (min)	***S. enterica* - λ CI 1ng atc**	***S. enterica* - λ CI 2ng atc**	***S. enterica* - λ CI 3ng atc**	***S. enterica* - λ CI 4ng atc**
120	p=1,000; f=0,00	p=0,308; f=1,32	p=0,008; f=-7,55	p=0,051; f=0,90
240	p=0,068; f=-3,11	p=0,017; f=-5,59	p=0,000; f=-48,00	p=0,005; f=-1,32
360	p=0,004; f=-10,44	p=0,001; f=-20,00	p=0,000; f=-50,20	p=0,004; f=-4,11
480	p=0,001; f=-35,34	p=0,002; f=-15,36	p=0,001; f=-32,94	p=0,008; f=-7,62

Table 5.7. Statistical significance of growth changes due to different induction time points (2 and 4h after inoculation) of λ CI and P22 C2 in *E. coli* and λ CI in *S. enterica*.

FDR-corrected t-test were carried out at four time points to test if growth in minimal media with glucose in the presence of a repressor was significantly different from growth in its absence; p-values and t-values are shown (red values indicate no significant difference).

Time (min)	*E. coli* - λ CI 2h	*E. coli* - λ CI 4h	*E. coli* - P22 C2 2h	*E. coli* - P22 C2 4h	*S. enterica* - λ CI 2h	*S. enterica* - λ CI 4h
120	p=1,000; f=1,19	p=0,095; f=3,20	p=0,160; f=0,86	p=0,285; f=4,44	p=0,267; f=2,01	p=1,000; f=0,00
240	p=0,907; f=-0,04	p=0,026; f=4,44	p=0,033; f=0,39	p=0,095; f=27,46	p=0,001; f=2,30	p=1,000; f=0,00
360	p=0,092; f=0,50	p=0,370; f=4,51	p=0,002; f=-0,06	p=0,451; f=15,75	p=0,001; f=0,88	p=0,026; f=0,00
480	p=0,033; f=4,54	p=0,031; f=6,94	p=0,002; f=1,27	p=0,001; f=29,17	p=0,044; f=0,56	p=0,940; f=0,00

Table 5.8. Statistical significance of fluorescence differences in growth competitions.

Fluorescence was measured in 1:1 mixtures of cells containing plasmids with either a phage repressor (λ CI**,** P22 C2 or λ CI dim mutant) or LacI and a fluorescence marker. FDR-corrected t-test were carried out at four time points to test if the number of *lacI*-containing cells was significantly increased in the presence of inducer as compared to its absence when cells were grown in minimal media with glucose; p-values and t-values are shown (red values indicate no significant difference). The LacI control in the lower table gives the fluorescence difference between absence and presence of LacI in cells containing the LacI-Venus plasmid.

Time (min)	*E. coli* - λ CI vs. LacI Venus	*E. coli*- P22 C2 vs. LacI Venus	*S. enterica* - λ CI vs. LacI Venus	*S. enterica* - P22 C2 vs. LacI Venus
120	p=0,021; f=-15,986	p=0,017; f=-26,851	p=0,047; f=-4,555	p=0,813; f=-0,494
240	p=0,025; f=-12,498	p=0,017; f=-28,534	p=0,047; f=-5,683	p=0,898; f=-0,146
360	p=0,035; f=-8,654	p=0,017; f=-21,504	p=0,047; f=-4,464	p=0,898; f=-0,170
480	p=0,030; f=-10,625	p=0,021; f=-15,064	p=0,047; f=-7,430	p=0,898; f=-0,165
Time (min)	*E. coli* - λ CI dim mutant vs. LacI Venus	*E. coli*- LacI Venus control	*S. enterica* - λ CI dim mutant vs. LacI Venus	*S. enterica* – LacI Venus control
120	p=0,325;f=-2,055	p=0,059;f=5,869	p=0,373;f=-1,688	p=0,802;f=0,442
240	p=0,284;f=-2,362	p=0,014;f=19,614	p=0,723;f=-0,790	p=0,052;f=6,531
360	p=0,284;f=-2,423	p=0,014;f=30,801	p=0,792;f=-0,628	p=0,108;f=4,137
480	p=0,419;f=-1,471	p=0,014;f=22,877	p=0,813;f=-0,482	p=0,234;f=2,527

Table 5.9. Statistical significance of growth changes due to mutant λ CI in *E. coli* and *S. enterica*.

FDR-corrected t-test were carried out at four time points to test if growth in minimal media with glucose in the presence of mutated λ CI (dim… dimerization mutant, bind… binding mutant, coop… cooperativity mutant) was significantly different from growth

in its absence; p-values and t-values are shown (red values indicate no significant difference).

Time (min)	***E. coli* - λ CI dim mutant**	***S. enterica* - λ CI dim mutant**	***E. coli* - λ CI bind mutant**	***S. enterica* - λ CI bind mutant**	***E. coli* - λ CI coop mutant**	***S. enterica* - λ CI coop mutant**
120	p=0,234; f=0,23	p=0,049; f=0,05	p=0,099; f=-3,464	p=0,041; f=8,000	p=1,000; f=-0,27	p=0,033; f=0,00
240	p=0,034; f=0,03	p=0,017; f=0,02	p=0,078; f=4,914	p=0,041; f=8,598	p=0,015; f=2,37	p=0,014; f=0,00
360	p=0,029; f=0,03	p=0,015; f=0,01	p=0,011; f=26,348	p=0,168; f=2,313	p=0,017; f=0,92	p=0,015; f=0,00
480	p=0,045; f=0,04	p=0,016; f=0,02	p=0,079; f=4,323	p=0,616; f=-0,589	p=0,059; f=0,59	p=0,092; f=0,00

Table 5.10. Statistical significance of growth changes due to overexpression of a chaperone gene in *E. coli* and *S. enterica*.

FDR-corrected t-test were carried out at four time points to test if the growth defect caused by phage repressors in minimal media with glucose could be alleviated by additional expression of a chaperone (Tig); p-values and t-values are shown (red values indicate no significant difference).

Time (min)	***E. coli* - λ CI**	***E. coli* - P22 C2**	***S. enterica* - λ CI**	***S. enterica* - P22 C2**
120	p=0,547; f=0,997	p=0,667; f=0,500	p=0,547; f=0,945	p=0,547; f=-1,000
240	p=0,547; f=1,045	p=0,547; f=1,000	p=0,563; f=0,832	p=0,564; f=-0,756
360	p=0,547; f=2,986	p=0,547; f=1,000	p=0,547; f=0,994	p=0,547; f=-1,589
480	p=0,079; f=14,172	p=0,547; f=2,000	p=0,547; f=1,392	p=0,547; f=-1,782

5.6 Methods

Plasmids and strains

λ CI, P22 C2, HK022 CI and 434 CI were cloned under the control of a $P_{LtetO-1}$ promoter onto a low copy number km^R plasmid (pZS)[99]. The plasmids were transformed either into MG1655 derived *E. coli* cells (strain BW27785, CGSC#: 7881)[122] or into LT2 derived *S. enterica* cells or into KF1 derived *E. albertii* cells with a tetracycline cassette inserted in the P22 attachment site (LT2 attP22::tetRA). In control experiments the phage

repressor on the plasmid was replaced by a fluorescence marker gene (*gfp*). Titration of λ CI was tested by transforming *E. coli* cells containing the pZS21- λ CI plasmid with a high-copy number pZE plasmid (50-70 copies)[99], which carries the natural λ CI operators O_{R1}, O_{R2}, O_{L1} and O_{L2}, i.e. 200-280 operators per cell.

For the competition experiments the phage repressor was replaced by LacI and a constitutive fluorescent marker *venus-yfp*[120] was introduced into the plasmid. In order to test for misfolding of repressor proteins, we used a high copy number plasmid containing a chaperone gene (*tig*[217]) – which exists in *E. coli* and *S. enterica* (UniProt) -under the control of a P_{Lac} promoter from the ASKA(-) library[218].

To monitor the induction of the stress response, we used a strain with a fast-maturing yellow fluorescent protein (YFP)[120] fused to the promoter of sulA (PsulA-yfp), which was placed on the chromosome using lambda red recombineering[219]. SulA is strongly upregulated as a part of the stress response[220]. The PsulA-yfp strain was then transformed with the pZS21- λ CI plasmid. We checked induction of the reporter by exposing cells to UV light for 30 seconds.

To visualize the chromosome we used a previously characterized fusion between a green fluorescent protein (GFP) and a major NAP (Nucleosome Assembly Protein) Fis (GFP-Fis) under the control of a P_{Lac} promoter[221]. The construct was placed on a pZE12 plasmid[99] and transformed into *E. coli* cells containing the pZS21- λ CI plasmid.

We used a seqA-gfp translational fusion under the control of the natural seqA promoter to monitor replication as seqA binds hemimethylated GATC sequences in the wake of the advancing replication fork, marking newly synthesized DNA[222]. The fusion protein was inserted into the HK022 attachment site on the *E. coli* chromosome using CRIM plasmids[177].

λ *CI mutants*

Based on previous studies, we produced three different λ repressor mutants that were cloned into the same low copy number plasmid (pZS) under the control of a $P_{LtetO-1}$ promoter: 1. A repressor mutant that cannot form dimers (S228N)[134] and hence not bind DNA effectively anymore; 2. A repressor mutant that can dimerize but not form higher-order oligomers, i.e. that cannot bind cooperatively (Y210N)[134] and 3. A repressor that is defective in its ability to bind DNA (N52D)[223].

Growth measurements

All cells were grown overnight at 37°C in M9 medium supplemented with 0.2% glucose and 50µg/ml kanamycin (except specified differently). Those cultures were used to dilute (1:100) 6 replicates without inducer and 6 replicates with 25ng aTc in 96 well plates and were grown at 37°C on the shaker at 220 rpm. Populations were measured (OD_{600}) every 30min or every 60min using Biotek H1 plate reader for overall 10h. Population growth was also measured in LB, and M9 medium supplemented with 0.5% Casamino acids and either 0.5% glycerol or 0.2% glucose. When indicated inducer concentration was change 1, 2, 3 and 4ng aTc or the induction time was varied from the inoculation point (0h) to 2h or 4h post-inoculation (early- and mid-exponential

phase). The chaperone gene was induced using 1mM IPTG and fis-gfp was expressed by adding 0.1mM IPTG to the medium.
Repressor concentration at 25ng aTc was measured using the Promega Nano-Glo HiBiT Lytic Detection System and gave about 500 dimers per cell in LB. The HiBiT peptide tag was attached at the N-terminal end (which is involved in DNA binding) using a (GGGS)2 linker (sequence: GGTGGTGGTTCTGGTGGTGGTTCT) to assure accessibility of the tag for interaction with the detection reagent.

Competition assays
A single colony for each host strain (*E. coli* or *S. enterica*) – plasmid (pZS21- λ CI, pZS21- λ CI dimerization mutant, pZS21-P22 C2, pZS21-LacI-Venus) combination was picked from a freshly streaked plate and grown overnight in minimal media with glucose supplied with kanamycin. Strains containing a phage repressor plasmid were mixed approximately 1:1 with a corresponding strain carrying pZS21-LacI-Venus and diluted 1:100 into fresh medium and grown in 96-wellplates for 10h. Fluorescence was measured every 30min. and compared between strains that were induced with aTc and strains that were not induced. There was no significant difference in fluorescence between the presence and absence of LacI expression in LacI-Venus-containing cells for *S. enterica* and a slight disadvantage in the presence of LacI in *E. coli* (Table 8), which however only strengthens the advantage of LacI-expressing cells in competition with phage repressor-expressing cells.
The selection coefficients were calculated using $\ln[(R^+_t/R^-_t)^* (R^+_0/R^-_0)]$, where R^+_t and R^-_t represent densities of cells with and without inducer aTc (presence or absence of repressor expression) at time t=20h respectively, and R^+_0 and R^-_0 represent densities at the beginning of the experiment.

Microscope fluorescence measurements
A Nikon Ti-E microscope equipped with a thermostat chamber (TIZHB, Tokai Hit), 100× oil immersion objective (Plan Apo λ, N.A. 1.45, Nikon), cooled CCD camera (ORCA-Flash, Hamamatsu Photonics) and LED excitation light source (DC2100, Thorlabs) was used for the microscopy fluorescence measurements of PsulA-yfp and SeqA-gfp. The microscope was controlled by micromanager (https://micro- manager.org). The cells were grown overnight in minimal media with glucose, diluted 1:100 in fresh media and grown to early exponential phase in the presence of the inducer aTc. Phase, YFP or GFP fluorescence (where appropriate) and RFP fluorescence images were taken simultaneously at 3-min time-lapse interval. Multiple patches of cells were monitored simultaneously in a single experiment. We used a custom macro of ImageJ (http://imagej.nih.gov/ij/) for the image analysis.
For TIRF microscopy of Fis-gfp, agarose pads were prepared by casting agarose molten in low-autofluorescence medium (minimal media with glucose and 25ng aTc) between two glass slides spaced by two cover slips. A 2x2mm square was cut out, placed onto a glass slide, and 100ul of bacterial culture was spotted onto it and allowed to dry. The pad was framed by a double-sided sticky 9x9mm frame seal (Bio-Rad SLF0201) and covered with a high-precision and clean room grade cover slip (Schott Nexterion glass D, thickness 0.170+/- 0.005mm). Imaging was done at 37°C with a temperature-controlled Olympus IX83 total internal reflection fluorescence microscope equipped with a water-cooled Hamamatsu ImageEM C9100-13 camera, a 100x 1.49NA objective lens and an additional 2x magnification tubular lens giving an effective pixel size of

80nm/pixel. Images were acquired every 3min. using a single-band pass GFP filter and a diode 488nm laser set to an output of 0.4mW and low camera gain settings. Fis-gfp was imaged at 65nm penetration depth.

ChIP sequencing

To perform ChIP sequencing experiments, λ CI was cloned with an HA-Tag at their carboxy-terminal end and transformed into both host strains. Samples were prepared according to Waldminghaus & Skarstad (2010)[224]; library preparation and Illumina Sequencing was performed at the VBCF NGS Unit (www.vbcf.ac.at). The obtained data was analyzed using Galaxy and RStudio, considering reads of non-overlapping 1000bp windows.

Binding footprints for genic and intergenic regions were obtained by scaling them and calculating binding for each percentile over all genic and intergenic regions annotated in *E. coli* and *S. enterica*. The analysis was performed for strains that expressed λ CI as well as a background control. In a further control reads in genic regions were randomly shuffled, giving a completely different footprint.

The number of reads in each bin was compared with the binding strength calculated from the λ CI energy matrix[97] by summing up binding affinities over the same 1000bp windows across the genomes if the affinities above a certain threshold. The threshold was optimized for giving the best Spearman correlation between the number of reads and the predicted binding strength. Binding strength was calculated using $1/1+\exp(E-\mu)$, with μ being the chemical potential and E the binding energy.

Statistical tests

Collected data was tested for normality (Shapiro-test) and subsequently we compared mean OD_{600} or fluorescence expression values using t-tests with FDR correction for multiple comparisons in RStudio. T-tests were performed for four different time points between samples grown in the presence and absence of inducer aTc (presence or absence of repressors) under indicated conditions.

Spearman correlation was calculated for the fit between model predictions of binding strength and the number of obtained ChIP reads per 1000bp window.

5.7 Author contributions

C.I. (Claudia Igler), T.F. (Tamar Friedlander), C.C.G. (Călin C. Guet) conceived the study together. C.I. designed and carried out the experiments and analyzed the data. C.F. (Claire Fourcade) helped with cloning, T.B. (Tobias Bergmiller), R.C. (Remy Chait) and K.J. (Kirti Jain) performed fluorescence microscopy, T.W. (Torsten Waldminghaus) guided ChIPseq experiments and F.P. (Florian Pauler) ChiPseq data analysis. C.I. wrote the current draft of the manuscript.

6. TF interference produces transient promoter memory in response to signal timing

6.1 Abstract

Cells usually integrate over multiple input signals at individual promoters, which governs their phenotype in changing environments[60,121,147,148]. However, different signals generally do not arrive at a promoter at exactly the same time[149], possibly making gene expression dependent on the signaling history. By measuring fluorescence dynamics from a promoter that is controlled by two transcriptional regulators, we show that even small differences in signal arrival timing can lead to transient regulatory memory over several generations. The longevity of these transient patterns largely result from generic transcription factor interactions - namely interference with each other's binding - which constitutes an abundant feature in multiple signal integration. Furthermore, we show that after environmental change transient memory can influence growth rates significantly. Transient promoter dynamics represents a generic form of memory, which produces phenotypic diversity in clonal cells.

6.2 Introduction

Integration of multiple signals is a key feature of regulation at many promoters[147] and differences in signal arrival time occur commonly in any given environment[149]. Cellular phenotypes can be significantly influenced by such signaling delays[149], making them dependent on the signaling history. History-dependence, or memory, has generally been associated with specific mechanisms, such as feedback loops or epigenetic states[150], which lead to multi-stability[151]. However, little attention has been paid to regulatory mechanisms per se as a source of (transient) memory[152,153]. Only recently, it has been proposed that the kinetics of transcription factor (TF) binding[154] and transient regulatory processes[8,59,60] can shape cellular phenotypes. Here, we report that microbes can employ transient gene regulatory dynamics to realize a generic form of memory. We use a promoter that is controlled by two external signals and has a single equilibrium state to determine transient promoter dynamics in response to varying signaling histories: two TFs are either induced with a time delay between them, or both are induced concomitantly (Fig. 6.1A).

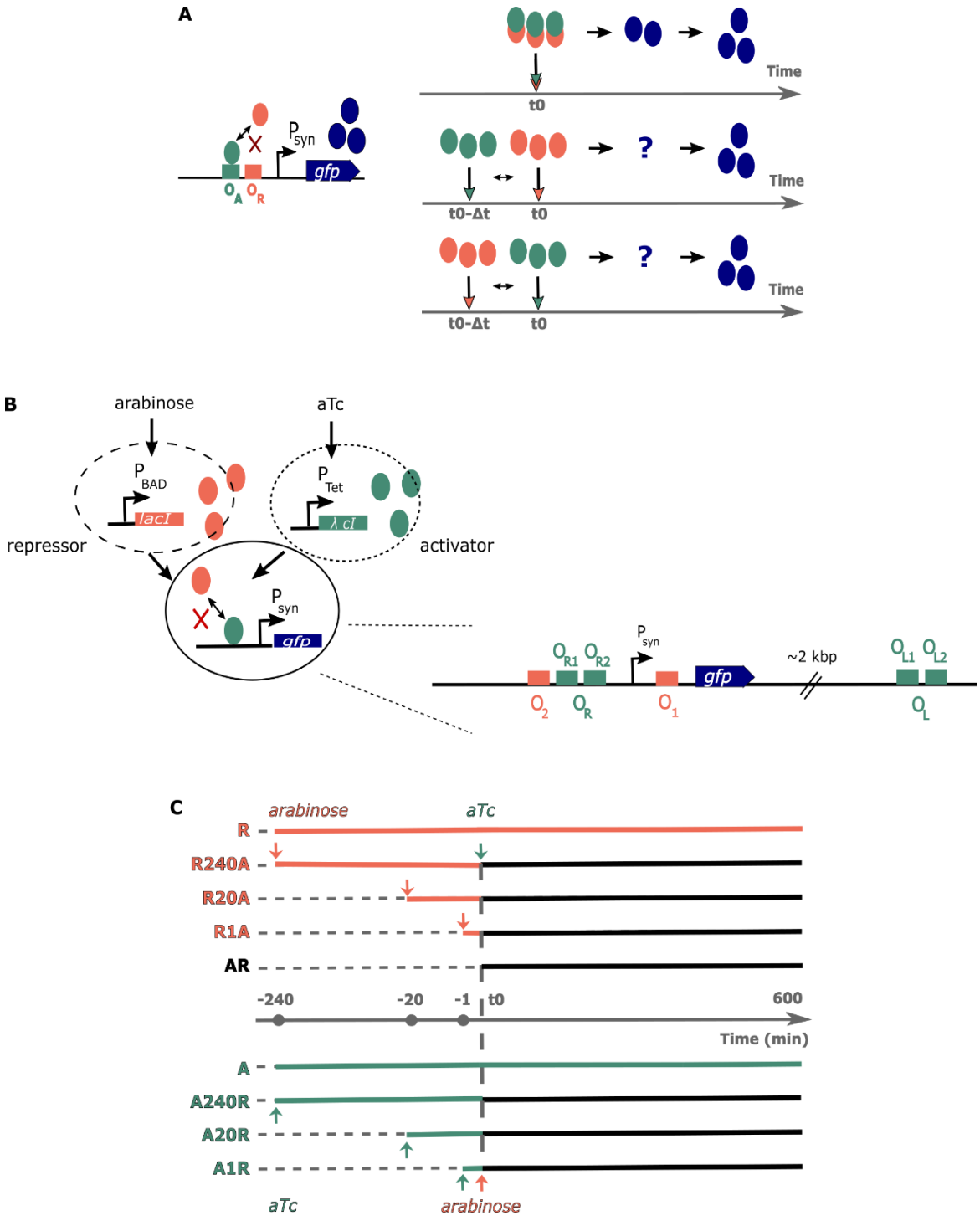

Figure 6.1. The experimental system and induction scheme.

A) *Expression of the fluorescent marker gene (gfp) is controlled by P_{syn} through competitive binding of an activator (λ CI, green) and a repressor (LacI, orange). Induction of both TFs at t0 leads to transient expression (two ovals) before reaching steady state (three ovals). A time delay Δt between the TFs produces unknown transient expression before reaching steady state.* ***B)*** *Circles indicate locations of the system components: on the chromosome (dashed line), on a low or medium copy number plasmid (dotted and solid line respectively). Arrows show promoters and colors show genes and their encoded TFs. P_{syn} consists of the region directly upstream of the fluorescence marker and two additional λ CI operators 2.3kbp away. λ CI operators are*

green (O_{R1} & O_{R2} and O_{L1} & O_{L2}) and LacI operators orange (O_1 and O_2). ***C)*** *Timeline of the induction procedure (grey arrow): gray dots as well as the green and orange arrows (activator and repressor respectively) give the induction time points of the first (240, 20 and 1min) and second inducer (vertical dashed line shows t0). Grey, dashed lines indicate no induction, green lines activation only, orange lines repression only and black lines concomitant activation and repression.*

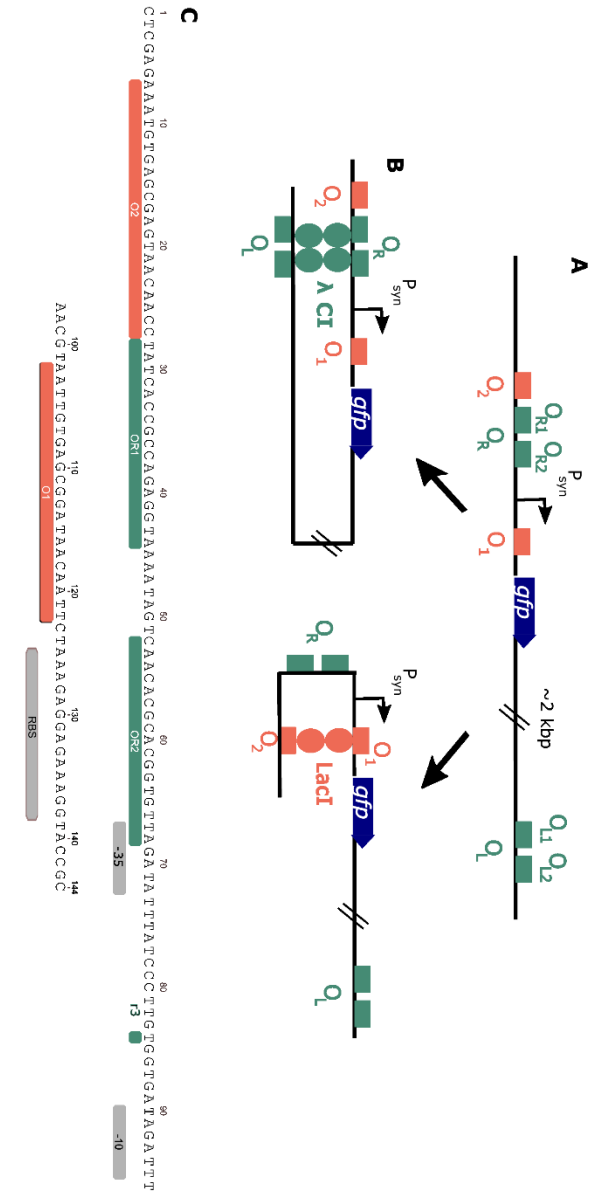

Figure 6.2. The experimental system.

A) *P_{syn} consists of the region directly upstream of the fluorescence marker and two additional λ CI operators 2.3kbp away[121]. λ CI operators are green (O_{R1} & O_{R2} and O_{L1} & O_{L2}) and LacI operators orange (O_1 and O_2).* ***B)*** *λ CI binds O_R (and O_L) through*

cooperative binding of two dimers (green ovals). Cooperative binding of four λ CI dimers results in a long-distance DNA loop[156] (left). LacI forms tetramers, where each dimer subunit (orange ovals) binds one operator, forming a short (~94bp) DNA loop[10] (right). ***C)*** *DNA sequence of the P_{syn} region. The region between the beginning of O_{R1} and the beginning of O_1 (28 – 101bp) was taken from the wildtype P_{RM}, except for the introduction of a mutation (r3) that destroys the O_{R3} operator (to avoid repression by CI). LacI operator O_1 was introduced at a distance from -10 as in the wildtype P_{lacZ} promoter and at a distance to O_2 that favors looping[174].*

6.3 Results

Synthetic promoter design and experimental layout

As it is difficult to determine promoter dynamics directly, we can instead observe if gene expression trajectories differ significantly from the expectation based on fast regulatory equilibration. Due to its high stability[155], DNA looping could be one process that slows down promoter state equilibration. To test the timescales of complex promoter regulation, we constructed a synthetic promoter (P_{syn}) that is regulated by two well-studied TFs, λ CI and LacI[10,121](Fig. 6.1B), which both form DNA loops (Fig. 6.2): λ CI forms a long-distance loop with a second operator region[156] , whereas LacI forms a short-distance loop[10]. DNA looping is not necessary for gene regulation, but increases the efficiency of either regulatory state[10,157,158] and interferes strongly with looping by the other TF[155,159]. LacI represses expression of a fluorescence marker from P_{syn} and its expression is controlled by an *arabinose*-inducible P_{BAD} promoter. The activator of P_{syn}, λ CI, was degradation-tagged due to its toxicity at high concentrations and placed under the control of an *aTc*-inducible P_{Tet} promoter (Fig. 6.1C).

We determined transient gene expression trajectories from P_{syn} for various signaling histories by measuring population fluorescence levels over 14h: Three conditions ('R->A': 'R240A', 'R20A', 'R1A') were induced with *arabinose* (LacI expression) for 240, 20 and 1 min (1min ~= synthesis time of 1-2 TF molecule(s)) before *aTc* (λ CI expression) was added at t0 (Fig. 6.1C). Likewise, three conditions ('A->R': 'A240R', 'A20R', 'A1R') were induced first with *aTc* and after the specified time delay also with arabinose (Fig. 6.1C). As controls, one sample was not induced ('basal'), one was activated ('A'), one was repressed ('R') and one was induced for both TFs concomitantly ('AR') (Fig. 6.3A,B) in order to determine expression trajectories for fast promoter equilibration in the P_{syn} system.

Transient promoter memory

By comparing fluorescence production after t0 from 'R->A' delay conditions with that of the 'AR' control we found significant memory in P_{syn} regulation for 'R240A' and 'R20A' (Fig. 6.3C, Table 6.1). Strikingly, for 'R240A' the fluorescence increase was very slow, contrasting strongly with the steep increase of 'AR'. For 'A->R' delay conditions we tested if fluorescence decay rates were longer than dilution rates; meaning that the timescales of promoter state equilibration were slower than those of dilution. We

compared regression slopes for fluorescence decay and inverse growth (dilution) over 3h after t0 and found that 'A20R' and 'A1R' decay was significantly slower than dilution (Fig. 6.3D, 6.4, Table 6.2). Generally, significant expression divergence from 'AR' was seen for 4-11h (~ 3-8 doublings) after t0 with 'A->R' conditions and for more than 11h after t0 with two 'R->A' conditions (Fig. 6.5A,B, Table 6.3). The observed memory in gene expression levels lasted for several cell generations, stemming from long transient promoter timescales.

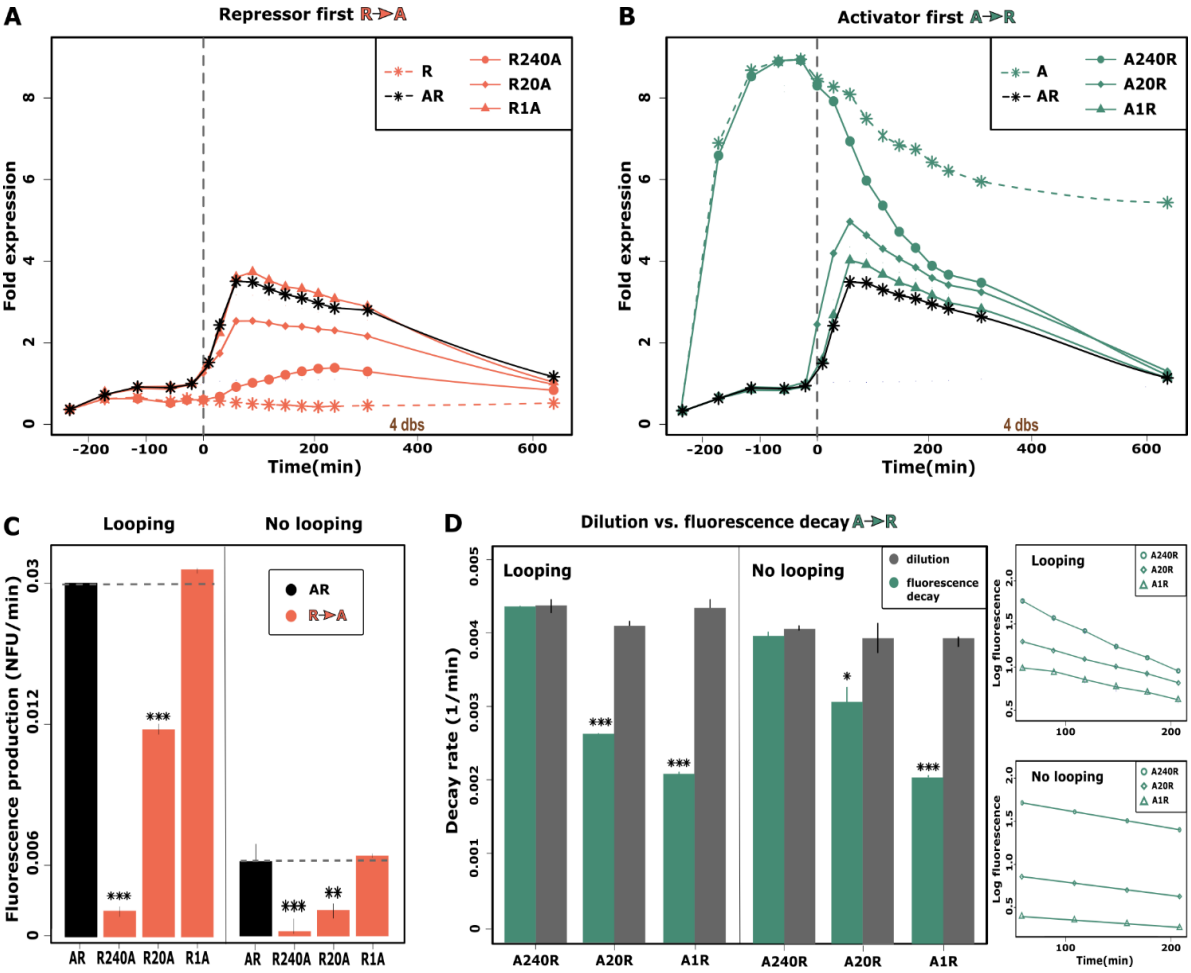

Figure 6.3. Experimentally observed transient memory in gene expression.

(A, B) *OD_{600}-normalized fluorescence units (NFU) are shown for 10h after t0 (x-axis shows time in minutes). Controls (dashed lines) and delay conditions are indicated by symbols. The second TF was induced at t0 for all cultures (dashed vertical line). 4 population doublings after t0 are indicated at 6h. Error bars are standard errors of the mean. 'AR' is shown in black, 'R->A' and 'R' conditions are orange* ***(A, C)*** *and 'A->R' and 'A' conditions are green* ***(B, D)****.* ***(C, D)*** *Memory measures for experiments in systems with and without DNA looping.* ***C)*** *Bars show fluorescence production for 'AR' (dashed line) and 'R->A' delay conditions, calculated over 1.5h after t0. Stars indicate significant difference to 'AR'.* ***D)*** *Rates for dilution (inverse growth) and observed fluorescence decay for 'A->R' delay conditions were determined through linear regression (from t0.5 to t3) of log growth and log fluorescence (shown on the right).*

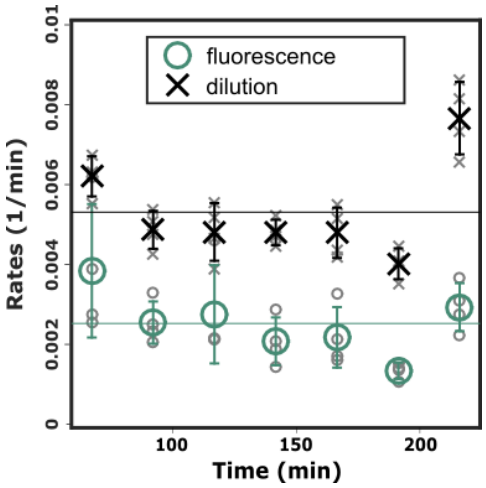

Figure 6.4. Mean decay and dilution rates for 'A240R', 'A20R', 'A5R', 'A1R'at 7 time points after t0.

Time is given in minutes after t0. Grey symbols give the individual dilution (crosses) and fluorescence decay (circles) rates for 'A->R' conditions. Mean decay rates were calculated from experimental fluorescence measurements (big green circles) and mean dilution rates were calculated from experimental OD measurements as the inverse of the growth rate (big black crosses). Error bars give standard deviations. The green and black lines give the mean decay and dilution rates respectively.

Simultaneously with plate-reader population measurements we performed FACS analysis of single cell fluorescence, but found no appreciable bimodality in the fluorescence distributions (Fig. 6.6). Hence, our observations did not result from subpopulations that were either fully activated or fully repressed in varying proportions, but from transient regulatory dynamics within individual cells. FACS and population measurements showed a fluorescence drop in the 'A' control to a lower, yet unmistakably activated, level (Fig. 6.3B). As fluorescence remained constant thereafter and the drop rate does not align with 'A->R' or 'R->A' fluorescence decay or dilution rates, the observed transient memory effects remain valid.

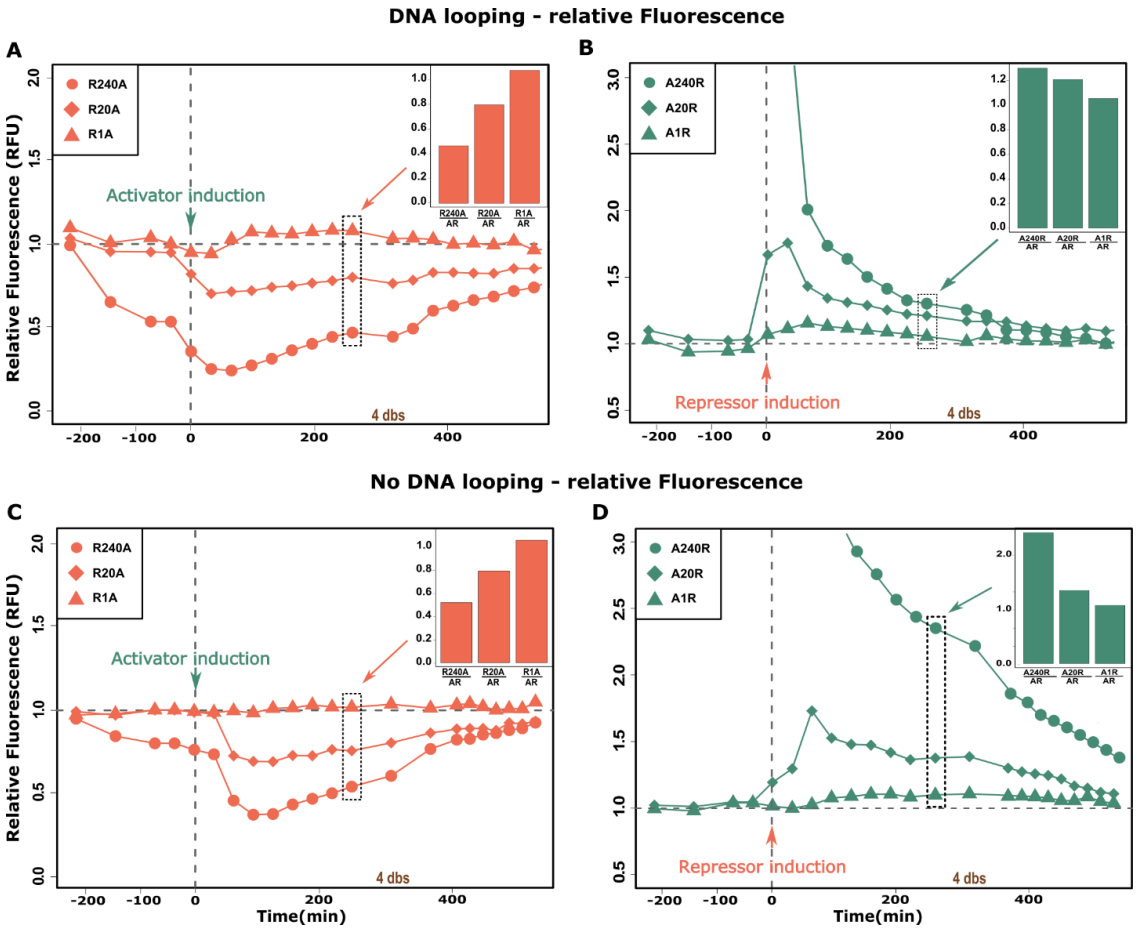

Figure 6.5. Relative fluorescence trajectories of experimentally observed transient memory.

The x-axis shows time in minutes and the y-axis relative OD_{600}-normalized fluorescence with respect to 'AR'. Delay conditions are indicated by symbols. ***A), C)*** *'R->A' and 'R' conditions are orange and* ***B), D)*** *'A->R' and 'A' conditions are green. The second TF was induced at t0 for all conditions (dashed vertical line). 4 population doublings after t0 are indicated at 6h. 'AR'-relative fluorescence values are shown for experiments* ***A), B)*** *with DNA looping,* ***C), D)*** *without DNA looping. The horizontal dashed line at 1 indicates conditions with fluorescence equal to 'AR'. Bar plots show the relative fluorescence values at t4.*

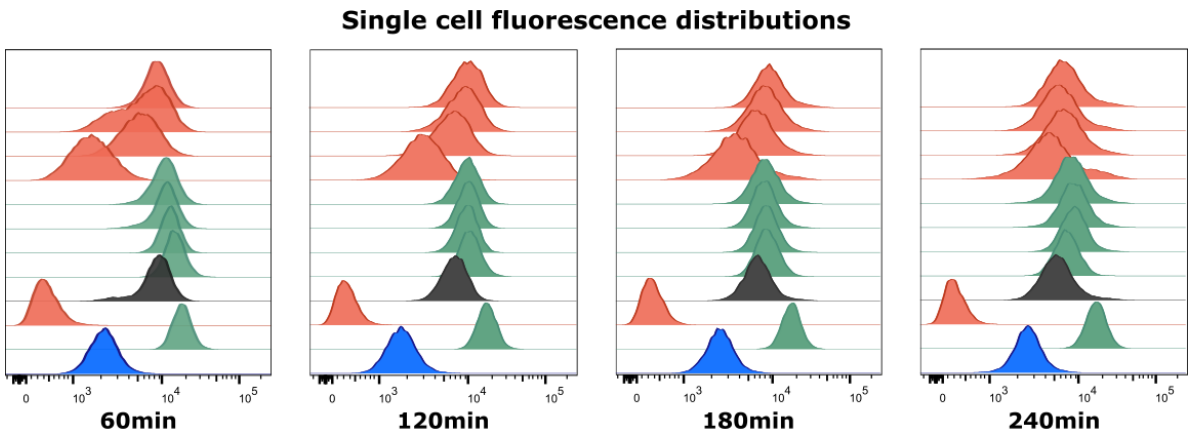

Figure 6.6. FACS data for transient memory at P_{syn} in the experiments with DNA looping.

Distributions of fluorescence values over 4h after t0 are shown. The order of conditions from bottom to top is as follows: 'basal', 'A', 'R', 'AR', 'A240R', 'A20R', 'A5R', 'A1R', 'R240A', 'R20A', 'R5A', 'R1A'. The conditions show no bimodality, although the coefficient of variation is higher for several 'A5R', 'A1R', and 'R->A' conditions after induction of the second TF. FACS measurements show no bimodality and especially for 'A->R' conditions the variation in gene expression distributions already resembles that of the 'A' control after one hour. 'R->A' conditions show a broader distribution that in some cases overlaps 'AR' but the mean expression differs significantly and reproducibly from AR (except for 'R5A').

Reaction rate model of transient dynamics

In order to gain a better understanding of the promoter dynamics causing the transient expression patterns, we built a reaction rate model, which calculates gene expression levels from the probabilities of promoter states (i.e. binding configurations of TFs) and the transitions between states, determined by TF binding affinities[51,57,160] (Fig. 6.7). Model parameters were fitted to experimental data, but were kept within physiological ranges (Table 6.4). Due to the transient promoter dynamics observed experimentally, we did not use equilibrium transition rates, i.e. we modeled binding and unbinding reactions individually. Further, several studies demonstrated interference between TFs, such as binding in close proximity (steric binding hindrance)[12,161] and concentration-dependent TF unbinding rates (facilitated dissociation)[14,136]. Hence, we introduced 'interference' states (Methods), in which both TFs can bind at the promoter at the same time. In the full model we assume that binding rates in those states are decreased (due to steric occlusion (SO) between TFs), whereas unbinding rates are increased proportionally to the concentration of the competing TF (facilitated dissociation (FD)). This model resulted in good agreement between simulations and experimental data (Fig. 6.8, Table 6.5). Our model proposes a mechanism for the non-equilibrium effects at P_{syn}: TF interference and the production of TFs results in slow timescales, which are not well seperated from dilution (production) rates (Table 6.4), and contrast with the greatly increased unbinding rates due to facilitated dissociation.

We compared this 'interference' model to a thermodynamic model, which is the most commonly used reaction rate model for gene expression but assumes equilibrium in

the system[57,162,163], and a non-equilibrium model that does not include 'interference' states. Both models produced a worse fit to our data (Fig. 6.8, 6.9, Table 6.5).

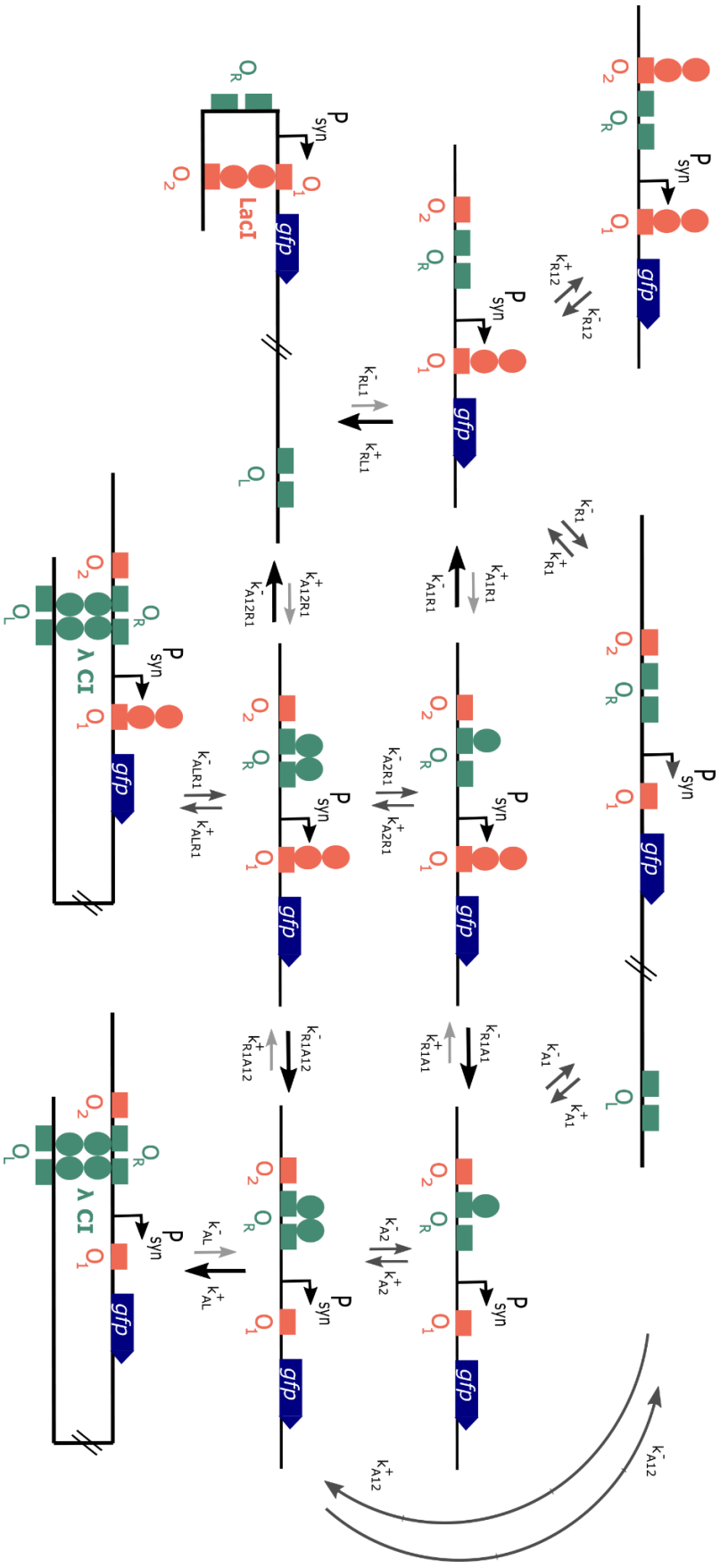

Figure 6.7. Regulatory states (binding configurations) of the non-equilibrium model with 'interference'.

λ CI operators (O_{R1} & O_{R2} and O_{L1} & O_{L2}) and λ CI dimers are shown in green, whereas LacI operators (O_1 and O_2) and LacI tetramers (each oval represents a dimer) are shown in orange. Arrows show transitions between individual regulatory states, with the size and coloring of the arrow indicating the speed of the reaction (big and dark arrows = fast reactions). For simplicity, repressor binding to one operator is only shown for O_1, but the same states and transitions were included for O_2. Rate names correspond to Table 6.4.

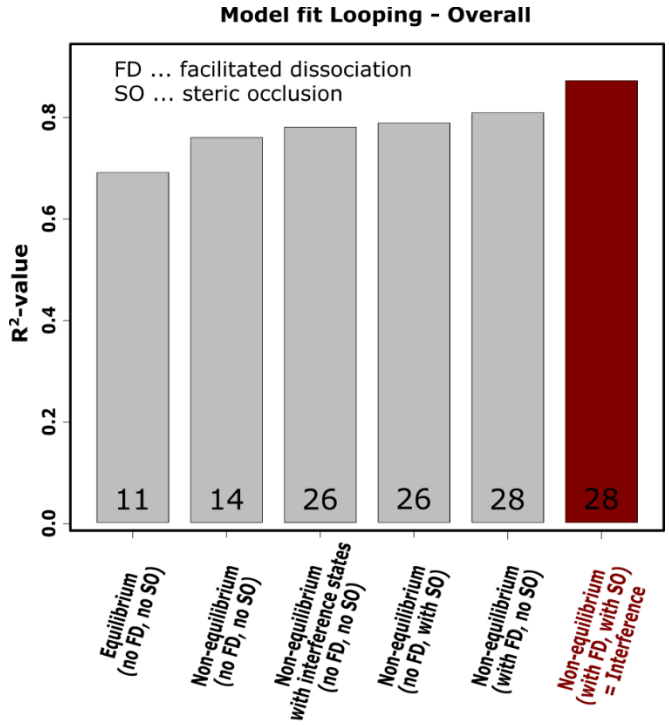

Figure 6.8. Model fit with experimental data.

Barplots show the model fit with experimental data for different model versions: equilibrium model, non-equilibrium without interference states (both TF bound), non-equilibrium model with interference states without FD (facilitated dissociation) or SO (steric occlusion), with either of them or the full model containing both (interference model). R^2-values were calculated as described in Methods. Numbers at the bottom of the bars give the number of fitted parameters per model.

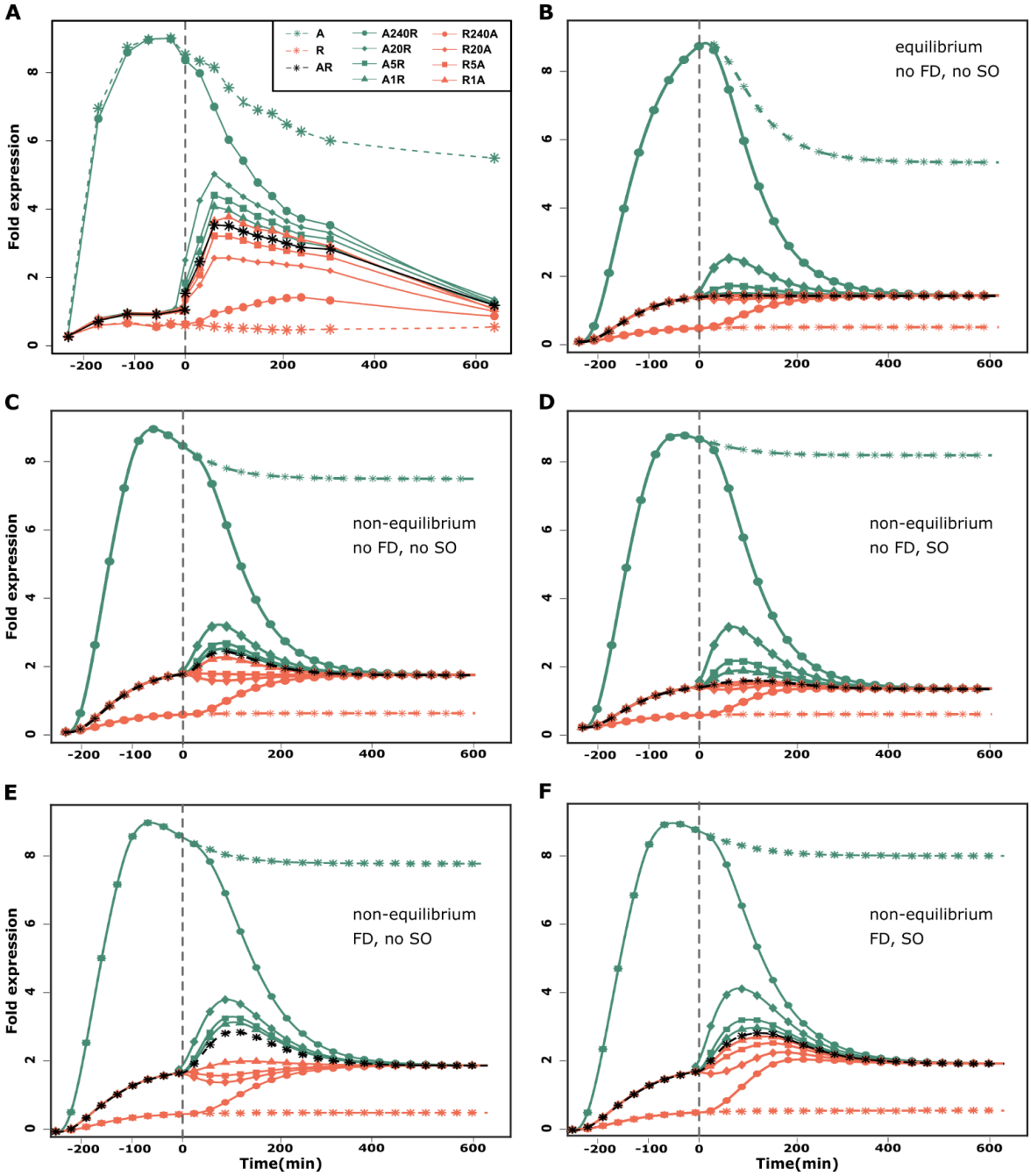

Figure 6.9. Qualitative fit for different model versions with experimental data.

The x-axis shows time in minutes and the y-axis fluorescence either in OD_{600}-normalized fluorescence units relative to the P_{syn} basal expression level (fold expression) (experiment) or in arbitrary units (AU) (simulations). 'AR' is shown in black, 'A->R' and 'A' conditions are green and 'R->A' and 'R' conditions orange. Control conditions (dashed lines, star symbols) and delay conditions are indicated by symbols (circles, diamonds, squares, triangles for 240, 20, 5, 1min delays respectively). The second TF was induced at t0 for all conditions (dashed vertical line). ***A)*** *Experimental observations of transient memory. Error bars are standard errors of the mean. Model simulations of fluorescence dynamics with the* ***B)*** *equilibrium model,* ***C)*** *non-equilibrium model,* ***D)*** *non-equilibrium model with SO (steric occlusion,*

E) *non-equilibrium model with FD (facilitated dissociation), and the* ***F)*** *non-equilibrium model with SO and FD (interference model). Especially the fluorescence patterns of 'R->A' conditions after t0 (e.g. the initial peak in fluorescence) are only captured by the simulations in* ***F)****. Interference states increase the duration of transient dynamics compared to the equilibrium model but decrease the duration of states that could be assumed to be highly stable (e.g. DNA loops).*

Growth effects of transient memory

Transient memory, as observed in our system, can lead to variation in population growth after environmental changes[164,165]. We studied growth effects due to signaling delays by replacing the *gfp* marker in the P_{syn} system with the tetracycline efflux pump, *tetA,* which determines growth rate in the presence of the antibiotic tetracycline (Fig. 6.10). The described induction scheme remained the same, but tetracycline was added together with the second inducer at t0. Four delay conditions showed significantly different growth curves and rates than 'AR' (Fig. 6.11A,B, Table 6.6, 6.7). Due to toxicity of high *tetA* expression in the presence of tetracycline[166], growth is not proportional to higher *tetA* levels but a trade-off between the toxicity cost and the benefit of pumping out tetracycline (Fig. 6.12). Nonetheless, long-term population growth was significantly influenced by transient promoter memory.

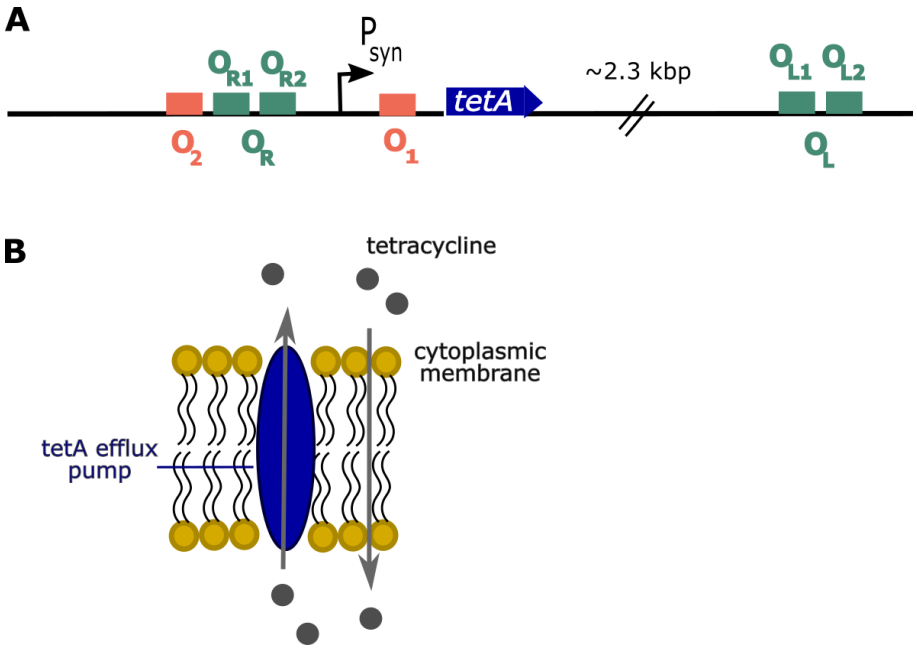

Figure 6.10. System for measuring growth differences due to the expression of tetracycline resistance.

A) *P_{syn} with tetA instead of gfp.* ***B)*** *TetA encodes an efflux pump that is incorporated into the cytoplasmic membrane and confers resistance by pumping tetracycline out of the cell[166].*

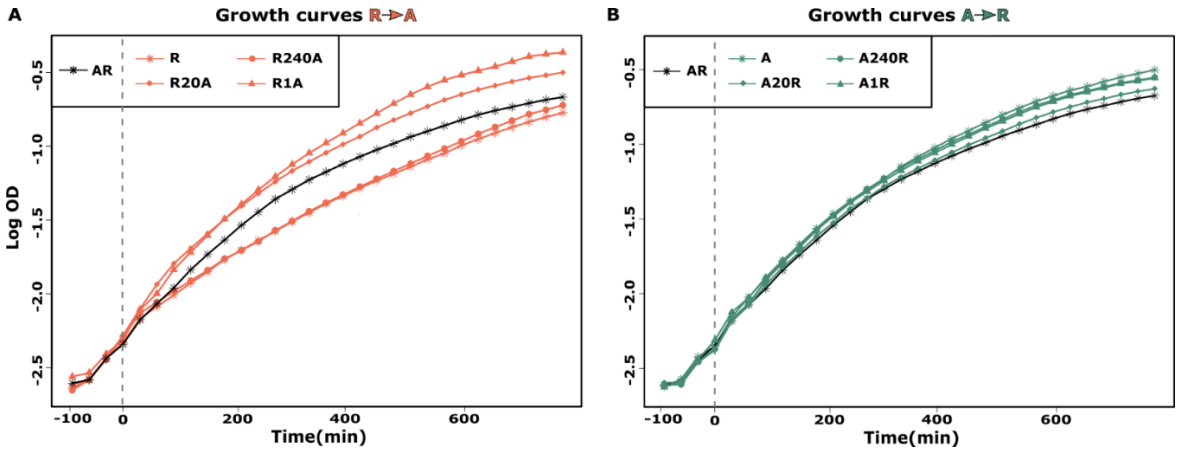

Figure 6.11. Experimental growth curves in the presence of tetracycline with tetA expressed from P_{syn}.

The x-axis shows time in minutes and the y-axis shows growth measurements (log OD_{600}). 'AR' control conditions are black and delay conditions are indicated by symbols. ***A)*** *'R->A' and 'R' conditions orange,* ***B)*** *'A->R' and 'A' conditions are green. The second TF was added together with tetracycline at t0 for all cultures (dashed vertical line). All conditions, except 'A20R', are significantly different from 'AR' for most measurement points after t4 (Table 6.6).*

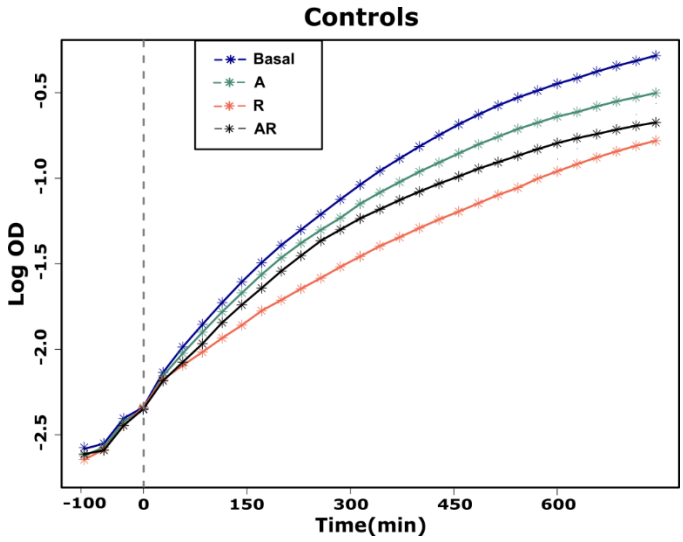

Figure 6.12. Growth curves of control conditions in the presence of tetracycline for P_{syn} controlling *tetA*.

'Basal' is shown in blue, 'AR' in black, 'A' in green and 'R' in orange. Error bars show variation in growth rates. Cells were grown in triplicates in 96 well plates in M9 with 0.5% CAA on a shaker at 30°C. The second TF was induced at the same time as tetracycline (5ug/mL) was added to the media, which is indicated by the dashed vertical line. The increased growth of 'basal' over 'A' and 'AR' conditions indicates a trade-off between the cost of tetA overproduction in the presence of tetracycline and the expression of enough efflux pumps to confer resistance to the antibiotic.

Transcription factor interference as basis of transient memory

The memory of the initial regulatory state at P_{syn} could result from stability conferred by DNA looping. We therefore tested the generality of transient memory by removing the operators that enable looping by either TF (Fig. 6.13). Indeed, even without DNA looping, almost all delay scenarios showed transient promoter dynamics (Fig. 6.3C,D, Table 6.1, 6.2) and fluorescence patterns that differed significantly from 'AR' (Fig. 6.5C,D, Table 6.8) for 5-9h. By adjusting the regulatory states in our non-equilibrium model with 'interference' accordingly, we obtained a satisfactory fit between simulations and experiments (Fig. 6.14, Table 6.9).

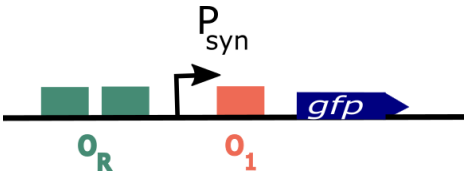

Figure 6.13. P_{syn} layout without operators for looping.

λ CI operators are shown in dark orange (O_{R1} & O_{R2}) and the LacI operator is shown in green (O_1). Operators involved in DNA looping by either TF were removed.

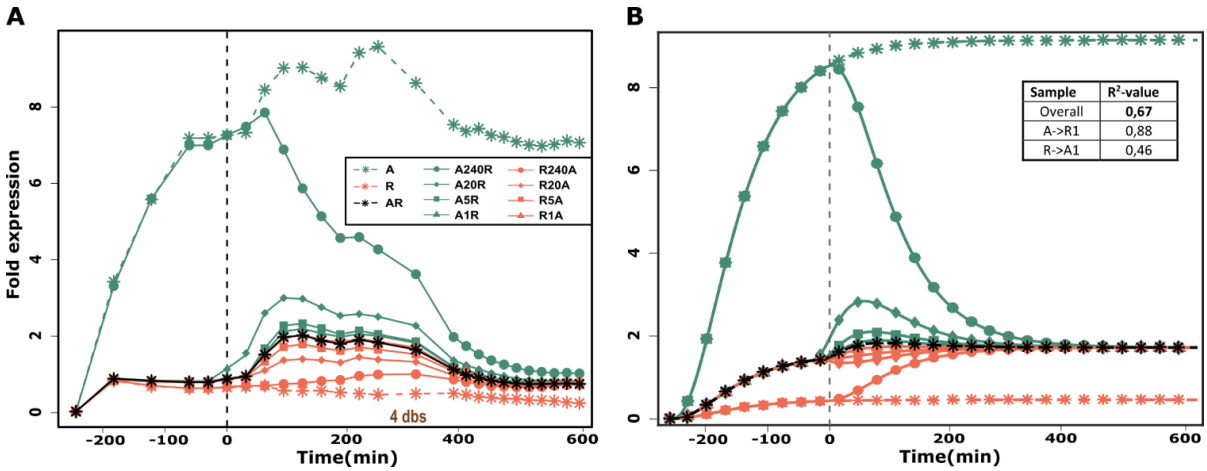

Figure 6.14. Transient memory in gene expression at P_{syn} without DNA looping for time delays of 240, 20, 5 and 1min.

The x-axis shows time in minutes and the y-axis fluorescence either in OD_{600}-normalized fluorescence units (NFU) (experiment) or in arbitrary units (AU) (simulations). ***A)*** *Experimental observations of transient memory at P_{syn} without DNA looping. Colors and symbols correspond to Fig. 6.8A with green curves 'A->R' and orange curves 'R->A' conditions. Notably, there is no drop in activation in a system without looping.* ***B)*** *Model simulations of transient memory at P_{syn} without DNA looping. Colors correspond to the ones used in A). The simulations produced the best qualitative and quantitative fit for the system without looping, indicating the increasing accuracy of modelling transients with decreasing system complexity.*

As looping is not an essential feature, transient memory could be conferred by slow production of TFs in response to external signals or by direct binding interference between TFs (Fig. 6.2). Model predictions show that the interference between RNA

polymerase binding (which in our model is equivalent to activator binding) and repressor binding leads to transient memory - which is abolished if repressor binding is weakened (Fig. 6.15). We tested this prediction by introducing a single mutation in the repressor operator that weakens the binding affinity[167]. We measured the memory effect by comparing fluorescence production rates after t0 and found no significant difference between the delay conditions and the 'AR' control (Fig. 6.15, Table 6.1). This confirms the model prediction of memory loss due to decreased binding competition between RNA polymerase and the repressor. Moreover, either of the three models (thermodynamic, non-equilibrium, interference) gave an equally good fit to experimental data (Table 6.9). As the rate of TF production should remain the same as in previous experiments, promoter dynamics arise to a significant extent from interference between TFs, thereby providing a generic form of transient memory at complex promoters.

6.4 Discussion

Bacterial physiological adaptation and decision making is determined by the ability to display history-dependent phenotypes[164,168–170]. We show that phenotypic memory can be caused by delays in signal arrival times as even delays of one minute can result in transient promoter dynamics, and thus different gene expression trajectories, for several cell generations. Transient memory can significantly influence adaptation: either favoring it by sampling various expression states[171], or constraining it due to long response times[172]. The inherent dynamics of simple TF-TF and TF-DNA interactions at a single promoter are sufficient to produce considerable phenotypic variation that does not originate from stochasticity, but from DNA allostery[173]. This surprising complexity in dynamics for a system with a single stable equilibrium leads to challenges for formal descriptions of gene regulation: even though we used well-described TFs, a detailed model containing complex interaction and interference mechanisms was necessary to obtain a good fit to experimental data. TF interference likely is a common feature at many *E. coli* promoters as they frequently contain closely spaced operators for more than one TF (almost 50% of all known operators overlap another operator)[13]. Moreover, our findings challenge our understanding of even well-studied regulatory mechanisms like DNA looping or TF binding. The architecture of complex promoters can be a rich source of variation in gene expression[154], enabling diverse transient promoter dynamics in clonal cells that manifest over time-scales of several generations.

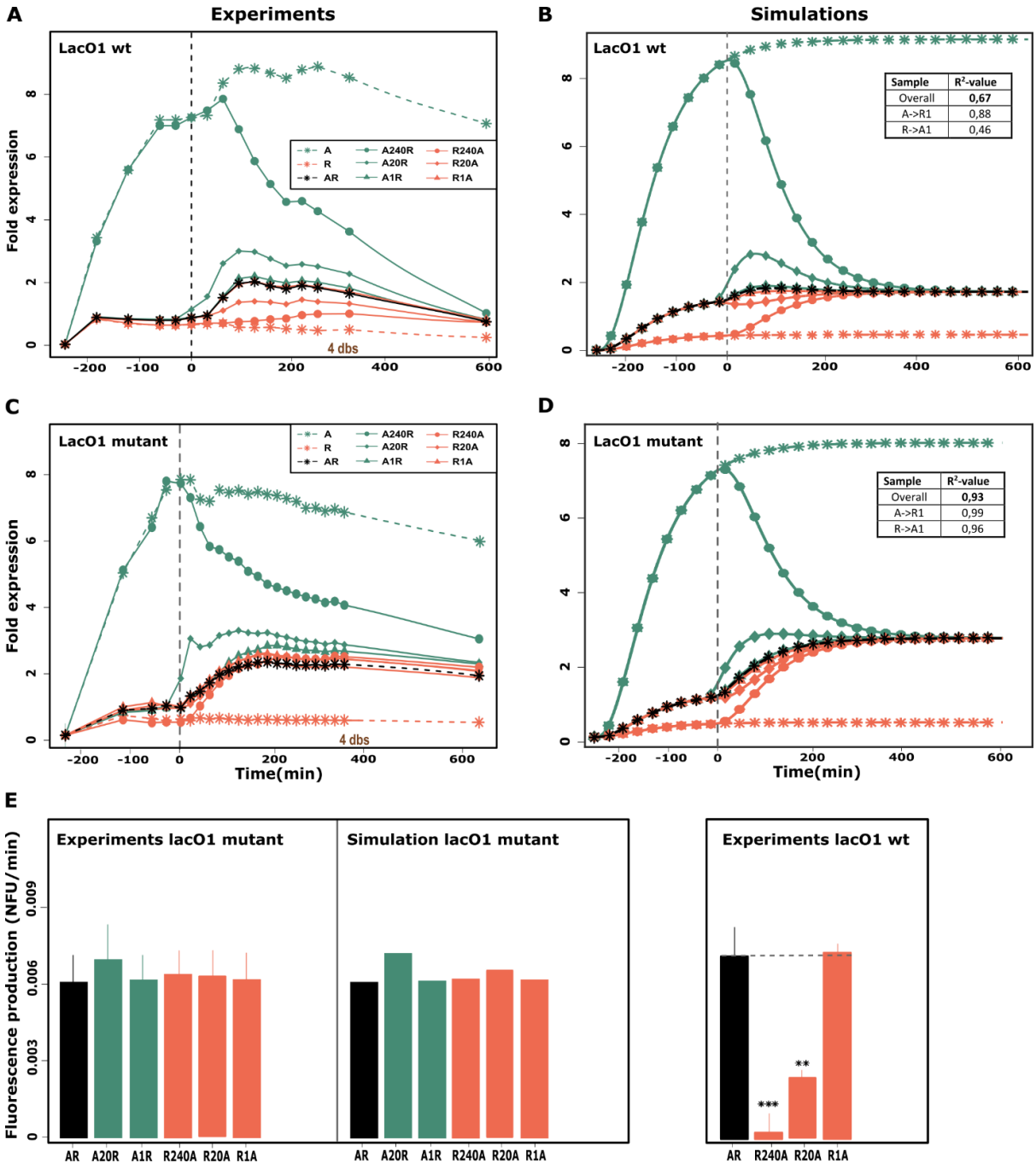

Figure 6.15. Transient memory in gene expression at P_{syn} without DNA looping disappears only for a $lacO_1$ mutant.

(A-D) *The x-axis shows time in minutes and the y-axis fluorescence either in OD_{600}-normalized fluorescence units relative to basal expression (fold expression) (experiment) or in arbitrary units (AU) (simulations).* ***(A, C)*** *Experimental observations of transient memory at P_{syn} without DNA looping with wildtype $lacIO_1$* ***(A)*** *or mutated $lacIO_1$* ***(C)****. Colors and symbols correspond to Fig. 2 with green curves 'A->R' and orange curves 'R->A' conditions. Notably, there is no drop in activation in systems without looping.* ***(B, D)*** *Model simulations of transient memory at P_{syn} without DNA looping with wildtype $lacIO_1$* ***(B)*** *or mutated $lacIO_1$* ***(D)****. Colors correspond to the ones used in A) and C).* ***E)*** *Bars show fluorescence production for 'AR' (dashed line), 2 'A->R' and 3 'R->A' conditions, which was calculated over 1.5h after t0 relative to the starting*

concentration at t0. Stars indicate significant difference to 'AR'. Experiments and simulations with the lacO_1 mutant (left) are compared to the experiments with the lacO_1 wildtype (right).

6.5 Tables

Table 6.1. FDR adjusted p-values comparing fluorescence increase after t0 for the 'AR' control, 'R->A' and (for the lacO_1 mutant) 'A->R' conditions.

The slope of the fluorescence increase was calculated from t0 over three time points (t0 to t1.5) and t-test were performed between 'AR' and 'R->A' conditions as well as 'AR' and 'A->R' conditions for the lacO_1 mutant without looping. P-values that are above the significance level α = 0.05 are shown in red.

	Fluorescence increase in AR vs. R->A			**Fluorescence increase in AR vs. A->R**	
Experiment	**'R240A'**	**'R20A'**	**'R1A'**	**'A20R'**	**'A1R'**
Looping	p=7,47 E-13; f=20,58	p=1,95 E-04; f=5,05	p=0,10; f=-1,60	/	/
No looping	p=3,5 E-04; f=11,45	p=1,8 E-03; f=7,02	p=0,99; f=0,02	/	/
No looping with lacO_1 mutant	p=0,99; f=0,95	p=0,99; f=0,99	p=0,99; f=0,96	p=0,99; f=0,65	p=0,99; f=0,96
Low concen-tration of TFs	p=0,005; f=6,61	p=0,75; f=1,11	p=0,84; f=-0,21	/	/

Table 6.2. FDR adjusted p-values comparing 'A->R' fluorescence decay against dilution over 3h after t0.

T-tests were performed for regression coefficients that were obtained from fluorescence decrease due to dilution or due to the decay observed in experiments. Decay curves were obtained by fitting an exponential function through experimental data points and correcting for the approximate equilibrium level. Dilution curves were calculated as exponential function that would be expected due to experimental growth rates. P-values that are above the significance level α = 0.05 are shown in red.

	Fluorescence dilution vs. decay A->R		
Experiment	**'A240R'**	**'A20R'**	**'A1R'**
Looping	p=0,96; f=0,05	p=2,4 E-05; f=-5,85	p=2,28 E-06; f=-7,47
No looping	p=0,08; f=0,06	p=0,03; f=-5,60	p=0,004; f=-4,37
Low concentration of TFs	p=0,001; f=6,39	p=0,001; f=-8,25	p=0,001; f=-7,66

Table 6.3. FDR adjusted p-values comparing 'A->R' or 'R->A' fluorescence values against the 'AR' control at several time points in experiments with looping.

T-tests were performed at different time points after t0 between 'A->R' or 'R->A' fluorescence and 'AR' fluorescence. Values that are above the significance level α = 0.05 are shown in red.

Time (min)	'A240R'	'A20R'	'A1R'	'R240A'	'R20A'	'R1A'
60	p=3,91E-14; f=-28,64	p=1,29E-06; f=-8,24	p=0,005; f=-3,56	p=9,08E-09; f=23,79	p=0,001; f=5,30	p=0,129; f=-1,78
120	p=1,348E-12; f=-21,58	p=1,692E-06; f=-8,02	p=0,012; f=-3,15	p=1,92E-08; f=19,51	p=0,020; f=2,98	p=0,028; f=-2,77
180	p=1,44E-11; f=-18,39	p=7,79E-07; f=-8,60	p=0,025; f=-2,74	p=1,45E-07; f=15,10	p=0,018; f=3,06	p=0,010; f=-3,47
240	p=2,04E-10; f=-15,33	p=1,22E-06; f=-8,30	p=0,078; f=-2,07	p=1,45E-07; f=14,95	p=0,028; f=2,78	p=0,008; f=-3,65
330	p=0,037; f=-2,54	p=0,000458; f=-4,91	p=0,072; f=-2,12	p=0,001; f=5,44	p=0,043; f=2,51	p=0,283; f=-1,22
390	p=0,072; f=-2,11	p=0,000969; f=-4,49	p=0,131; f=-1,77	p=0,001; f=4,86	p=0,040; f=2,55	p=0,121; f=-1,84
510	p=0,718; f=-0,42	p=0,004; f=-3,76	p=0,060; f=-2,25	p=0,001; f=5,52	p=0,005; f=3,90	p=0,458; f=-0,83
640	p=0,803; f=0,29	p=0,026; f=-2,73	p=0,057; f=-2,29	p=0,005; f=4,00	p=0,006; f=3,83	p=0,404; f=0,94

Table 6.4. Parameter values (and their units) used in the reaction rate model.

Parameters were either selected from literature or fit to the experimental data, but kept within boundaries found in literature (see Methods). The last column gives the rate label as used in Fig. 6.7. Parameters colored in green were removed in the 'no looping' model and repressor off rates for the *lacO*$_1$ mutant in orange.

Parameter	**Value**	**Unit**	**Label**
scaling constant that combines cell volume & Avogadro number	1E+09		
protein degradation rate	0,003	s^{-1}	
GFP mRNA degradation rate	0,000375	s^{-1}	
GFP protein degradation rate	0,00003	s^{-1}	
GFP maturation rate	0,005	s^{-1}	
GFP translation rate	0,0133	s^{-1}	
dilution rate	0,000144	s^{-1}	
basal promoter strength	0,0068	s^{-1}	

activated promoter strength	0,075	s^{-1}	
promoter strength for repressor bound at O1 and activator bound at OR1	0,01	s^{-1}	
repressed promoter strength	0,003	s^{-1}	
LacI copy number (chromosome)	3	molec	
CI copy number (low-copy plasmid)	5	molec	
Pswitch copy number (medium-copy plasmid)	20	molec	
PBAD promoter strength	0,0358	s^{-1}	
Ptet promoter strength	0,0387	s^{-1}	
araC degradation rate	0,0002	s^{-1}	
araC molecule number	50	molec	
tetR molecule number	50	molec	
araC dissociation constant	0,002	s^{-1}	
tetR dissociation constant	0,0018	s^{-1}	
CI dimerization forward	0,6	s^{-1}	
CI dimerization reverse	5	s^{-1}	
CI degradation rate by ClpX	0,005	s^{-1}	
CI binding rate OR1	0,18	s^{-1}	k_{A1}^{+}
CI unbinding rate OR1	0,5	s^{-1}	k_{A1}^{-}
CI binding rate for two dimers with cooperativity	0,2	s^{-1}	k_{A12}^{+}
CI unbinding rate for two dimers	0,3	s^{-1}	k_{A12}^{-}
CI binding rate for a second dimer at OR2 with cooperativity	0,2	s^{-1}	k_{A2}^{+}
CI unbinding rate for a second dimer at OR2	0,3	s^{-1}	k_{A2}^{-}
CI cooperativity factor	2.2		
CI cooperativity factor for LacI bound as well	2		
LacI tetramerization forward	0,5	s^{-1}	
LacI tetramerization reverse	5	s^{-1}	
LacI dimerization forward	0,6	s^{-1}	
LacI dimerization reverse	2	s^{-1}	
LacI binding rate O1	0,4	s^{-1}	k_{R1}^{+}
LacI unbinding rate O1	0,1 / 0,3	s^{-1}	k_{R1}^{-}

LacI binding rate O2	0,6	s^{-1}	k_{R2}^{+}
LacI unbinding rate O2	0,4	s^{-1}	k_{R2}^{-}
Looping parameters			
activated, looped promoter strength	0,05	s^{-1}	
CI binding rate with looping from bound dimers at OR	0,1	s^{-1}	k_{AL}^{+}
CI unbinding rate with looping from bound dimers at OR	0,6	s^{-1}	k_{AL}^{-}
CI looping factor	2		
LacI binding rate with looping from bound O1	528	s^{-1}	k_{RL1}^{+}
LacI unbinding rate with looping from bound O1	0,015	s^{-1}	k_{RL1}^{-}
LacI binding rate with looping from bound O2	510	s^{-1}	k_{RL2}^{+}
LacI unbinding rate with looping from bound O2	0,035	s^{-1}	k_{RL2}^{-}
rate of changing into less activated looping state	1,5	s^{-1}	
rate of changing back into more activated looping state	0,0005	s^{-1}	
Steric occlusion parameters			
binding rate of CI at OR1&OR2 if LacI is bound to O1	0,0005	s^{-1}	k_{A12R1}^{+}
unbinding rate of CI at OR1&OR2 if LacI is bound to O1	0,2083	s^{-1}	k_{A12R1}^{-}
binding rate of LacI at O1 if CI is bound to OR1&OR2	0,00075	s^{-1}	k_{R1A12}^{+}
unbinding rate of LacI at O1 if CI is bound to OR1&OR2	0,004	s^{-1}	k_{R1A12}^{-}
binding rate of CI at OR2 if LacI is bound to O1 and CI to OR1	0,2	s^{-1}	k_{A2R1}^{+}
unbinding rate of CI at OR2 if LacI is bound to O1 and CI to OR1	0,5	s^{-1}	k_{A2R1}^{-}
binding rate of LacI at O2 if CI is bound to OR1&OR2	0,000133	s^{-1}	
unbinding rate of LacI at O2 if CI is bound to OR1&OR2	0,5938	s^{-1}	
binding rate of CI at OR1&OR2 if LacI is bound to O2	0,01	s^{-1}	
unbinding rate of CI at OR1&OR2 if LacI is bound to O2	0,2083	s^{-1}	
binding rate of CI at OR1&OR2 if LacI is bound to O1 and O2	0,000005	s^{-1}	
unbinding rate of CI at OR1&OR2 if LacI is bound to O1 and O2	1,25	s^{-1}	
CI looping rate if CI is bound at OR1&OR2 and LacI at O1	0,0002	s^{-1}	k_{ALR1}^{+}
CI unlooping rate if CI is bound at OR1&OR2 and LacI at O1	0,175	s^{-1}	k_{ALR1}^{-}
Facilitated dissociation parameters			
factor for facilitated dissociation due to CI concentration	20		
factor for facilitated dissociation due to LacI concentration	30		

Table 6.5. Comparison of R^2-values for the model fit in systems with and without DNA looping.

R^2-values were calculated between experimental data (18 or 6 replicates) and model trajectories for t0 to t6. Overall values were calculated for samples with both TFs ('AR', 'A->R', 'R->A').

		Looping				
Conditions	**Equilibrium (no FD, no SO)**	**Non-equilibrium without interference states (no FD, no SO)**	**Non-equilibrium with interference states (no FD, no SO)**	**Non-equilibrium (no FD, SO)**	**Non-equilibrium (FD, no SO)**	**Non-equilibrium (FD and SO) =Interference**
Overall	**0,74**	**0,76**	**0,83**	**0,84**	**0,83**	**0,90**
A	0,93	0,93	0,93	0,97	0,91	0,97
R	0,94	0,84	0,83	0,98	0,84	0,92
AR	0,63	0,73	0,85	0,93	0,88	0,93
AR240	0,82	0,79	0,74	0,94	0,77	0,97
AR20	0,99	0,96	0,73	-0,10	0,54	0,60
AR5	0,62	0,65	0,79	0,92	0,81	0,92
AR1	0,75	0,77	0,95	0,94	0,96	0,93
RA240	0,58	0,66	0,79	0,91	0,82	0,90
RA20	0,67	0,70	0,88	0,94	0,91	0,92
RA5	0,59	0,68	0,81	0,91	0,84	0,90
RA1	0,60	0,69	0,82	0,91	0,86	0,89

	No looping		
Conditions	**Equilibrium**	**Non-equilibrium**	**Interference**
Overall	**0,47**	**0,53**	**0,67**
A	0,64	0,64	0,64
R	0,05	0,04	0,04
AR	0,92	0,91	0,91
AR240	0,23	0,66	0,97
AR20	0,99	0,99	0,95
AR5	0,28	0,21	0,88
AR1	0,90	0,54	0,86
RA240	0,44	0,88	0,88
RA20	0,51	0,50	0,64
RA5	0,22	0,25	0,03
RA1	0,22	0,26	0,41

Table 6.6. FDR adjusted p-values comparing 'A->R' or 'R->A' OD_{600} values against the 'AR' control value at several time points for growth with tetracycline.

T-tests were performed at different time points after t0 between 'A->R' or 'R->A' fluorescence and 'AR' fluorescence. Values that are above the significance level α = 0.05 are shown in red.

Time (min)	'A240R'	'A20R'	'A1R'	'R240A'	'R20A'	'R1A'
240	p=0,013; f=-9,26	p=0,456; f=-0,92	p=0,012; f=-9,82	p=0,038; f=5,10	p=0,119; f=-2,64	p=0,014; f=-8,72
300	p=0,030; f=-6,01	p=0,413; f=-1,03	p=0,056; f=-4,22	p=0,017; f=7,93	p=0,042; f=-4,77	p=0,005; f=-15,33
360	p=0,010; f=-10,96	p=0,379; f=-1,13	p=0,043; f=-4,91	p=0,016; f=8,02	p=0,021; f=-6,97	p=0,003; f=-18,90
420	p=0,007; f=-12,63	p=0,250; f=-1,63	p=0,043; f=-4,93	p=0,012; f=9,44	p=0,011; f=-10,11	p=0,001; f=-29,25
480	p=0,018; f=-7,94	p=0,250; f=-1,62	p=0,045; f=-4,79	p=0,014; f=8,89	p=0,012; f=-9,69	p=0,001; f=-28,70
540	p=0,022; f=-7,19	p=0,202; f=-1,92	p=0,027; f=-6,30	p=0,026; f=6,26	p=0,014; f=-8,84	p=0,002; f=-24,99
600	p=0,027; f=-6,46	p=0,202; f=-1,91	p=0,023; f=-6,92	p=0,033; f=5,51	p=0,010; f=-10,50	p=0,002; f=-27,27
660	p=0,029; f=-6,12	p=0,214; f=-1,83	p=0,027; f=-6,32	p=0,055; f=4,15	p=0,011; f=-10,23	p=0,003; f=-20,98
720	p=0,030; f=-5,99	p=0,212; f=-1,84	p=0,027; f=-6,30	p=0,100; f=2,93	p=0,008; f=-12,16	p=0,002; f=-22,69
780	p=0,031; f=-5,88	p=0,204; f=-1,90	p=0,023; f=-6,97	p=0,144; f=2,34	p=0,010; f=-10,63	p=0,001; f=-37,51

Table 6.7. FDR adjusted p-values comparing growth rates of 'A->R' or 'R->A' conditions to 'AR'.

T-test were performed for growth rates calculated between t2 and t8 over 6 replicates.

'A240R'	'A20R'	'A1R'	'R240A'	'R20A'	'R1A'
p=0,045; f=-5,83	p=0,219; f=-1,94	p=0,040; f=-6,92	p=0,040; f=7,77	p=0,058; f=-4,64	p=0,011; f=-27,03

Table 6.8. FDR adjusted p-values comparing 'A->R' or 'R->A' fluorescence values against the 'AR' control value at several time points in experiments without looping of DNA.

T-tests were performed at different time points after t0 between 'A->R' or 'R->A' fluorescence and 'AR' fluorescence. Values that are above the significance level α = 0.05 are shown in red.

Time (min)	**'A240R'**	**'A20R'**	**'A1R'**	**'R240A'**	**'R20A'**	**'R1A'**
60	p=0,001; f=15,85	p=0,002; f=5,88	p=0,310; f=0,25	p=1,72E-06; f=15,85	p=0,005; f=5,88	p=0,837; f=0,25
120	p=0,001; f=39,14	p=0,004; f=5,22	p=0,019; f=-0,31	p=2,62E-07; f=39,14	p=0,008; f=5,22	p=0,808; f=-0,31
180	p=0,001; f=28,80	p=0,005; f=4,54	p=0,011; f=-0,99	p=1,77E-05; f=28,80	p=0,012; f=4,54	p=0,432; f=-0,99
240	p=0,001; f=15,39	p=0,005; f=3,93	p=0,001; f=-0,71	p=6,61E-05; f=15,39	p=0,019; f=3,93	p=0,570; f=-0,71
340	p=0,012; f=4,13	p=0,017; f=3,08	p=0,045; f=-0,91	p=0,016; f=4,13	p=0,041; f=3,08	p=0,468; f=-0,91
380	p=0,017; f=6,25	p=0,026; f=3,74	p=0,035; f=-0,26	p=0,005; f=6,25	p=0,022; f=3,74	p=0,835; f=-0,26
420	p=0,019; f=6,70	p=0,017; f=5,89	p=0,103; f=-1,96	p=0,004; f=6,70	p=0,005; f=5,89	p=0,141; f=-1,96
460	p=0,019; f=10,88	p=0,022; f=7,27	p=0,076; f=-0,77	p=0,001; f=10,88	p=0,003; f=7,27	p=0,545; f=-0,77
500	p=0,022; f=10,62	p=0,048; f=6,59	p=0,017; f=-0,35	p=0,001; f=10,62	p=0,004; f=6,59	p=0,785; f=-0,35
540	p=0,022; f=3,95	p=0,019; f=1,77	p=0,126; f=-1,87	p=0,019; f=3,95	p=0,173; f=1,77	p=0,156; f=-1,87

Table 6.9. Comparison of R^2-values for the model fit in experiments with the lacO_1 mutant.

R^2-values were calculated between experimental data (6 replicates) and model trajectories for t0 to t6. Overall values were calculated for samples with both TFs ('AR', 'A->R', 'R->A').

	No looping with *lacO_1* mutant	
Conditions	**Equilibrium**	**Non-equilibrium**
Overall	**0,918081**	**0,916304**
B	0,977199	0,977219
A	0,825327	0,824652

R	0,851155	0,835213
AR	0,926568	0,911066
AR240	0,883105	0,899491
AR20	0,994914	0,987054
AR5	0,834139	0,854511
AR1	0,984979	0,974295
RA240	0,927155	0,935269
RA20	0,968008	0,954429
RA5	0,926338	0,926142
RA1	0,977199	0,977219

Table 6.10. Strains and plasmids.

Strain name	**Current phenotype**
ASE006	BW27784 ΔlacI785
CI088	ASE006-Pcon-tetR integrated on the chromosome at the phage P21 attachment site
CI125	CI088 with araC – pBAD – LacI recombined into galK
CI158	CI125 containing plasmids pZE3-Pswitch and pZS21-CI-Lite
CI180	CI125 containing plasmids pZE3-Psw_noloop and pZS21-CI-Lite
pZS21-λ CI-Lite	Low copy plasmid with λ CI under the control of a $P_{LtetO-1}$ promoter, and tagged for degradation with an ssrA tag
pZE3-Pswitch	Medium-copy plasmid containing the synthetic P_{syn} controlling a fluorescence gene and the P_L promoter region from phage λ
pZE3-Psw_noloop	Medium-copy plasmid containing the synthetic P_{syn} without the binding sites for looping controlling a fluorescence gene
PBAD-LacI	Plasmid PBAD24 with AraC and the PBAD promoter region controlling LacI
pAH81frt-cat – Pcon-tetR	CRIM plasmid with chloramphenicol resistance gene flanked by FRT sites, carrying P_{con}-tetR for integration at the phage P21 attachment site

Table 6.11. Primers.

Primers	
Pswitch_R0	TTCTCGAGCAGTgactg
Pswitch_F0	cagtcACTGCTCGAGAATTGTGAGCGCTCACAA

Pswitch_R17	TCTGGCGGTGATAAAATTGTGAGCGCTCACAA
Pswitch_F33	TTTTATCACCGCCAGAGGTAAAATAGTCAACACGC
Pswitch_R49	GGTAAATATCTAACACCGTGCGTGTTGACTATTTTACC
Pswitch_F68	ACGGTGTTAGATATTTACCCCTTGCAGTGATAGATTT
Pswitch_R87	GTTACTCGCTCACATTTAAATCTATCACTGCAAGG
Pswitch_F105	AAATGTGAGCGAGTAACAACCAATTCATTAAAGAGGA
Pswitch_R122	GCATGCGGTACCTTTCTCCTCTTTAATGAATTGGTT
Pswitch_F142	GAAAGGTACCGCATGCGTAAAGGAGAAGAACcatt
Pswitch_R158	gccagacataatgtttgaaatgGTTCTTCTCCTTTAC
Pswitch_R177	tcaaacattatgtctggc
Pswitch_shortF0	ACTGCTCGAGAATTGTGAGCGC
Pswitch_shortR0	GTTCTTCTCCTTTACGCATGCG
PL3-toPswitch_fwd	ATCTGGATCCCAAGGTGTTCTGGTCGG
PL3-toPswitch_rev	TATCACTAGAGTTGGTTATCTGTATGTT
PL3-_fwd	GATAACCAACTCTAGTGATAAATTATCTCTGGC
PL3-_rev	AGCTTCTAGACCTGCTGATGTGCTCAGTATC
R0	CGCTCACATTTCTCGAG
F0	CTCGAGAAATGTGAGCGAGTAACAACCTATCACCG
R17	TTGACTATTTTACCTCTGGCGGTGATAGGTTGTTACT
F35	CCAGAGGTAAAATAGTCAACACGCACGGTGTTAGA
R54	TCACCACAAGGGATAAATATCTAACACCGTGCGTG
F70	TATTTATCCCTTGTGGTGATAGATTTAACGTAATTGTGAGC
R89	TCCTCTTTAGAATTGTTATCCGCTCACAATTACGTTAAATCTA
F111	GGATAACAATTCTAAAGAGGAGAAAGGTACCGCATGC
R132	ggtaGTTCTTCTCCTTTACGCATGCGGTACCTTTC
F148	GTAAAGGAGAAGAACtaccactgagatgtatgatggc
galK_term_araC_fwd	CCGGAGTGTAAGAAATGAGTCTGAAAGAAAAAACACAATCTC CCAATTATGACAACTTGACGGC
galK_LacI_term_rev	CGGTACGGCTGACCATCGGGTGCCAGTGCGGGAGTTTCGT CGCAAAAAGGCCATCCGTCAG
Pcon_fwd	ACTG GCATGC AATTCACCGTCGTTG TTGACA TTTTTAAGCTTGGCGGT TATAATGGTACC ATAAGGAGGTGGATCCGGCA TAAATATGGCTGGTTCTCG

Pcon_rev	CGAGAACCAGCCATATTTATGCCGGATCCACCTCCTTATGGTACC ATTATAACCGCCAAGCTTAAAAATGTCAACAACGACGGTGAATTG CATGCCAGT
GFP_fwd	ATGCGTAAAGGAGAAGA
GFP_rev	ACGTGGATCCTTATTTGTATAGTTCATCCAT
LacI_fwd_EcoRI	ACGTGAATTCATGGTGAATGTGAAACCAGTAACG

6.6 Methods

P_{syn} architecture

We constructed a synthetic P_{syn} under the control of two inducible TFs: an activator, λ CI, and a repressor, LacI (Fig. 6.1B). P_{syn} was obtained from the phage λ P_{RM} promoter by mutating the λ O_{R3} operator to eliminate repression by CI[11] and introducing two operators for LacI binding[10], positioning the strong LacI operator as in the natural promoter (P_{LacZ}) and the weaker operator at a distance optimal for DNA looping[174] (Fig. 6.2). Both TFs can loop the DNA when bound to their operators (Fig. 6.2), which increases the efficiency of the respective regulatory state[10,158]. The P_{syn} controls the expression of a green fluorescent marker (*gfpmut3*). *E. coli* contains a high frequency of promoters that are regulated by more than one TF and most of these promoter architectures contain operators separated by less than 10-20 bps[13], making our synthetic system a relevant model for natural promoters.

In the modified P_{syn}, which did not allow for looping, we removed the λ phage P_L promoter region (O_{L1} and O_{L2}) as well as the second operator for LacI, O_2 (Fig. 6.13). Due to their locations, those operators should not have any impact on promoter regulation on their own. CI-mediated looping on plasmids has been shown before[15] and looping in our original P_{syn} system was tested by measuring fluorescence activation (repression) with and without the P_L promoter region for CI (the second operator, O_2, for LacI). In both cases regulatory activity was significantly decreased in the absence of operators supporting DNA looping. Fluorescence increase due to activation was much slower without looping, whereas repression efficiency was overall considerably decreased (compare 'A' / 'R' curves in Fig. 6.8A, 6.14A). To test the predictions of memory loss we introduced a mutation into the lacI operator O_1, that weakened lacI binding but didn't abolish it completely[167].

As the region between λ P_R and P_L was taken from the bacteriophage genome and contains two genes involved in superinfection exclusion (*rexA*, *rexB*), we cloned a version of P_{syn} where we replaced those two genes with half of the *galK* gene and half of the *galT* gene and reproduced the transient memory dynamics in this system (Fig. 6.16). For the growth measurements in the presence of an antibiotic, *gfpmut* was replaced with *tetA.*

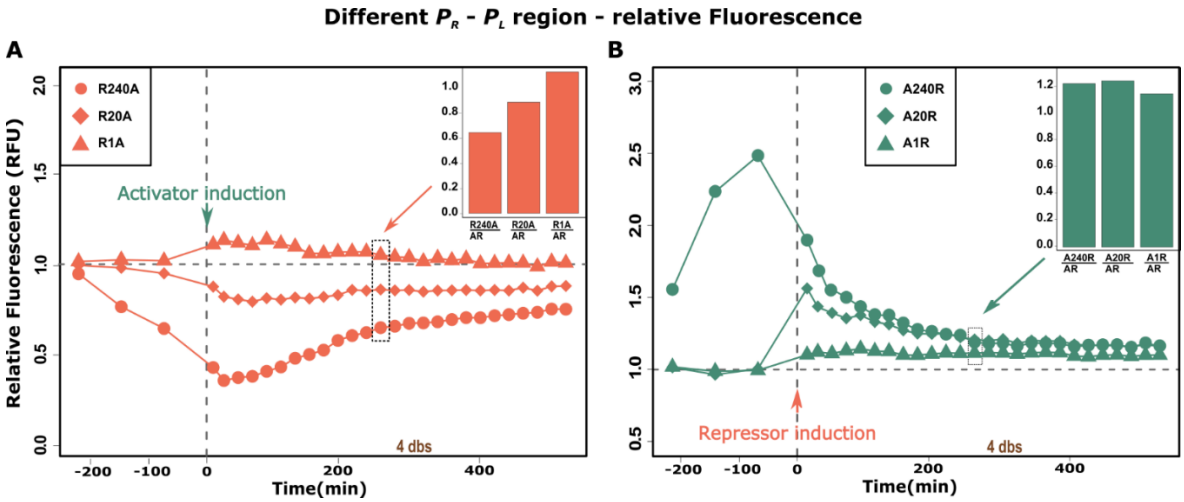

Figure 6.16. Relative fluorescence showing transient memory in gene expression for time delays of 240, 20 and 1 minute at P_{syn}, where the P_R-P_L region was replaced with parts of the *galK* and *galT* genes.

*Shown are OD_{600}-normalized fluorescence values that were divided by 'AR' fluorescence values over the time course of the experiment (x-axis shows time in minutes). Green and orange curves show the 'R->A' **(A)** and 'A->R' **(B)** conditions respectively. The dashed vertical line indicates the time point when the second inducer was added. The dashed horizontal line at 1 indicates that fluorescence is equal to the 'AR' control. Bar plots show the relative fluorescence values at t4.*

Plasmid and strain construction

All strains, plasmids and primers used are listed in Table 6.10 and 6.11. The host strain used in all experiments was obtained from BW27784 (CGSC# 7881) by P1 transduction of a *lacI::kan* deletion cassette from the Keio collection strain JW0336 and subsequent removal of the kanamycin resistance gene by using pCP20[99]. The *lacI* gene was cloned into the MCS of pBAD24 [175] and from the resulting plasmid the region AraC-P_{BAD}-LacI was recombined with λ red onto the chromosome replacing *galK*[176]. *tetR* was placed under the control of a synthetic, constitutive promoter P_{con}[118] and inserted into the P21 attachment site on the chromosome of BW27784 using the CRIM plasmid system[177]. λ CI was tagged for degradation (ssrA tag)[178] to avoid its accumulation in the cell, which proved to be toxic and put under the control of a $P_{LtetO-1}$ promoter (short P_{Tet}) onto a low copy number km^R plasmid pZS[99]. P_{BAD} and $P_{LtetO-1}$ were chosen as promoters for the input signals as they are both tightly regulated, fast and strong promoters[99,179,180]. The P_{syn} promoters with and without looping were assembled synthetically from oligos and fused to *gfpmut3*. The genomic region starting after the stop codon of CI until (and including) the O_{L1} of P_L were amplified from phage λ DNA to keep the distance and environment as similar as possible to the natural regulatory context[158]. This genomic region includes two genes *rexA* and *rexB* that provide superinfection exclusion in the presence of lytic phages[181] and hence should not interfere with our promoter system. The P_{syn} - *gfpmut3* and the Lambda DNA piece containing P_L were cloned together into a cm^R pZE plasmid[99]. The main operator for LacI is placed after the transcription start site as in the WT *lac* operon, thereby we could avoid changing the CI O_R sites or RNA polymerase (RNAP) binding sites. The CI

repressor operator sites were at WT locations with regard to P_{RM} but we introduced mutations into O_{R3} (r3) and O_{L3} (OL3-4) to avoid repression by CI[121,182]. As the basal level of λ P_{RM} is in general relatively low and was further impaired by the *lac* O_1 binding site, we introduced a strong ribosomal binding site upstream of *gfp* to increase the difference between basal and repressed fluorescence levels. The origins of the pZE and pZS* plasmid families are compatible and can therefore be stably propagated in cultures over long timescales. With these constructs activated fluorescence levels were 5.6-fold higher than basal at equilibrium and repressed levels were 2.7-fold lower than basal.

Fluorescence assays

All cells were grown overnight at 30°C in M9 medium supplemented with 0.5% casamino acids, 0.5% glycerol, 50μg/ml kanamycin and 30μg/ml chloramphenicol. Those cultures were used to inoculate twelve conditions (with 18 replicates each for 'looping' and 6 replicates each for 'no looping' and 'no looping with lacO_1 mutant') 1:100 in 96 well plates and were grown at 30°C at 220 rpm. We found that growth rates were most comparable at these conditions, among replicates, different induction scenarios, as well as between different experiments. To test if high concentrations of TFs affected our observations or the growth rates, we repeated the experiments for 6 replicates with low inducer concentrations and found similar growth rates as well as transient memory (Table 6.1, 6.2). Similarly, the model showed transient memory if activator concentration and degradation rate were varied (Fig. 6.17).
We also measured transient dynamics at 25°C and 35°C and found increased memory at 25°C and decreased memory at 35°C (Fig. 6.18), as would be expected for an effect that is dependent on TF binding and hence DNA replication.
Inducers were added at 25ng/ml (5ng/ml for low concentrations) for *aTc* and 0.05% (0.001% for low concentrations) for *arabinose* after cultures reached an initial OD_{600} of 0.1 in the following way: As controls, one sample was not induced at all ('basal expression'), one was activated with *aTc* ('A), one was repressed with *arabinose* ('R') and one was induced for both TFs at the same time ('AR'). Four conditions were induced with *aTc* (λ CI expression) and after a specific time delay *arabinose* (LacI expression) was added to the medium as well. The time delays between inductions of the two TFs were 240 ('A240R'), 20minutes ('A20R'), 5minutes ('A5R') and 1minute ('A1R'). The last four conditions were induced first with *arabinose* (LacI expression) and after 240 ('R->240A'), 20minutes ('R20A'), 5minutes ('R5A') and 1minute ('R1A') with *aTc* (λ CI expression). Induction kinetics for a P_{tet}-based promoter and P_{BAD} are expected to be on the order of minutes[175,183].

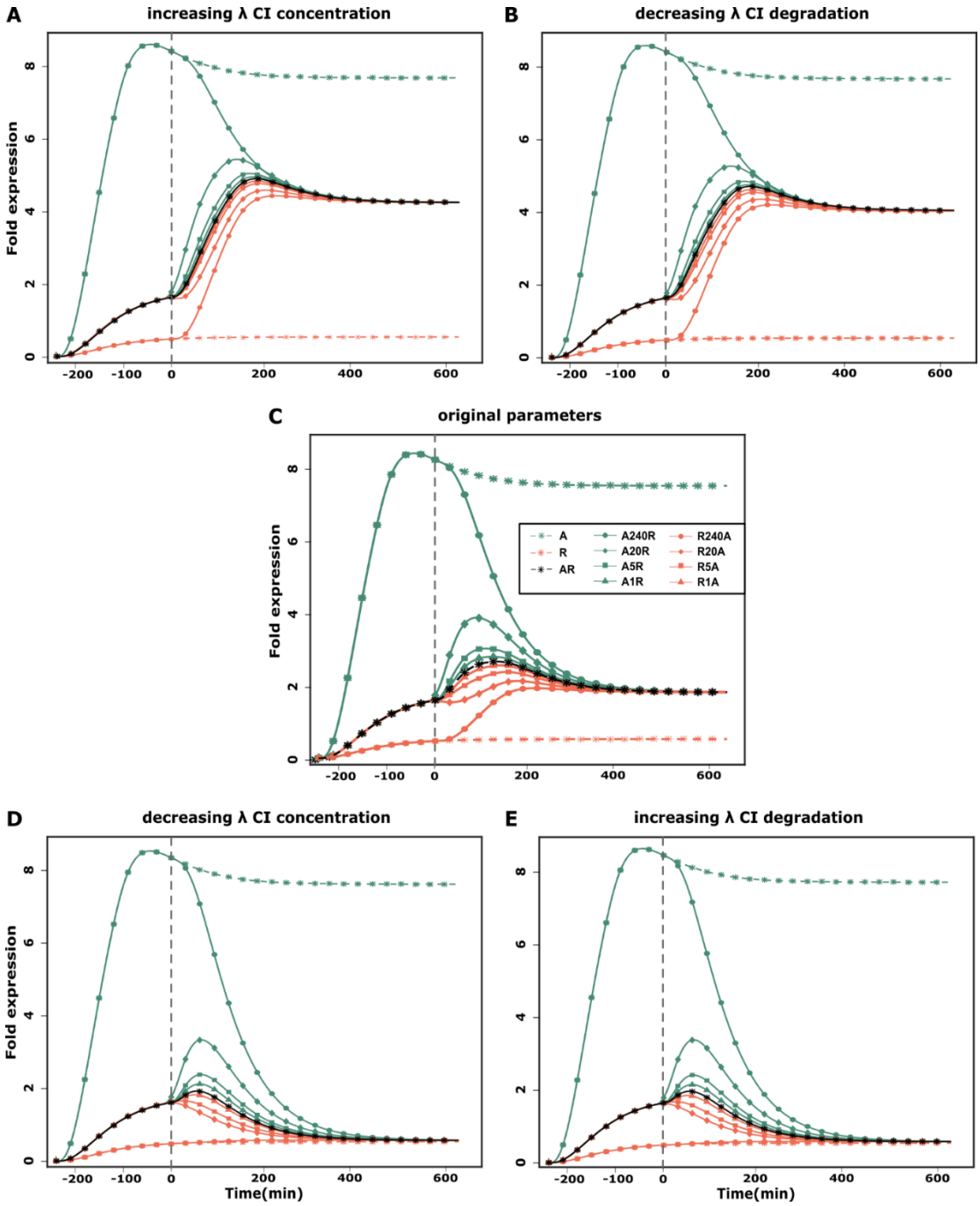

Figure 6.17. Memory effects for varying activator concentration and activator degradation rate on gene expression levels.

Fluorescence (in AU) over 14h (x-axis shows time in minutes) is shown for different delay scenarios in simulations with ***C)*** *original CI concentration and degradation rate,* ***A)*** *increasing or* ***D)*** *decreasing CI concentration,* ***B)*** *decreasing or* ***E)*** *increasing CI degradation rate. Colors correspond to Fig. 6.8A with 'AR' in black, 'A->R' in green and 'R->A' conditions in orange. Increasing (decreasing) CI concentration has a similar effect to decreasing (increasing) CI degradation rate. Although the steady state expression level varies if activator concentration are varied, the memory effect remains in all cases (varying repressor concentration leads to opposing changes in steady state levels but similar memory effects).*

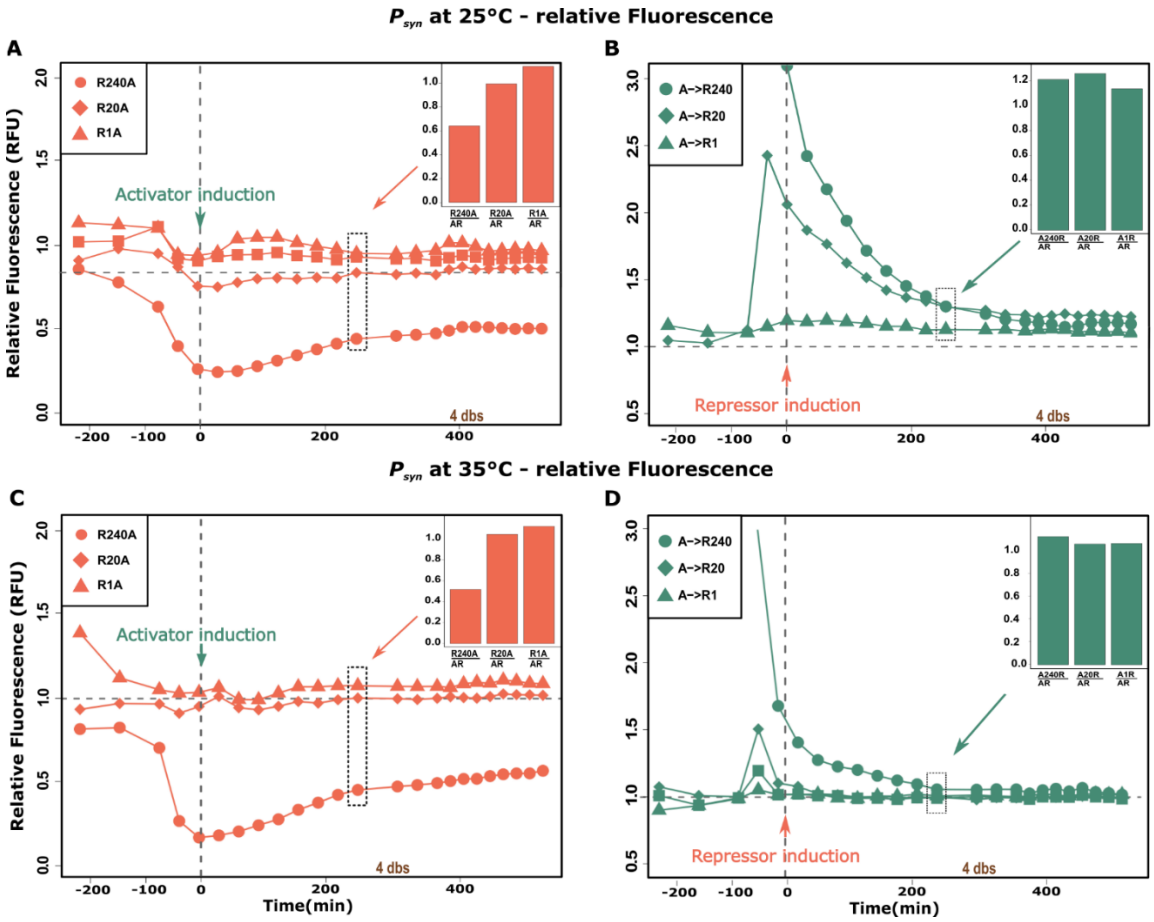

Figure 6.18. Experimental observations of transient memory in gene expression at A, B) 25°C or C,D) 35°C.

TF were induced with time delays of 240, 20 or 1min. Shown are OD_{600}-normalized fluorescence values that were divided by 'AR' fluorescence values over the time course of the experiment (x-axis shows time in minutes). Green and orange curves show the 'R->A' ***(A, C)*** *and 'A->R'* ***(B, D)*** *conditions respectively. The dashed vertical line indicates the time point when the second inducer was added. The dashed horizontal line at 1 indicates that fluorescence is equal to the 'AR' control. Bar plots show the relative fluorescence values at t4.*

Cultures were diluted every 6h to keep them in exponential phase, grown on the shaker at 30°C and measured using Biotek H1 plate reader at first for 240 every 60min., then for 10h ('looping' experiment) or 6h ('no looping' and 'low concentration' experiments) every 30min. We tested if the second dilution affected the observed memory and found no significant differences in relative fluorescence values for any of the conditions (Fig. 6.19). Growth was continued overnight in the plate reader with measurements every 20min. In the first 14h of the 'looping' experiment measurements were taken using plate reader and FACS. In the plate reader fluorescence and OD_{600} measurements were taken and fluorescence measurements were normalized by the OD_{600} measurements to account for differences in cell density. Data were analyzed using R statistical software. Promoter activities were calculated from fluorescence and OD_{600} measurements using equations (6), (11) and (13) from[184]. The cell doubling time was approximately 0.75/h. All fluorescence values were given

as fold-changes of the basal expression level. Relative fluorescence ratios were calculated by dividing OD_{600} -normalized fluorescence values of the individual scenarios with delayed induction ('A->R' or 'R->A) by OD_{600} -normalized fluorescence values of the 'AR' control at indicated time points.
Flow cytometry analysis was carried out on the FACS Canto II Analyzer (BD Biosciences, San Jose, CA) equipped with FACSDiva software (version 6.1.3, BD Biosciences). Sensitivity of the lasers was determined within the daily setup using BD FACS 7-color setup beads. Conditions were diluted into M9 buffer (1x M9 salts supplemented with 1mM $MgSO_4$ and 0.1 mM $CaCl_2$) to yield ~500 events/sec at medium flow rate. For GFP fluorescence a 488-nm argon excitation laser and a 530/30 nm band-pass filter was used. For each culture 20,000 events were collected and data was analyzed with FlowJo software.

Growth measurements in the presence of tetracycline
Cells were grown overnight and diluted in the morning for 3 replicates of each population as described in the previous section for fluorescence measurements. However, induction of 'A', 'R', 'A240R' and 'R240A' was done immediately after inoculation in order to avoid dilution throughout the experiment. Otherwise the induction schemes were carried out as described before and after addition of the second inducer 5ug/mL tetracycline were added to each sample (results were comparable at 15ug/mL). Conditions were grown and measured (OD600) every 30min in the plate reader.
It is highly unlikely that tetA pumps out *aTc* in significant amounts, even though it is a tetracycline derivative, as tetA only confers very weak resistance to *aTc*[185]. Further, we tested *aTc* induction of a P_{Tet}- controlled GFP in the absence and presence of tetA and fluorescence induction rates were not significantly different (p=0.3615, f=-0.925).

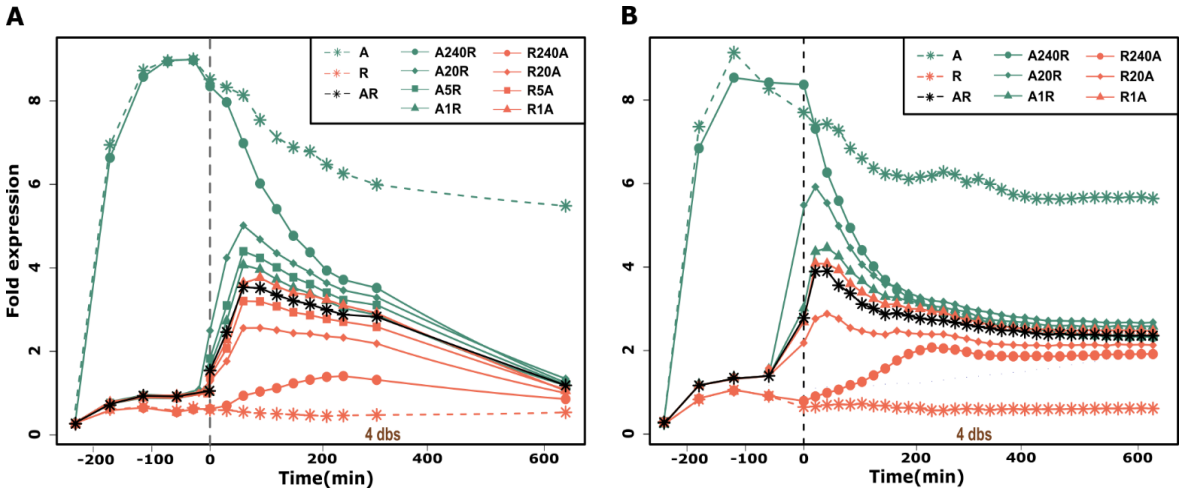

Figure 6.19. Experimental observations of transient memory in gene expression for P_{syn} with looping A) with dilution at t-1 and t6 or B) with dilution at t-1 only.

The x-axis shows time in minutes centered at induction of the second inducer and the y-axis shows OD_{600}-normalized fluorescence (NFU) with relative errors given at each measurement point. Colors and symbols correspond to Fig. 8A with 'AR' in black, 'A->R' in green and 'R->A' conditions in orange. Cells were grown in 3x sextuples in 96 well plates in M9 with 0.5% CAA on a shaker at 30°C. The first TF was induced 240, 20,

5 or 1 minute earlier as indicated. The second TF was induced at the same time, t0, for all conditions, which is indicated by the dashed vertical line.

Statistical tests

Collected data was tested for normality (Shapiro-test) and subsequently we compared mean expression values through t-tests with FDR correction for multiple comparisons in RStudio. At first mean expression values were compared between basal and activated or repressed states ($p < 0.0001$, f=-19.52 / 9.28) as well as between activated and repressed ($p < 0.0001$, f=19.46) to make sure they were significantly different. For memory assessment means between delay conditions and the 'AR' control were compared at indicated time points. We tested if the dilution rate was significantly different from the drop in fluorescence for 'A->R' conditions by calculating the slope of the logarithmic OD values and comparing it with the negative slope of the logarithmic, normalized fluorescence values (that were corrected for their descent to the equilibrium value, not to zero expression) for six replicates during the exponential growth phase (t0 to t3) by using t-tests and FDR correction. We also calculated the mean and standard deviation over all decay and dilution rates for A->R conditions in the following way:

$$rate = \frac{\frac{\log(x(t))}{\log(x(t+\Delta t))}}{\Delta t},$$

With x indicating either OD or fluorescence values at time point t or t+Δt.

We calculated maximum fluorescence production for the 'AR' control and 'R->A' conditions after t0 by finding the maximum fluorescence value and subtracting the fluorescence value at t0. Error bars were calculated as relative errors by using error propagation. Further, we compared the calculated slope of increase in fluorescence from t0 to t1.5 between the 'AR' control and R->A conditions by using t-tests with FDR correction.

We determined the fit between a certain model version (the equilibrium, non-equilibrium without interference states, non-equilibrium with interference states, non-equilibrium with SO (steric occlusion) or FD (facilitated dissociation) or both (interference model)) and the experimental data (looping, no looping and no looping with *lacO$_1$* mutation) for 11 time points after t0 for all conditions individually (Fig. 6.8, 6.9, Table 6.5, 6.9) in Matlab using the following version of Fraction of Variance explained: $$FVE = \frac{SS_{er}}{"SS_{tot}"} = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i {y_i}^2}.$$

Reaction rate models

Reaction rate models calculate the expression dynamics of a promoter system by representing the plausible state configurations and transitions between promoter states[186,187]. The regulatory states of the promoter arise from different binding states of regulatory proteins[84] (i.e. activator, repressor and/ or RNAP can be bound or not bound in various combinations at operator sites) and different DNA configurations (e.g. looped or not looped). Gene expression levels are then assumed to be proportional to the probability of RNA polymerase (RNAP) being bound at the

promoter, which is dependend on the binding state of the other TFs at the promoter (e.g. activators enable RNAP binding and repressors inhibit it)[51,57,160].

For instance, the production of a protein x from an active (RNAP / activator bound) promoter p_{ON} with rate α and the degradation of x with rate β would be described by the following reactions: $p_{ON} \xrightarrow{\alpha} p_{ON} + x$

$$x \xrightarrow{\beta} \emptyset,$$

resulting in the differential equation for x: $\dot{x} = \alpha p_{ON} - \beta x.$ (1)

Different regulatory states are associated with varying transcriptional activity and overall promoter output *x* results from summing the product of a state's probabilty and its corresponding transcription rate:

$$\dot{x} = (\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \cdots) - \beta x, \quad (2)$$

with α_i being the transcription rate of state i and p_i the probability of being in state i. The time evolution of regulatory state probabilities, protein concentrations and GFP production and maturation were calculated using differential equations. GFP mRNA and tRNA production and degradation as well as protein folding into its fluorescent state were explicitely modelled[188].

The probability of being in a particular promoter regulatory state, and hence p_{ON}, was calculated differently for the three models used in this study. We will demonstrate the difference using the simple example reaction of a repressor R binding to an unbound promoter U yielding a bound promoter B:

$$U + R \underset{k-}{\overset{k+}{\rightleftarrows}} B \quad (3)$$

Thermodynamic models are reaction rate models that assume that the promoter state is at equilibrium: $\dot{p}_{ON} = 0$. Hence, the description of expression dynamics from p_{ON} is only dependent on equilibrium TF concentrations and TF-DNA binding affinities. Instead of individually modelling forward and backward transitions between two states of the promoter, this assumption enables the use of equilibrium constants to describe transitions (i.e. the forward and backward rate between two states are lumped into one equilibrium constant $K_R = \frac{backward\ rate}{forward\ rate}$). For the above reaction of repressor binding to the promoter (3), we obtain the following equation for promoter expression:

$$p_{ON} = \frac{1}{1 + \frac{[R]}{K_R}} \quad (4)$$

with the equilibrium dissociation constant $K_R = \frac{k-}{k+}$.

For the non-equilibrium equations, we considered forward and backward rates (*k+*, *k-*) seperately as the promoter state is not assumed to equilibrate immediately ($\dot{p}_{ON} \neq 0$):

$$\dot{p}_{ON} = k^- \cdot B - k^+ \cdot R \cdot U \quad (5)$$

The non-equilibrium interference model additionally includes states where both TFs can bind at the same time, including looping through the activator, but not the repressor (as the repressor loop likely is too tight to allow for binding of the activator). In the full model we assumed that steric occlusion (SO) between TFs would decrease association rates, and that facilitated dissociation (FD) would make the dissociation rate of one TF dependent on the concentration c_A of the other TF[14,189], e.g. an activator, that binds in close vicinity of the repressor and modifies repressor unbinding (k^-):

$$p_{ON} = k^-(c_A) \cdot B - k^+ \cdot R \cdot U \qquad (6)$$

The concentration-dependence according to FD was implemented by multiplying the dissociation rate by the number of molecules of competing TF present. We allowed for binding of both TFs at the same time as well as for DNA looping by the activator λ CI, but not looping by the repressor LacI as such a tight loop would likely inhibit binding to operators within the loop[190]. DNA association rates for those 'interference' transitions were decreased as we assumed that the TFs would sterically interfere with each other's binding.

We used this model (for a diagram of states and transitions see Fig. 6.7) to simulate fluorescence levels from P_{syn} over 14h in Matlab R2015a and compared the obtained values to experimental data. All fluorescence values were normalized by basal expression from P_{syn} as in the experimental meeasurements. Fluorescence ratios were calculated by dividing fluorescence levels of individual scenarios with delayed induction ('A->R' or 'R->A) by fluorescence levels of the 'AR' control. Binding rates and TF concentrations that were used in the simulations were fitted to the experimental data using a simulated annealing algorithm (non-equilibrium model with interference, FD and SO for the looping case: 28 parameters, and the equilibrium model with interference: 11 parameters), but were kept within realistic ranges as found in previous studies at microbial promoters[63,95,99,118,155,175,182,188,191–201] (Table 6.4). The degradation tag of the activator, λ CI, was taken into account and accordingly faster than degradation by dilution of the other proteins[178].

6.7 Author contributions

C.I. (Claudia Igler), A.S. (Ali Sezgin), A. N.-S. (Anna Nagy-Staron), G.T. (Gašper Tkačik), C.C.G. (Călin C. Guet) conceived the study together. C.I. designed and carried out the experiments and analyzed the data. K.T. (Kathrin Tomasek) carried out FACS experiments. C.I. wrote the code and ran the model. A.S., G.T. and J.M. (John Marko) helped with model conceptualization. C.I. wrote the initial draft of the manuscript and revised it together with G.T., A.S., A. N.-S., K.T., J.M. and C.C.G.

7. Conclusion

7.1 Network evolution – local and global determinants

Transcriptional regulation lies at the foundation of cellular function and behavior, which emerges from complex interconnections between individual transcriptional units[4,6]. In order to gain a better understanding of emerging cellular behavior, we first have to understand how its parts function and come together, and the last decades have revealed an astounding richness and intricacy in gene regulatory mechanisms of even the simplest organisms like bacteria and viruses (see Chapter 1). Thus it has proven to be useful to study isolated systems in order to elucidate the molecular basis of gene regulatory processes.

We have used this concept to provide insights into the biophysical mechanisms that allow or constrain the evolution of transcriptional regulation. Instead of trying to entangle the complex features and properties of whole networks that can potentially play a role in their evolution, we focused on the connections between individual components: TFs and their binding sites. Surprisingly, very few biophysical parameters are sufficient to determine the potential of a TF to rewire regulatory connections. In our study we used very well characterized TFs, which allowed us to explore the significance of all biophysical parameters in the system, but our findings indicate that only the wild type binding energy and the energy matrix (describing the energy penalty of binding site mutations) of a TF need to be determined in order to find its evolutionary potential. For approximate estimations, one can forgo the determination of the energy matrix and replace it with the general estimation of 1-3k_BT energy penalty per mutation[24]. This insight provides simple, conceptual tools that can be used to obtain a predictive understanding of regulatory network changes based on the binding properties of TFs. These implications will be discussed further below under 'Robustness and Evolvability of TFs for regulatory rewiring'.

By its very nature, the rewiring of local connections requires the consideration of at least one other transcriptional unit, but in reality there is the whole genome to acknowledge for the potential of new binding sequences. Hence, in order to obtain a more comprehensive picture of the constraints acting on regulatory network evolution, one has to go beyond isolated units. Rewiring of transcriptional connections will inevitably lead to crosstalk between two TFs and as our experiments indicate, it is highly likely that crosstalk can have unexpected consequences for gene expression through interactions at non-specific sites. As many (local) TFs are present in the cell at low numbers (e.g. LacI is present at ~10 copies per cell), titration due to non-specific sites can significantly affect the availability of a TF for specific regulation. Especially for promiscuous TFs non-specific binding sites can act as traps, hindering the target search[41,42] and decreasing the free concentration of TF available for specific binding[58]. Competition between two TFs at non-specific binding sites can however i) increase free TF concentration, and therefore occupancy at the target promoter as there are less non-specific sites free for binding or ii) it can increase the dissociation rate at non-specific sites due to binding of the other TF in close vicinity. This effect is likely to occur

with TFs sharing similar binding footprints, meaning that they are likely to engage in binding crosstalk. Accordingly, depending on a TFs non-specific binding strength, the amount of crosstalk with other cellular TFs and the level of molecular crowding on DNA, non-specific binding can affect promoter occupancy in opposite ways. The fitness effects of this outcome will again depend on the specific gene in question. Further, regulatory crosstalk can be beneficial for adaptation as it provides plasticity to the regulatory network, although it will be ultimately selected against to avoid interference with the newly evolved, specific regulation[241].

Non-specific interactions with the genome can also have other, indirect consequences on the cell, for example by affecting growth rates. We found a substantial burden on the cell when expressing two gratuitous TF from low copy plasmids. The promiscuous repressor showed a serious growth arrest in a strain that it evolved with as well as in a strain it does not usually encounter. The more specific repressor still affected growth of the non-host strain substantially. This suggests that although the more promiscuous TF has a higher evolutionary potential, there is a trade-off due to unwanted interactions with the rest of the genome leading to potentially high fitness costs. Therefore, a more comprehensive understanding of regulatory network evolution will require combining investigations of specific target site evolution and non-specific background crosstalk.

7.2 Dynamics matter

Ignoring other TFs or non-specific binding could be one of the reasons why previous studies reported a discrepancy between binding affinities determined through *in vitro* methods and gene expression patterns that are observed *in vivo*. Another reason however, could be the importance of transient dynamics (the path to steady state expression) in gene regulation, which has been largely ignored when trying to connect signaling input to promoter output. Many studies have successfully classified specific types of promoter logic and the many ways to implement them[242]. However, knowing the architecture of a regulatory system is generally not enough to determine its steady state gene expression[67], even less the transient behavior until the steady state. It is becoming increasingly clear that transient gene expression is playing an important role in cellular behavior, especially with regard to signaling cascades[60]. As we observed, transient gene expression can last for several hours, especially in slowly growing cells, which increases the likelihood that the environment changes and a new signal reaches the cell within that time. In the extreme scenario where cellular processes rarely reach steady state on relevant timescales, these processes would be optimized and selected for their transient gene expression only. Indeed, we find that variability in transient gene expression patterns of growth-determining genes can result in significant growth differences.

We show that transient gene expression can arise in complex regulatory systems due to delays in the arrival timing of input signals. Especially small delays of a few minutes between signals are very likely to occur regularly in nature, either due to differences

in timing of external signals (for example decreased diffusion, larger distance or slower uptake of one signal) or due to differences in intracellular processes along the signaling chain (for example promoter strength, gene length, protein production and folding). This makes signaling delays and hence transient gene expression patterns a relevant feature for cellular programs and although this variation does not stem from stochasticity, it can provide similar advantages. One classical example is bet-hedging, where phenotypic variability might reduce fitness in the current environment but provide an advantage by anticipating future environmental changes[23]. In addition, phenotypic variability (plasticity) can accelerate evolutionary adaptation as it facilitates the search for a genotype producing a new phenotype that is better suited for a specific environment. Subsequently, selection can stabilize this new phenotype genetically, allowing it to become the new dominant gene expression phenotype[171,243]. This shows that transient dynamics in gene expression can add another layer of behaviors to the phenotype space and should not be lightly ignored in order to obtain a more comprehensive picture of the mapping between promoter architecture and gene expression phenotype.

7.3 Transcription factor interference

TFs, which are binding in close vicinity of each other (either to specific binding sites in a promoter region or non-specifically anywhere along the DNA) can interfere with each other's cis-regulatory activities in various ways: 1) Direct TF competition: The binding of one TF can partially or completely obstruct the binding site of another operator, thereby sterically excluding the other TF from binding[12]. Further, partial dissociation of a bound TF can provide access for another TF to bind nearby, which would then impede rebinding and lead to dissociation of the first TF. This mechanism (facilitated dissociation) can substantially increase off-rates over those measured *in vitro*[14]. 2) Indirect TF interference: TF binding can induce structural changes in DNA or affect DNA supercoiling[16]. Non-specifically bound TFs can also pose obstacles to the 1D-target search along DNA[244]. We found evidence that TF interference significantly impacts gene regulation through direct competition, either locally, by hindering the binding of other TFs in the promoter region, or globally, by affecting free TF concentrations.

The mean number of TF binding sites (in operons with at least one known site) is 3.5, demonstrating the prevalence of co-regulation at bacterial promoters. Most of these sites are separated from their neighbors by less than 15 base pairs and almost half of all binding sites overlap with another one[13]. An elementary example of competition through binding site overlap is given by repressor binding sterically excluding RNAP from the promoter as in the *lac* promoter system. This shows that TF competition seems to be an abundant and important feature in gene regulation, increasing the capacity for complex signal integration drastically[12].

The very nature of TF competition implies a dynamic aspect in this interaction, which cannot be captured through steady state promoter logic, and can lead to an increase

in gene expression variability[245]. We find that TF competition also makes the promoter output sensitive to signal timing, thereby providing another way of encoding information about the environment in the promoter architecture: TFs whose binding sites are too close together to allow for independent binding will produce a memory of the signaling history through their binding dynamics. By moving those sites further apart, the memory will be lost and the phenotypic variability reduced, which would allow for more rapid and precise responses in trade for losing the ability to integrate two inputs based on their timing - which is a property that could provide beneficial information to the cell. In our experiments the causes for the interference are most likely steric hindrance and facilitated dissociation. Steric hindrance will decrease the association rates of the second TF, increasing the occupancy of the promoter by the first TF. Our system exhibits TF competition through binding site overlap between RNAP and the repressor on one hand and the activator and the repressor on the other hand. Moreover, the formation of DNA loops could stabilize the retention of the first regulatory state. However, using previously measured values for looping timescales, showed that stable looping should result in drastically longer memory of the first regulatory state as compared to our experimental results. On the other hand, DNA looping increases the local TF and DNA concentration drastically, which could result in the dissociation of TF forming the loop through facilitated dissociation[14,246], thereby opposing the stabilizing effect. This would also explain why there were no significant differences in memory to a system without DNA looping. We found that interactions between TFs can result in unexpected gene expression dynamics and deserve a closer examination.

7.4 Modeling complexity

Thermodynamic models are very useful and simple models, which capture the biophysics of TF binding quite comprehensively if the molecular states are known and the system is at steady state. Gene expression is approximated through the binding probability of RNAP at the target promoter, which depends on the binding energy of RNAP as well as the TFs binding at the promoter. We found that these simple models provided excellent results when extracting basic mechanistic insights from well-defined systems, where molecular mechanisms and parameters are well known. The model fits the experimental data very well and allowed us to draw conclusions about the importance of each of the parameters on the evolution of the system under question.

Similarly, we used the same model to investigate global crosstalk between two well-characterized repressors. Although the fit is not expected to be very good with the mutants in this system approaching non-specific binding,as the energy penalties in this range have not been biochemically determined, a simplified model provided us with a mechanism that can explain the experimental observations. Even in the extreme case of weak non-specific binding across the genome, we found a surprising amount of correlation between the predicted binding strength from a thermodynamic model

and the number of reads for a specific region from ChIP sequencing (which is indicative of how strongly the region was bound by the TF).

However, in order to describe the dynamics of TF competition at a complex promoter we had to use a differential equation model based on non-equilibrium reaction rates. Coupling the differential equation model with a thermodynamic description of the underlying regulatory dynamics produced a much worse fit as the equilibrium assumption was not fulfilled. TF competition leads to much slower association rates, which are not well separated from the promoter production and protein dilution rates. Moreover, they are at odds with the increased – and therefore very fast – dissociation rates due to facilitated dissociation. Hence, the system cannot be treated as an equilibrium system and the reaction rate model with non-equilibrium transition rates resulted in a much-improved fit. It was however not necessary to include stochastic effects in our model to obtain a satisfactory fit to experimental data.

Overall, we found that the current workhorse for gene regulatory model class - thermodynamic models - is providing important insights into basic mechanisms of regulation and its evolution in simple and well-defined systems. However, more complex promoter systems will require different approaches, especially because we find that it will not always be enough to only consider steady state expression of the system. Transient dynamics seem to be of much more importance to regulatory complexity than currently recognized and only a combination of experimental evidence and appropriate modeling approaches will be able to reveal its exact relevance.

7.5 Model systems

We employed two of the best-characterized repressors, LacI and Lambda CI, to explore new mechanistic concepts, which determine regulatory architecture. This allowed us to combine experimental approaches with well-informed modeling approaches and pre-determined system parameters. Although these synthetic, isolated systems are not reflecting the much more intricate connectivity of natural regulatory systems, they have proven time and again to be a highly advantageous concept in order to elucidate molecular principles of regulatory processes[7,67].

Using phage-based regulatory components for evolutionary studies has the additional advantage that they are related to each other through horizontal gene transfer. Yet, for example Lambda and P22 have diverged over time in two different host species. Hence, it is possible to not only study how those two systems diverged but also their interactions with host and non-host genomes. The latter seems to be a significant constraint in regulatory evolution considering our findings regarding crosstalk and non-specific binding. Hence, a more comprehensive picture needs to encompass how the cellular and genomic background affect regulatory processes.

7.6 New design principles

For many years, particularly spurred by the era of high-throughput and genomic-sequencing approaches, much effort has been put into answering the question of whether there are fundamental design principles of regulatory architectures that would allow us to i) infer their biological function directly from the DNA sequence[67] and ii) predict new network designs. One example of design rules are regulatory structures that are over- or under-represented in comparison to random network design[13,25]. However, the regulatory architecture is still not enough to draw a conclusion as to its function or output. Yet, it seems that there is a finite number of ways to obtain a specific output function[67], giving rise to the hope that we can identify certain constraints that act on network architecture and evolution. One difficulty in elucidating fundamental constraints is that design rules are being studied at different levels and a unifying scheme is needed to obtain a comprehensive design space[247].

Two properties that play an important role in determining network function and adaptation are robustness and evolvability. Many studies have tried to elucidate the factors that make networks and cellular processes robust and /or evolvable as both properties are crucial for cellular fitness, yet often seem to be opposed to one another. To some extent this paradox arises however from the specific definitions of these properties and the scale they are measured at, both of which varies throughout the literature[27,28]. Below we will first discuss a broader concept of robustness and evolvability in gene expression with regard to signal integration, and then more specifically robustness and evolvability of TFs and the implications for regulatory network evolution and function.

Robustness and evolvability of gene expression with regard to the signal input

The simplest unit in regulation is the transcriptional unit, consisting of a gene sequence that is bounded by a promoter region and a terminator[4]. The promoter region encodes the molecular mode of control through the binding sites for RNA polymerase and the binding sites for regulatory proteins (TFs). The arrangement of binding sites for TFs (including RNAP) determines the number of constraints that the system encounters[70]. The more TFs interfere or interact with each other, the harder it becomes to infer the promoter output function from the regulatory architecture alone. What seems like a constraint on the architectural level could however lead to more flexibility on the signaling level.

On the architectural level robustness is conferred by non-overlapping binding sites and no interference between TFs as this makes the binding sites more flexible with regard to changes in their spacing and ordering[70]. The more binding sites overlap with one another, the more constrained their arrangement becomes. This robustness leads at the same time to evolvability in the binding site arrangement, as the spacing and ordering can vary without changing the output of the system significantly. Hence, the

regulatory architecture can explore a neutral genotype space[76] until it reaches genotypes that have a different phenotype.

If we now consider robustness with regard to the dynamics of the signaling input at a promoter, we find that the steady state of a transcriptional unit is robust to changes in timing and even ordering of the signals: regardless if the underlying binding site architecture includes binding site overlap, TF interference, or looping, the system reaches the same steady state for signals that have long or short delays between them, or arrive in different succession. This appears to be an extremely relevant type of robustness in natural systems as signaling delays can easily arise due to external or internal factors (e.g. slow diffusion or protein production).

Considering evolvability with respect to signal timing, there is a clear difference between promoter systems with and without binding site overlap or any form of TF interference. While systems without interference between TFs require genetic changes to exhibit new phenotypes that can be selected on, systems with interference will display phenotypic variability in transient expression dynamics. These transient dynamics can last for several cell generations, enabling appreciable phenotypic variability (plasticity), which selection can act on, thereby increasing evolvability[171,243]. Moreover, transient dynamics due to TF interference arise from history-dependence, which allows for selection to act not only on the current phenotype but on the current phenotype given a certain history. This can be useful in fluctuating environments where transitions are correlated[248].

While promoter architectures that allow for TF interference seem to be more constrained in their design than ones with independent TF binding, considering the dynamics of the input function as the selective feature reveals a benefit of TF interference under certain conditions. For systems that are required to reach their steady state fast and which have a strong constraint on the amount of gene expression, TF interference would be disadvantageous[172]. However, many transcriptional units might not reach steady state on timescales that are relevant for the cell[59,60] and, hence, selection will frequently act on transient expression dynamics. TF interference allows for encoding of information of the signaling history reliably in transient expression dynamics, which – as opposed to stochastic gene expression phenotypes – allows for phenotypic variability that is informed by the history of the cell, possibly facilitating cellular physiological adaptation. Hence, systems with TF interference are evolvable in that selection can act on an advantageous transient phenotypes (produced by certain signal dynamics), while still retaining robustness with regard to the steady state. In this way, complex and - to some extent - variable gene expression patterns can be encoded in regulatory architectures without changing the components but only by strengthening or weakening specific interactions. Similarly, even though TF interference makes the system more sensitive to changes in binding site spacing and ordering, those changes can easily be compensated for by tuning the binding affinity or concentrations of TFs in order to obtain a certain steady state expression level.

Robustness and evolvability of TFs for regulatory rewiring

As gene expression patterns and promoters adapt over time, they will not only change the properties of their regulatory components, but also the components themselves. These components – TFs and their binding sites – connect transcriptional units to form gene regulatory networks controlling cellular behavior, hence understanding the constraints on these components in changing their connections over time is crucial to understanding cellular evolution. Two properties crucially describing these constraints for TFs are robustness and evolvability.

"Proteins thus give DNA its functions – its personality – in cells, turning a DNA molecule from just naked DNA, into a chromosome" (Marko, 2015)[249].

Indeed, we find that the potential for rewiring individual network connections, as characterized by robustness and evolvability, is determined by inherent biophysical characteristics of the TF, not the binding site - even though the binding site is likely the component changed by mutations due to higher constraint on TF mutations[87]. This reduces the number of cellular components dictating the capacity for rewiring by one order of magnitude from ~2600 promoter regions (~2000 of them regulated by TFs)[13] to ~270 TFs that have to be considered[250]. We did not find a general number of mutations or base pair matches that is necessary to enable binding of a TF to a certain sequence. Rather, if certain biophysical characteristics of a TF are known (most importantly the wild type binding energy and the average energy per mutation), it is possible to infer the average number of mutations necessary in order to gain a specific amount of repression using their sigmoidal relationship. As this inference is only based on TF binding to one operator (without knowledge of the specific DNA sequence), no details about the regulatory system are necessary to obtain this number.

If a larger number of TFs are classified for their robustness and evolvability it might become possible to classify them into certain categories according to their evolutionary potential – which will be different from the current definition of TF families as Lambda CI and P22 C2 belong to the same TF family, but showed opposing evolutionary potential. This new evolutionary classification might be tied to the way a TF recognizes its DNA sequence as binding energy is one of the most important determinants for rewiring, and Lambda CI and P22 C2 are known to differ in their recognition method: the former uses direct hydrogen bonding, while the latter is also reliant on structural DNA information. This might indicate two different modes of rewiring for these two classes: whereas one class can gain access to new binding sites if only very few mutations accumulate in a non-cognate site, the other class is more likely to be reliant on duplication of its binding site in order to gain new regulatory connections. Overall, robustness and evolvability of individual system components (TFs) determine the constraints that are placed on global network evolution through few TF-intrinsic characteristics.

Robustness and evolvability of cellular programs with regard to genomic context

The focus in designing regulatory architectures and understanding their evolution is generally placed on specific regulatory interactions. However, the cytoplasm is crowded with proteins and between 10-50% of DNA is bound by non-cognate proteins[41]. Therefore, in addition to specific regulatory requirements, non-specific binding along the DNA is likely to put constraints on regulatory function and architecture.

Non-specific binding can affect regulation at target promoters in several ways: Non-specifically bound proteins can constitute obstacles on DNA, thereby limiting 1D-diffusion[41,42], or trapping bound TFs[160]. This can impose certain structural architectures in order to circumvent this problem: for example, DNA loops formed by a single protein complex (e.g. LacI looping)[42] or the co-localization of TFs and the genes they regulate[251]. Moreover, the need to find a specific binding site within a sea of non-specific sites (including ones that might only be one or two base pair mutations away), leads to specific constraints on the binding energies of TFs - which are in good agreement with experimentally obtained values[24,85]: the binding energy per nucleotide is 2-3 k_BT and the difference between non-specific binding and the best binding sequence is 15-16 k_BT. The latter is a trade-off between the constraints of getting trapped in non-specific sites and being 'programmable' to new sites[85].

Furthermore, the avoidance of non-specific binding leads to a constraint on the number of TFs per cell[42]. This constraint can be imposed by overall binding of TFs and impeding of cellular processes, but also by TFs binding to non-cognate binding sites. The latter can impose crosstalk between a cognate and a non-cognate TF, leading to miss-regulation of genes and therefore to expression of genes at the wrong time or place, or in the wrong amounts. This can occur at much lower TF concentrations than the random non-specific binding across the genome, as the binding difference between these non-cognate sequences and the cognate sequence is likely not as high. Hence, one constraint will be acting on the total number of DNA binding proteins in the cell, but another one will be affecting individual TFs, depending on their proclivity to bind DNA and the presence of related binding sites. Similarly, the acquisition of horizontally transferred genes, plasmids or prophages can prove to be toxic to the cell due to non-specific binding and interference with the cellular program[216]. Therefore, cells have to balance the possible benefit of obtaining a useful gene versus the danger of obtaining a toxic protein and allow or limit horizontal gene transfer accordingly.

We found that the fitness cost for the cell also depends on environmental conditions, as non-specific binding was increasingly detrimental at slower growth. This adds another layer of complexity to the deduction of evolutionary constraints on regulatory evolution as they might differ not only due to specific environmental effects on the regulatory network under investigation but also due to environmental effects on global cellular parameters and growth. The slow-growth-scenario seemed to be most stringent in our study, yet seems to be a realistic condition, as most bacteria tend to grow at a doubling time from 1-100h (with a median of ~7h) in nature[252]. Moreover,

poor environments or stress conditions often induce the uptake of foreign DNA[253], leading to cells with slow growth having to cope with foreign TFs. The evolutionary constraints that arise in nature might therefore be quite different from the ones that arise under fast growth in usual lab conditions.

Although non-specific binding might increase the potential of TFs to acquire new binding sites, overall non-specific binding seems to put strong constraints on the regulatory architectures of the cell by limiting i) the number of TFs, ii) horizontal transfer of genes, iii) binding energies, iv) target search by TFs, and iv) DNA conformation. Consequently, more promiscuous binders are more likely to disrupt cellular programs and constrain adaptation. This is in direct opposition to what we found at specific binding sites, where TF robustness and evolvability of promiscuous binders leads to preservation or gain of promoter function. Hence, TF robustness and evolvability seem to experience a trade-off between specific and non-specific sites.

7.7 Summary

As a central feature of living organisms, gene regulation has been the focus of research for many decades. Even though we have learned much about the molecular mechanisms of TF regulation at promoters, the link between those mechanisms and other features that constrain gene regulatory design, like regulatory evolution and TF dynamics, have remained mysterious.

The search for general design principles in regulatory architecture has focused more on the occurrence of specific advantageous patterns, than the molecular rules that delimit the possible parameter space[24]. Nevertheless, the biophysical characteristics of TF binding can provide us with fundamental rules and trade-offs that are acting on the architecture of regulatory networks. Properties that might seem beneficial if considering a specific regulatory function might be in trade-off with spurious functions arising from the cellular background. Furthermore, the surprising complexity of signal integration at many promoters makes it necessary to consider not only TF binding, but also the biophysics of TF-TF interactions to understand the resulting regulatory dynamics.

Robustness and evolvability have been employed previously as tools to determine global design rules, yet their implications depend on the regulatory level that is studied and the specific regulatory feature it is applied to. For example, higher robustness and evolvability of a TF facilitate network evolution, but can pose a considerable cost on cellular growth. Nevertheless, those properties can be helpful in determining general constraints on cellular networks by finding these trade-offs and the network features governing fitness.

Overall, considering the most fundamental level of gene regulation, namely the molecular interactions of individual regulatory components, can provide a thorough understanding that can then be employed in studies of higher-level functions, such as resistance evolution and complex cellular group behaviors.

References

1. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3,** 318–356 (1961).
2. Ptashne, M. *et al.* How the lambda repressor and cro work. *Cell* **19,** 1–11 (1980).
3. Lwoff, A. Interaction among virus, cell, and organism. *Science* **152,** 1216–20 (1966).
4. Savageau, M. a. Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos* **11,** 142–159 (2001).
5. Hirsh, J. & Schleif, R. The araC promoter: Transcription, mapping and interaction with the araBAD promoter. *Cell* **11,** 545–550 (1977).
6. Balleza, E. *et al.* Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol. Rev.* **33,** 133–151 (2009).
7. Friedman, D. I. & Court, D. L. Bacteriophage lambda: alive and well and still doing its thing. *Curr. Opin. Microbiol.* **4,** 201–7 (2001).
8. Hammar, P. *et al.* Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nat. Genet.* **46,** 405–8 (2014).
9. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14,** 638–650 (2016).
10. Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B. & Müller-Hill, B. Quality and position of the three lac operators of E. coli define efficiency of repression. *EMBO J.* **13,** 3348–55 (1994).
11. Cui, L., Murchland, I., Shearwin, K. E. & Dodd, I. B. Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2922–7 (2013).
12. Hermsen, R., Tans, S. & Ten Wolde, P. R. Transcriptional regulation by competing transcription factor modules. *PLoS Comput. Biol.* **2,** 1552–1560 (2006).
13. Rydenfelt, M., Garcia, H. G., Cox, R. S. & Phillips, R. The Influence of Promoter Architectures and Regulatory Motifs on Gene Expression in Escherichia coli. *PLoS One* **9,** e114347 (2014).
14. Kamar, R. I. *et al.* Facilitated dissociation of transcription factors from single DNA binding sites. *Proc. Natl. Acad. Sci. U. S. A.* **114,** E3251–E3257 (2017).
15. Ding, Y. *et al.* DNA supercoiling: a regulatory signal for the λ repressor. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 15402–7 (2014).
16. Yeung, E. *et al.* Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Syst.* **5,** 11–24.e12 (2017).
17. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference - A crash course. *Trends Genet.* **21,** 339–345 (2005).
18. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309,** 2010–3 (2005).
19. Chalancon, G. *et al.* Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* **28,** 221–232 (2012).
20. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135,** 216–226 (2008).
21. Paulsson, J. Summing up the noise in gene networks. *Nature* **427,** 415–8 (2004).
22. Tkačik, G., Walczak, A. M. & Bialek, W. Optimizing information flow in small genetic networks. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **80,** 1–18 (2009).
23. Fraser, D. & Kaern, M. A chance at survival: gene expression noise and phenotypic

diversification strategies. *Mol. Microbiol.* **71,** 1333–40 (2009).
24. Lässig, M. From biophysics to evolutionary genetics: statistical aspects of gene regulation. *BMC Bioinformatics* **8,** S7 (2007).
25. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31,** 64–8 (2002).
26. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296,** 1466–70 (2002).
27. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J. & Doyle, J. Robustness of cellular functions. *Cell* **118,** 675–85 (2004).
28. Pigliucci, M. Is evolvability evolvable? *Nat Rev Genet* **9,** 75–82 (2008).
29. Ancel, L. W. & Fontana, W. Plasticity, Evolability and Modularity in RNA. *J. Exp. Zool.* **288,** 242–283 (2000).
30. Whitacre, J. M. Degeneracy: a link between evolvability, robustness and complexity in biological systems. *Theor. Biol. Med. Model.* **7,** 6 (2010).
31. Payne, J. L. & Wagner, A. Mechanisms of mutational robustness in transcriptional regulation. *Front. Genet.* **6,** 1–10 (2015).
32. Wagner, A. Robustness, evolvability, and neutrality. *FEBS Lett.* **579,** 1772–1778 (2005).
33. Li, H. & Johnson, A. D. Evolution of transcription networks-lessons from yeasts. *Curr. Biol.* **20,** R746–R753 (2010).
34. Lozada-Chávez, I., Janga, S. C. & Collado-Vides, J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res.* **34,** 3434–3445 (2006).
35. Isalan, M. *et al.* Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452,** 840–845 (2008).
36. Nocedal, I., Mancera, E. & Johnson, A. D. Gene regulatory network plasticity predates a switch in function of a conserved transcription regulator. 1–20 (2017). doi:10.7554/eLife.23250
37. Pougach, K. *et al.* Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nat. Commun.* **5,** 1–11 (2014).
38. Madan Babu, M., Teichmann, S. a & Aravind, L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J. Mol. Biol.* **358,** 614–33 (2006).
39. Nocedal, I. & Johnson, A. D. How Transcription Networks Evolve and Produce Biological Novelty. *Cold Spring Harb. Symp. Quant. Biol.* **80,** 265–274 (2015).
40. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* **139,** 1366–1375 (2009).
41. Flyvbjerg, H., Keatch, S. A. & Dryden, D. T. F. Strong physical constraints on sequence-specific target location by proteins on DNA molecules. *Nucleic Acids Res.* **34,** 2550–2557 (2006).
42. Li, G.-W., Berg, O. G. & Elf, J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* **5,** 294–297 (2009).
43. Dey, P. & Bhattacherjee, A. Role of Macromolecular Crowding on the Intracellular Diffusion of DNA Binding Proteins. *Sci. Rep.* **8,** 1–11 (2018).
44. Mirny, L. *et al.* How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A Math. Theor.* **42,** 434013 (2009).
45. Zabet, N. R. & Adryan, B. The effects of transcription factor competition on gene regulation. *Front. Genet.* **4,** 1–10 (2013).
46. Morelli, M. J., Allen, R. J. & Rein Ten Wolde, P. Effects of macromolecular crowding on genetic networks. *Biophys. J.* **101,** 2882–2891 (2011).
47. Hao, N. *et al.* Road rules for traffic on DNA - Systematic analysis of transcriptional roadblocking in vivo. *Nucleic Acids Res.* **42,** 8861–8872 (2014).

48. Bremer, H. & Dennis, P. P. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* **3,** (2008).
49. Nielsen, H. J., Youngren, B., Hansen, F. G. & Austin, S. Dynamics of Escherichia coli Chromosome Segregation during Multifork Replication. *J. Bacteriol.* **189,** 8660–8666 (2007).
50. Nordström, K. & Dasgupta, S. Copy-number control of the Escherichia coli chromosome: A plasmidologist's view. *EMBO Rep.* **7,** 484–489 (2006).
51. Shea, M. A. & Ackers, G. K. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J. Mol. Biol.* **181,** 211–30 (1985).
52. von Hippel, P. H., Revzin, A., Gross, C. A. & Wang, A. C. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. *Proc. Natl. Acad. Sci. U. S. A.* **71,** 4808–12 (1974).
53. Schlitt, T. & Brazma, A. Current approaches to gene regulatory network modelling. *BMC Bioinformatics* **8 Suppl 6,** S9 (2007).
54. Ay, A. & Arnosti, D. N. Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit. Rev. Biochem Mol Biol.* **46,** 137–151 (2011).
55. Albert, R. & Othmer, H. G. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J. Theor. Biol.* **223,** 1–18 (2003).
56. Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A. & Alon, U. Plasticity of the cis-regulatory input function of a gene. *PLoS Biol.* **4,** e45 (2006).
57. Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* **15,** 116–24 (2005).
58. Bakk, A. & Metzler, R. In vivo non-specific binding of λ CI and Cro repressors is significant. *FEBS Lett.* **563,** 66–68 (2004).
59. Braun, E. & Brenner, N. Transient responses and adaptation to steady state in a eukaryotic gene regulation system. *Phys. Biol.* **1,** 67–76 (2004).
60. Bolouri, H. & Davidson, E. H. Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics. *Proc. Natl. Acad. Sci.* **100,** 9371–9376 (2003).
61. Joh, R. I. & Weitz, J. S. To lyse or not to lyse: transient-mediated stochastic fate determination in cells infected by bacteriophages. *PLoS Comput. Biol.* **7,** e1002006 (2011).
62. Verd, B., Crombach, A. & Jaeger, J. Classification of transient behaviours in a time-dependent toggle switch model. *BMC Syst. Biol.* **8,** 43 (2014).
63. Santillán, M. & Mackey, M. C. Why the lysogenic state of phage lambda is so stable: a mathematical modeling approach. *Biophys. J.* **86,** 75–84 (2004).
64. Santillán, M., Mackey, M. C. & Zeron, E. S. Origin of bistability in the lac Operon. *Biophys. J.* **92,** 3830–42 (2007).
65. Garcia, H. G., Sanchez, A., Kuhlman, T., Kondev, J. & Phillips, R. Transcription by the numbers redux: experiments and calculations that surprise. *Trends Cell Biol.* **20,** 723–733 (2010).
66. Tkačik, G. & Walczak, A. M. Information transmission in genetic regulatory networks: a review. *J. Phys. Condens. Matter* **23,** 153102 (2011).
67. Lim, W. A., Lee, C. M. & Tang, C. Design Principles of Regulatory Networks: Searching for the Molecular Algorithms of the Cell. *Mol. Cell* **49,** 202–212 (2013).
68. Hormoz, S. Cross talk and interference enhance information capacity of a signaling pathway. *Biophys. J.* **104,** 1170–1180 (2013).
69. Sanchez, A., Garcia, H. G., Jones, D., Phillips, R. & Kondev, J. Effect of Promoter

Architecture on the Cell-to-Cell Variability in Gene Expression. *PLoS Comput. Biol.* **7,** e1001100 (2011).
70. Levo, M. & Segal, E. In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.* **15,** 453–468 (2014).
71. Tyson, J. J. & Novák, B. Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem.* **61,** 219–40 (2010).
72. Sneppen, K., Krishna, S. & Semsey, S. Simplified Models of Biological Networks. (2010). doi:10.1146/annurev.biophys.093008.131241
73. Englesberg, E., Irr, J., Power, J. & Lee, N. Positive control of enzyme synthesis by gene C in the L-arabinose system. *J. Bacteriol.* **90,** 946–57 (1965).
74. Ciliberti, S., Martin, O. C. & Wagner, a. Innovation and robustness in complex regulatory gene networks. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 13591–6 (2007).
75. Steinacher, A., Bates, D. G., Akman, O. E. & Soyer, O. S. Nonlinear Dynamics in Gene Regulation Promote Robustness and Evolvability of Gene Expression Levels. *PLoS One* **11,** e0153295 (2007).
76. Payne, J. L. & Wagner, A. The Robustness and Evolvability of Transcription Factor Binding Sites. *Science (80-.).* **343,** 875–877 (2014).
77. Tuğrul, M., Paixão, T., Barton, N. H. & Tkačik, G. Dynamics of Transcription Factor Binding Site Evolution. *PLOS Genet.* **11,** e1005639 (2015).
78. Prud'homme, B., Gompel, N. & Carroll, S. B. Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci.* **104,** 8605–8612 (2007).
79. Ward, J. J. & Thornton, J. M. Evolutionary models for formation of network motifs and modularity in the Saccharomyces transcription factor network. *PLoS Comput. Biol.* **3,** 1993–2002 (2007).
80. Tuch, B. B., Li, H. & Johnson, A. D. Evolution of eukaryotic transcription circuits. *Science (80-.).* **319,** 1797–1799 (2008).
81. Maerkl, S. J. & Quake, S. R. Experimental determination of the evolvability of a transcription factor. (2009).
82. Sayou, C., Monniaux, M., Nanao, M. H. & Moyroud, E. Promiscuous Intermediate Underlies the Evolution of LEAFY DNA Binding Specificity. 645–649 (2014).
83. Wagner, G. P. & Lynch, V. J. The gene regulatory logic of transcription factor evolution. *Trends Ecol. Evol.* **23,** 377–385 (2008).
84. von Hippel, P. H. & Berg, O. G. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 1608–12 (1986).
85. Gerland, U., Moroz, J. D. & Hwa, T. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12015–12020 (2002).
86. Mustonen, V., Kinney, J., Callan, C. G. & La, M. Energy-dependent fitness : A quantitative model for the evolution of yeast transcription factor binding sites. **105,** 12376–12381 (2008).
87. Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. *Protein Sci.* **25,** 1204–1218 (2016).
88. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8,** 610–618 (2007).
89. Carroll, S. B. Evolution at Two Levels: On Genes and Form. *PLoS Biol.* **3,** e245 (2005).
90. Ludwig, M. Z. Functional evolution of noncoding DNA. *Curr. Opin. Genet. Dev.* **12,** 634–639 (2002).
91. Moses, A. M., Chiang, D. Y., Pollard, D. A., Iyer, V. N. & Eisen, M. B. MONKEY: identifying conserved transcription-factor binding sites in multiple alignments

using a binding site-specific evolutionary model. *Genome Biol.* **5,** R98 (2004).
92. Berg, J., Willmann, S. & Lässig, M. Adaptive evolution of transcription factor binding sites. *BMC Evol Biol* **4,** 42 (2004).
93. Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13,** 59–69 (2012).
94. Pujato, M., MacCarthy, T., Fiser, A. & Bergman, A. The Underlying Molecular and Network Level Mechanisms in the Evolution of Robustness in Gene Regulatory Networks. *PLoS Comput. Biol.* **9,** (2013).
95. Sauer, R. T. *et al.* The lambda and P22 phage repressors. *J. Biomol. Struct. Dyn.* **1,** 1011–22 (1983).
96. Susskind, M. M. & Botstein, D. Molecular genetics of bacteriophage P22. *Microbiol. Rev.* **42,** 385–413 (1978).
97. Sarai, A. & Takeda, Y. Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 6513–7 (1989).
98. Hilchey, S. P., Wu, L. & Koudelka, G. B. Recognition of Nonconserved Bases in the P22 Operator by P22 Repressor Requires Specific Interactions between Repressor and Conserved Bases *. **272,** 19898–19905 (1997).
99. Lutz, R. & Bujard, H. 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,** 1203–10 (1997).
100. Degnan, P. H., Michalowski, C. B., Babić, A. C., Cordes, M. H. J. & Little, J. W. Conservation and diversity in the immunity regions of wild phages with the immunity specificity of phage lambda. *Mol. Microbiol.* **64,** 232–44 (2007).
101. Maerkl, S. J. & Quake, S. R. A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science (80-.).* **315,** 233–237 (2007).
102. Zhao, Y., Ruan, S., Pandey, M. & Stormo, G. D. Improved models for transcription factor binding site identification using nonindependent interactions. *Genetics* **191,** 781–790 (2012).
103. Weirauch, M. T. *et al.* Evaluation of methods for modeling transcription factor sequence specificity. *Nat. Biotechnol.* **31,** 126–134 (2013).
104. Klumpp, S. & Hwa, T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 20245–50 (2008).
105. Razo-Mejia, M. *et al.* Comparison of the theoretical and real-world evolutionary potential of a genetic circuit. *Phys. Biol.* **11,** 026005 (2014).
106. Lagator, M., Paixão, T., Barton, N. H., Bollback, J. P. & Guet, C. C. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *Elife* **6,** 1–16 (2017).
107. Kreamer, N. N., Phillips, R., Newman, D. K. & Boedicker, J. Q. Predicting the impact of promoter variability on regulatory outputs. *Sci. Rep.* **5,** 18238 (2016).
108. Luscombe, N. M. & Thornton, J. M. Protein-DNA interactions: Amino acid conservation and the effects of mutations on binding specificity. *J. Mol. Biol.* **320,** 991–1009 (2002).
109. Watkins, D., Hsiao, C., Woods, K. K., Koudelka, G. B. & Williams, L. D. P22 c2 repressor-operator complex: Mechanisms of direct and indirect readout. *Biochemistry* **47,** 2325–2338 (2008).
110. Gertz, J., Gerke, J. P. & Cohen, B. A. Epistasis in a quantitative trait captured by a molecular model of transcription factor interactions. *Theor. Popul. Biol.* **77,** 1–5 (2010).
111. Stormo, G. D. & Zhao, Y. Determining the specificity of protein–DNA interactions.

Nat. Rev. Genet. **11,** 751–760 (2010).
112. Draghi, J. A., Parsons, T. L., Wagner, G. P. & Plotkin, J. B. Mutational robustness can facilitate adaptation. *Nature* **463,** 353–5 (2010).
113. Wagner, A. The role of robustness in phenotypic adaptation and innovation. *Proc. R. Soc. B Biol. Sci.* **279,** 1249–1258 (2012).
114. Fattah, K. R., Mizutani, S., Fattah, F. J., Matsushiro, A. & Sugino, Y. A comparative study of the immunity region of lambdoid phages including Shiga-toxin-converting phages : molecular basis for cross immunity . 223–232 (2000).
115. Friedlander, T., Prizak, R., Guet, C. C., Barton, N. H. & Tkačik, G. Intrinsic limits to gene regulation by global crosstalk. *Nat. Commun.* **7,** 12307 (2016).
116. Duque, T. *et al.* Simulations of enhancer evolution provide mechanistic insights into gene regulation. *Mol. Biol. Evol.* **31,** 184–200 (2014).
117. Kinney, J. B., Murugan, A., Callan, C. G. & Cox, E. C. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9158–63 (2010).
118. Brunner, M. & Bujard, H. Promoter recognition and promoter strength in the Escherichia coli system. *EMBO J.* **6,** 3139–44 (1987).
119. Vilar, J. M. G. Accurate prediction of gene expression by integration of DNA sequence statistics with detailed modeling of transcription regulation. *Biophys. J.* **99,** 2408–13 (2010).
120. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20,** 87–90 (2002).
121. Ptashne, M. *A Genetic Switch: Phage Lambda Revisited*. (Cold Spring Harbor Laboratory Press, 2004).
122. Datsenko, K. a & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6640–5 (2000).
123. Koblan, K. S. & Ackers, G. K. Energetics of subunit dimerization in bacteriophage lambda cI repressor: linkage to protons, temperature, and KCl. *Biochemistry* **30,** 7817–21 (1991).
124. Yamane, T. & Hopfield, J. J. Experimental evidence for kinetic proofreading in the aminoacylation of tRNA by synthetase. *Proc. Natl. Acad. Sci.* **74,** 2246–2250 (1977).
125. Mckeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. **92,** 5042–5046 (1995).
126. Johnson, M. E. & Hummer, G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. *Proc. Natl. Acad. Sci.* **108,** 603–608 (2011).
127. Rowland, M. A. & Deeds, E. J. Crosstalk and the evolution of specificity in two-component signaling. *Proc. Natl. Acad. Sci.* **111,** 5550–5555 (2014).
128. Mora, T. Physical limit to concentration sensing amid spurious ligands. 1–9 (2018).
129. Kao-Huang, Y. *et al.* Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 4228–32 (1977).
130. Shultzaberger, R. K., Maerkl, S. J., Kirsch, J. F. & Eisen, M. B. Probing the informational and regulatory plasticity of a transcription factor DNA-binding domain. *PLoS Genet.* **8,** (2012).
131. Yang, J.-R., Liao, B.-Y., Zhuang, S.-M. & Zhang, J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc. Natl. Acad. Sci.* **109,** E831–E840 (2012).

132. Vavouri, T., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. *Cell* **138,** 198–208 (2009).
133. Todeschini, A. L., Georges, A. & Veitia, R. A. Transcription factors: Specific DNA binding and specific gene regulation. *Trends Genet.* **30,** 211–219 (2014).
134. Whipple, F. W., Kuldell, N. H., Cheatham, L. a & Hochschild, a. Specificity determinants for the interaction of lambda repressor and P22 repressor dimers. *Genes Dev.* **8,** 1212–1223 (1994).
135. Igler, C., Lagator, M., Tkačik, G., Bollback, J. P. & Guet, C. C. Evolutionary potential of transcription factors for gene regulatory rewiring. *Nat. Ecol. Evol.* **2,** (2018).
136. Paramanathan, T., Reeves, D., Friedman, L. J., Kondev, J. & Gelles, J. A general mechanism for competitor-induced dissociation of molecular complexes. *Nat. Commun.* **5,** 5207 (2014).
137. Aguilar-Rodríguez, J., Payne, J. L. & Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol. Evol.* **1,** 0045 (2017).
138. Gama-Castro, S. *et al.* RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.* **44,** D133-43 (2016).
139. Lewis, D. E. A., Gussin, G. N. & Adhya, S. New Insights into the Phage Genetic Switch: Effects of Bacteriophage Lambda Operator Mutations on DNA Looping and Regulation of PR, PL, and PRM. *J. Mol. Biol.* **428,** 4438–4456 (2016).
140. Aguilar-Rodríguez, J., Peel, L., Stella, M., Wagner, A. & Payne, J. L. The architecture of an empirical genotype-phenotype map. *Evolution (N. Y).* **72,** 1242–1260 (2018).
141. Borneman, A. R. *et al.* Divergence of Transcription Factor Binding Sites Across Related Yeast Species. *Science (80-.).* **317,** 815–819 (2007).
142. Tsai, Z. T.-Y. *et al.* Evolution of cis-regulatory elements in yeast de novo and duplicated new genes. *BMC Genomics* **13,** 717 (2012).
143. Pérez JC *et al.* How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes. *Genes Dev.* **28,** 1272–1277 (2014).
144. Oren, Y. *et al.* Transfer of noncoding DNA drives regulatory rewiring in bacteria. *Proc. Natl. Acad. Sci.* **111,** 16112–16117 (2014).
145. Botstein, D. A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* **354,** 484–491 (1980).
146. Benson, N. R. & Goldman, B. S. Rapid mapping in Salmonella typhimurium with Mud-P22 prophages. *J. Bacteriol.* **174,** 1673–81 (1992).
147. Barnard, A., Wolfe, A. & Busby, S. Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr. Opin. Microbiol.* **7,** 102–8 (2004).
148. Gama-Castro, S. *et al.* RegulonDB version 7.0: Transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.* **39,** 98–105 (2011).
149. Cortes, M. G., Trinh, J. T., Zeng, L. & Balázsi, G. Late-Arriving Signals Contribute Less to Cell-Fate Decisions. *Biophys. J.* **113,** 2110–2120 (2017).
150. Casadesús, J. & D'Ari, R. Memory in bacteria and phage. *BioEssays* **24,** 512–518 (2002).
151. Veening, J., Smits, W. K. & Kuipers, O. P. Bistability, Epigenetics, and Bet-Hedging in Bacteria. *Annu. Rev. Microbiol.* **62,** 193–210 (2008).
152. Joh, R. I. & Weitz, J. S. To Lyse or not to Lyse: Transient-mediated stochastic fate determination in cells infected by bacteriophages. *PLoS Comput. Biol.* **7,** (2011).
153. Lambert, G., Kussell, E. & Kussel, E. Memory and fitness optimization of bacteria

under fluctuating environments. *PLoS Genet.* **10,** e1004556 (2014).
154. Ezer, D., Zabet, N. R. & Adryan, B. Physical constraints determine the logic of bacterial promoter architectures. *Nucleic Acids Res.* **42,** 4196–4207 (2014).
155. Vilar, J. M. G. & Leibler, S. DNA Looping and Physical Constraints on Transcription Regulation. *J. Mol. Biol.* **331,** 981–989 (2003).
156. Révet, B., Von Wilcken-Bergmann, B., Bessert, H., Barker, A. & Müller-Hill, B. Four dimers of λ repressor bound to two suitably spaced pairs of λ operators form octamers and DNA loops over large distances. *Curr. Biol.* **9,** 151–154 (1999).
157. Anderson, L. M. & Yang, H. DNA looping can enhance lysogenic CI transcription in phage lambda. **105,** (2008).
158. Lewis, D., Le, P., Zurla, C., Finzi, L. & Adhya, S. Multilevel autoregulation of λ repressor protein CI by DNA looping in vitro. (2011). doi:10.1073/pnas.1111221108
159. Priest, D. G. *et al.* Quantitation of interactions between two DNA loops demonstrates loop domain insulation in E. coli cells. *Proc. Natl. Acad. Sci.* **111,** E4449–E4457 (2014).
160. Gerland, U., Moroz, J. D. & Hwa, T. Physical constraints and functional characteristics of transcription factor – DNA interaction. **2002,** (2002).
161. Hammar, P. *et al.* The lac Repressor Displays Facilitated Diffusion in Living Cells. *Science (80-.).* **336,** 1595–1598 (2012).
162. Bintu, L. *et al.* Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* **15,** 125–135 (2005).
163. Segal, E. & Widom, J. From DNA sequence to transcriptional behaviour : a quantitative approach. *Nat. Publ. Gr.* **10,** 443–456 (2009).
164. Wolf, D. M. *et al.* Memory in Microbes: Quantifying History-Dependent Behavior in a Bacterium. *PLoS One* **3,** e1700 (2008).
165. Gefen, O. & Balaban, N. Q. The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress. *FEMS Microbiol. Rev.* **33,** 704–17 (2009).
166. Hillen, W. & Berens, C. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu. Rev. Microbiol.* **48,** 345–69 (1994).
167. Betz, J. L., Sasmor, H. M., Buck, F., Insley, M. Y. & Caruthers, M. H. Base substitution mutants of the lac operator: in vivo and in vitro affinities for lac repressor. *Gene* **50,** 123–32 (1986).
168. Mitchell, A. *et al.* Adaptive prediction of environmental changes by microorganisms. *Nature* **460,** 220–224 (2009).
169. Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Memory and modularity in cell-fate decision making. *Nature* **503,** 481–486 (2013).
170. Mathis, R. & Ackermann, M. Response of single bacterial cells to stress gives rise to complex history dependence at the population level. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 4224–9 (2016).
171. Espinosa-Soto, C., Martin, O. C. & Wagner, A. Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits. *BMC Evol Biol* **11,** 5 (2011).
172. Roemhild, R. *et al.* Cellular hysteresis as a principle to maximize the efficacy of antibiotic therapy. *Proc. Natl. Acad. Sci.* **115,** 9767–9772 (2018).
173. Kim, S. *et al.* Probing Allostery Through DNA. *Science (80-.).* **339,** 816–819 (2013).
174. Han, L. *et al.* Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS One* **4,** e5621 (2009).
175. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J.*

Bacteriol. **177,** 4121–30 (1995).
176. Court, D. L., Sawitzke, J. A. & Thomason, L. C. Genetic Engineering using Homologous Recombination. *Ann. Rev. Gene.* **36,** 361–388 (2002).
177. Haldimann, A. & Wanner, B. L. Conditional-Replication, Integration, Excision, and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria. *J. Bacteriol.* **183,** 6384–6393 (2001).
178. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12,** 1338–47 (1998).
179. Le, T. T., Emonet, T., Harlepp, S., Guet, C. C. & Cluzel, P. Dynamical Determinants of Drug-Inducible Gene Expression in a Single Bacterium. *Biophys. J.* **90,** 3315–3321 (2006).
180. Johnson, C. M. & Schleif, R. F. In vivo induction kinetics of the arabinose promoters in Escherichia coli. *J. Bacteriol.* **177,** 3438–42 (1995).
181. Parma, D. H. *et al.* The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes Dev.* **6,** 497–510 (1992).
182. Dodd, I. B. *et al.* Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev.* **18,** 344–54 (2004).
183. Le, T. T., Emonet, T., Harlepp, S., Guet, C. C. & Cluzel, P. Dynamical Determinants of Drug-Inducible Gene Expression in a Single Bacterium. *Biophys. J.* **90,** 3315–3321 (2006).
184. Leveau, J. H. J. & Lindow, S. E. Predictive and Interpretive Simulation of Green Fluorescent Protein Expression in Reporter Bacteria. *J. Bacteriol.* **183,** 6752–6762 (2001).
185. Guay, G. g & Rothstein, D. m. Expression of the tetK gene from Staphylococcus aureus in Escherichia coli: comparison of substrate specificities of TetA(B), TetA(C), and TetK efflux proteins. *Antimicrob. Agents Chemother.* **37,** 191–198 (1993).
186. Ninio, J. Alternative to the steady-state method: derivation of reaction rates from first-passage times and pathway probabilities. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 663–7 (1987).
187. Arkin, A. & Ross, J. Computational functions in biochemical reaction networks. *Biophys. J.* **67,** 560–78 (1994).
188. de Jong, H., Ranquet, C., Ropers, D., Pinel, C. & Geiselmann, J. Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst. Biol.* **4,** 55 (2010).
189. Tsai, M.-Y., Zhang, B., Zheng, W. & Wolynes, P. G. Molecular Mechanism of Facilitated Dissociation of Fis Protein from DNA. *J. Am. Chem. Soc.* **138,** 13497–13500 (2016).
190. Becker, N. A., Greiner, A. M., Peters, J. P. & Maher, L. J. Bacterial promoter repression by DNA looping without protein-protein binding competition. *Nucleic Acids Res.* **42,** 5495–5504 (2014).
191. Kolodrubetz, D. & Schleif, R. Identification of araC protein and two-dimensional gels, its in vivo instability and normal level. *J. Mol. Biol.* **149,** 133–9 (1981).
192. Kamionka, A., Bogdanska-Urbaniak, J., Scholz, O. & Hillen, W. Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Res.* **32,** 842–7 (2004).
193. Levandoski, M. M. *et al.* Cooperative and Anticooperative Effects in Binding of the First and Second Plasmid OsymOperators to a LacI Tetramer: Evidence for Contributions of Non-operator DNA Binding by Wrapping and Looping. *J. Mol. Biol.* **260,** 697–717 (1996).

194. Cookson, N. a *et al.* Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol. Syst. Biol.* **7,** 561–561 (2011).
195. Bergmiller, T. *et al.* Biased partitioning of the multidrug efflux pump AcrAB-TolC underlies long-lived phenotypic heterogeneity. *Science (80-.).* **356,** 311–315 (2017).
196. Huang, Z., Senocak, F., Jayaraman, A. & Hahn, J. Integrated modeling and experimental approach for determining transcription factor profiles from fluorescent reporter data. *BMC Syst. Biol.* **2,** 64 (2008).
197. Maienschein-Cline, M., Warmflash, A. & Dinner, A. R. Defining cooperativity in gene regulation locally through intrinsic noise. *IET Syst. Biol.* **4,** 379–392 (2010).
198. Saiz, L. & Vilar, J. M. G. Stochastic dynamics of macromolecular-assembly networks. *Mol. Syst. Biol.* **2,** 2006.0024 (2006).
199. Schleif, R. AraC protein, regulation of the l-arabinose operon in Escherichia coli , and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* **34,** 779–796 (2010).
200. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers — the database of key numbers in molecular and cell biology. **38,** 750–753 (2010).
201. Morelli, M. J., Ten Wolde, P. R. & Allen, R. J. DNA looping provides stability and robustness to the bacteriophage lambda switch. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 8101–6 (2009).
202. Pardee, A. B., Jacob, F. & Monod, J. The genetic control and cytoplasmic expression of "Inducibility" in the synthesis of β-galactosidase by E. coli. *J. Mol. Biol.* **1,** 165–178 (1959).
203. Elf, J., Li, G.-W. & Xie, X. S. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science (80-.).* **316,** 1191–1194 (2007).
204. Burger, A., Walczak, A. M. & Wolynes, P. G. Abduction and asylum in the lives of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 4016–21 (2010).
205. Griffith, J., Hochschild, A. & Ptashne, M. DNA loops induced by cooperative binding of lambda repressor. *Nature* **322,** 750–2 (1986).
206. Valenzuela, D. & Ptashne, M. P22 repressor mutants deficient in co-operative binding and DNA loop formation. *EMBO J.* **8,** 4345–50 (1989).
207. Weiss, M. A., Pabo, C. O., Karplus, M. & Sauer, R. T. Dimerization of the operator binding domain of phage lambda repressor. *Biochemistry* **26,** 897–904 (1987).
208. Chen, Y., Golding, I., Sawai, S., Guo, L. & Cox, E. C. Population Fitness and the Regulation of Escherichia coli Genes by Bacterial Viruses. *PLoS Biol.* **3,** e229 (2005).
209. Pray, T. R., Burz, D. S. & Ackers, G. K. Cooperative non-specific DNA binding by octamerizing λcI repressors: a site-specific thermodynamic analysis. *J. Mol. Biol.* **282,** 947–958 (1998).
210. Sarkar-Banerjee, S. *et al.* Specifically bound lambda repressor dimers promote adjacent non-specific binding. *PLoS One* **13,** e0194930 (2018).
211. Cooper, S. & Ruettinger, T. Replication of deoxyribonucleic acid during the division cycle of Salmonella typhimurium. *J. Bacteriol.* **114,** 966–973 (1973).
212. Sauer, R. T., Ross, M. J. & Ptashne, M. Cleavage of the lambda and P22 repressors by recA protein. *J. Biol. Chem.* **257,** 4458–62 (1982).
213. Arends, S. J. R. & Weiss, D. S. Inhibiting Cell Division in. *Society* **186,** 880–884 (2004).
214. Gullbrand, B. & Nordström, K. FtsZ ring formation without subsequent cell division after replication runout in Escherichia coli. *Mol. Microbiol.* **36,** 1349–1359 (2000).
215. de Boer, P. A. J. Advances in understanding E. coli cell fission. *Curr. Opin. Microbiol.* **13,** 730–737 (2010).

216. Baltrus, D. A. Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.* **28,** 489–495 (2013).
217. Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Appl. Environ. Microbiol.* **66,** 884–9 (2000).
218. Kitagawa, M. *et al.* Complete set of ORF clones of Escherichia coli ASKA library (A complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res.* **12,** 291–299 (2005).
219. Costantino, N. & Court, D. L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15748–53 (2003).
220. Friedberg, E. C., Walker, G. C., Siede, W. & Wood, R. D. *DNA repair and mutagenesis*. (American Society for Microbiology Press, 2005).
221. Hadizadeh Yazdi, N., Guet, C. C., Johnson, R. C. & Marko, J. F. Variation of the folding and dynamics of the Escherichia coli chromosome with growth conditions. *Mol. Microbiol.* **86,** 1318–1333 (2012).
222. Waldminghaus, T. & Skarstad, K. The Escherichia coli SeqA protein. *Plasmid* **61,** 141–50 (2009).
223. Nelson, H. C. M. & Sauer, R. T. Interaction of mutant λ repressors with operator and non-operator DNA. *J. Mol. Biol.* **192,** 27–38 (1986).
224. Waldminghaus, T. & Skarstad, K. ChIP on Chip: surprising results are often artifacts. *BMC Genomics* **11,** 414 (2010).
225. Clokie, M. R. J., Millard, A. D., Letarov, A. V. & Heaphy, S. Phages in nature. *Bacteriophage* **1,** 31–45 (2011).
226. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* **11,** 1511–1520 (2017).
227. Leitet, C., Riemann, L. & Hagström, Å. Plasmids and prophages in Baltic Sea bacterioplankton isolates. *J. Mar. Biol. Assoc. UK* **86,** 567 (2006).
228. Espeland, E. M., Lipp, E. K., Huq, A. & Colwell, R. R. Polylysogeny and prophage induction by secondary infection in Vibrio cholerae. *Environ. Microbiol.* **6,** 760–763 (2004).
229. Fogg, P. C. M., Allison, H. E., Saunders, J. R. & Mccarthy, A. J. Bacteriophage Lambda : a Paradigm Revisited □. **84,** 6876–6879 (2010).
230. Refardt, D. Within-host competition determines reproductive success of temperate bacteriophages. *ISME J.* **5,** 1451–60 (2011).
231. Díaz-Muñoz, S. L. Viral coinfection is shaped by host ecology and virus–virus interactions across diverse microbial taxa and environments. *Virus Evol.* **3,** vex011 (2017).
232. Berngruber, T. W., Weissing, F. J. & Gandon, S. Inhibition of Superinfection and the Evolution of Viral Latency. *J. Virol.* **84,** 10200–10208 (2010).
233. Kameyama, L., Fernández, L., Calderón, J., Ortiz-Rojas, a & Patterson, T. a. Characterization of wild lambdoid bacteriophages: detection of a wide distribution of phage immunity groups and identification of a nus-dependent, nonlambdoid phage group. *Virology* **263,** 100–111 (1999).
234. Mitarai, N., Brown, S. & Sneppen, K. Population Dynamics of Phage and Bacteria in Spatially Structured Habitats Using Phage λ and Escherichia coli. *J. Bacteriol.* **198,** 1783–1793 (2016).
235. Dennehy, J. J., Abedon, S. T. & Turner, P. E. Host density impacts relative fitness of bacteriophage Phi6 genotypes in structured habitats. *Evolution* **61,** 2516–27

(2007).
236. Díaz-Muñoz, S. L. Viral coinfection is shaped by host ecology and virus–virus interactions across diverse microbial taxa and environments. *Virus Evol.* **3,** 1–14 (2017).
237. Abedon, S. T. Commentary: Communication between Viruses Guides Lysis–Lysogeny Decisions. *Front. Microbiol.* **8,** 488–493 (2017).
238. Colon, M., Chakraborty, D., Pevzner, Y. & Koudelka, G. Mechanisms that Determine the Differential Stability of Stx+ and Stx− Lysogens. *Toxins (Basel).* **8,** 96 (2016).
239. Thomason, L. C., Oppenheim, A. B. & Court, D. L. *Bacteriophages*. **501,** (Humana Press, 2009).
240. Ellis, E. L. & Delbrück, M. THE GROWTH OF BACTERIOPHAGE. *J. Gen. Physiol.* **22,** 365–84 (1939).
241. Friedlander, T., Prizak, R., Barton, N. H. & Tkačik, G. Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nat. Commun.* **8,** 216 (2017).
242. van Hijum, S. A. F. T., Medema, M. H. & Kuipers, O. P. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol. Mol. Biol. Rev.* **73,** 481–509, Table of Contents (2009).
243. Pigliucci, M., Murren, C. J. & Schlichting, C. D. Phenotypic plasticity and evolution by genetic assimilation. *J. Exp. Biol.* **209,** 2362–7 (2006).
244. Ezer, D., Zabet, N. R. & Adryan, B. Physical constraints determine the logic of bacterial promoter architectures. *Nucleic Acids Res.* **42,** 4196–4207 (2014).
245. Mitarai, N., Semsey, S. & Sneppen, K. Dynamic competition between transcription initiation and repression : Role of nonequilibrium steps in cell-to-cell heterogeneity. **022710,** 1–6 (2015).
246. Giuntoli, R. D. *et al.* DNA-Segment-Facilitated Dissociation of Fis and NHP6A from DNA Detected via Single-Molecule Mechanical Response. *J. Mol. Biol.* **427,** 3123–36 (2015).
247. Savageau, M. A., Coelho, P. M. B. M., Fasani, R. A., Tolla, D. A. & Salvador, A. Phenotypes and tolerances in the design space of biochemical systems. **106,** (2009).
248. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309,** 2075–8 (2005).
249. Marko, J. F. Biophysics of protein–DNA interactions and chromosome organization. *Phys. A Stat. Mech. its Appl.* **418,** 126–153 (2015).
250. Madan Babu, M. & Teichmann, S. a. Functional determinants of transcription factors in Escherichia coli: protein families and binding sites. *Trends Genet.* **19,** 75–9 (2003).
251. Junier, I. Conserved patterns in bacterial genomes: a conundrum physically tailored by evolutionary tinkering. *Comput. Biol. Chem.* **53 Pt A,** 125–33 (2014).
252. Gibson, B., Wilson, D. J., Feil, E. & Eyre-Walker, A. The distribution of bacterial doubling times in the wild. *Proc. R. Soc. B Biol. Sci.* **285,** (2018).
253. Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440,** 545–550 (2006).