

STUDYING THE COMPLEXITIES OF TRANSCRIPTIONAL REGULATION

by

Georg Rieckh

May, 2016

A Thesis

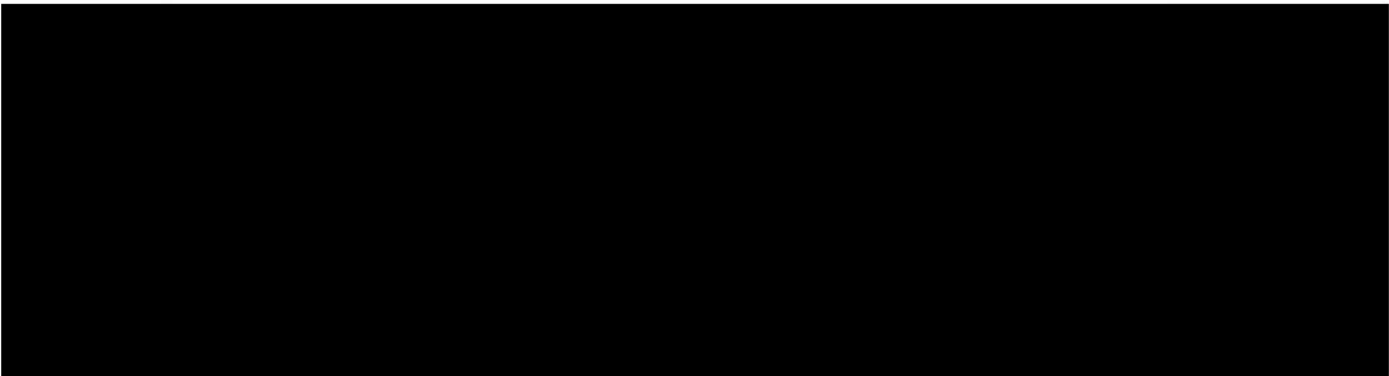
*Presented to the Faculty of the Graduate School of the
Institute of Science and Technology Austria, Klosterneuburg, Austria
in Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy*

Supervisor: Gašper Tkačik, IST Austria, Klosterneuburg, Austria

Committee Member: Călin Guet, IST Austria, Klosterneuburg, Austria

Committee Member: Terence Hwa, University of California at San Diego, La Jolla, CA, USA

Program Chair: Nicholas Barton, IST Austria, Klosterneuburg, Austria



© by Georg Rieckh, August, 2016

All Rights Reserved

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

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.

Signature: _____

Georg Rieckh

August, 2016

Biographical Sketch

Education

- 06/2009 Mag. phil. (MA) in Philosophy, University of Vienna
– Thesis in Philosophy of Language
- 09/2008 BSc in Mathematics, University of Vienna
– Theses in Logics and in Harmonic Analysis
- 05/2008 Mag. rer. nat. (MSc) in Physics, University of Vienna
– Thesis in Experimental Aerosol Science

Research Experience

- 06/2011 – present PhD student, Tkačik Group, IST Austria
- 04/2011 – 05/2011 Visiting scientist, Departement for Theoretical and Mathematical Physics, Universite Catholique de Louvain, Belgium
- 08/2008 – 03/2011 Research assistant, Computational Acoustics Group, Acoustics Research Institute, Austrian Academy of Sciences, Austria
- 05/2007 – 08/2007 Student assistant, Aerosol Physics Group, Institute for Experimental Physics, University of Vienna, Austria

Teaching Experience

- 10/2013 – 12/2013 Teaching assistant, Methods of Data Analysis, IST Austria
- 09/2012 Teaching assistant, Python for Biological Applications, IST Austria
- 09/2007 – 02/2008 Teaching assistant, Introduction to Physics, University of Vienna
- 03/2004 – 06/2007 Tutor, Introduction to Physics Labs, University of Vienna

List of Publications

1. S. Cepeda-Humerez, **G. Rieckh**, G. Tkačik – *Stochastic proofreading mechanism alleviates crosstalk in transcriptional regulation*, Physical Review Letters **115**:248101 (2015).
2. **G. Rieckh**, G. Tkačik – *Noise and information transmission in promoters with multiple internal states*, Biophysical Journal **106**:1194–204 (2014).
3. **G. Rieckh**, W. Kreuzer, H. Waubke, P. Balazs – *A 2.5D-Fourier-BEM model for vibrations in a tunnel running through layered anisotropic soil*, Engineering Analysis with Boundary Elements **36**, 960–7 (2012).
4. W. Kreuzer, H. Waubke, **G. Rieckh**, P. Balazs – *A 3D model to simulate vibrations in a layered medium with stochastic material parameters*, Journal of Computational Acoustics **19**, 139–54 (2011).
5. E. Ramharter, **G. Rieckh** – *Die Principia Mathematica auf den Punkt gebracht*, öbv&hpt Verlagsgesellschaft, 2007.

Abstract

The process of gene expression is central to the modern understanding of how cellular systems function. In this process, a special kind of regulatory proteins, called transcription factors, are important to determine how much protein is produced from a given gene. As biological information is transmitted from transcription factor concentration to mRNA levels to amounts of protein, various sources of noise arise and pose limits to the fidelity of intracellular signaling.

This thesis concerns itself with several aspects of stochastic gene expression: (i) the mathematical description of complex promoters responsible for the stochastic production of biomolecules, (ii) fundamental limits to information processing the cell faces due to the interference from multiple fluctuating signals, (iii) how the presence of gene expression noise influences the evolution of regulatory sequences, (iv) and tools for the experimental study of origins and consequences of cell-cell heterogeneity, including an application to bacterial stress response systems.

Table of Contents

Biographical Sketch	iii
List of Publications	iv
Abstract	v
0 Introduction	1
1 Noise and information transmission in complex promoters	5
1.1 Introduction	5
1.2 Channel capacity as a measure of regulatory power	9
1.3 Multi-state promoters as state transition graphs	10
1.4 Input noise	12
1.5 Experimentally accessible noise characteristics	14
1.6 Information transmission in simple gene regulatory elements	19
1.7 Improving information transmission with multi-state promoters	21
1.7.1 Cooperative regulation	21
1.7.2 Regulation with dual-role transcription factors	24
1.7.3 Promotor cycling	25
1.8 Experimentally measured promoter switching rates	27
1.9 Langevin method and promoters as state-transition graphs	27
1.9.1 Translating a state transition diagram into dynamic equations	27
1.9.2 Example: Dual regulation (3E-1a)	30
1.9.3 Comparison to other methods	32
1.10 Discussion	33

2	The broadcasting cell	36
2.1	Introduction	36
2.2	A simple model demonstrates the detrimental effects of cross-talk on signaling fidelity	38
2.3	Kinetic proofreading alleviates deleterious effects of cross-talk	38
2.3.1	Robustness to non-optimal input distributions	48
2.3.2	Possible molecular implementations of proofreading in transcriptional regulation	50
3	Biophysical aspects of noisy gene expression evolution	54
3.1	Introduction	54
3.2	Noise changes the fitness landscape	56
3.2.1	Effective fitness	56
3.2.2	Gene expression noise is not generally equivalent to a change in population size	58
3.2.3	Noise smoothens the effective fitness function and can shift the maximally fit phenotype	60
3.3	Genotype-phenotype maps based on thermodynamic models of gene expression	63
3.3.1	Simple form of neutral distribution of mean expression phenotypes	64
3.3.2	Joint densities of states for mean and noise in thermodynamic models . .	65
3.4	Combined effects of expression noise, density of states, and finite population size on genetic diversity	69
3.5	Effects of a joint genotype-phenotype map for noise and mean expression . . .	71
3.6	Dynamics of evolution for noisy promoters	74
3.6.1	Mutation rates in mismatch models	74
3.6.2	Population genetics: dynamics of adaptation in monomorphic populations	75
3.6.3	Fitness model with time averaging	77
3.6.4	Expression model for a small regulatory network	77
3.6.5	Trade-off in evolutionary dynamics	79
3.6.6	Discussion	82

4 A genetic platform to study single cell stress response	85
4.1 Introduction	85
4.2 A cloning strategy for the construction of chromosomal reporter constructs	85
4.2.1 Construction of a plasmid containing the platform	88
4.2.2 Chromosomal integration of promoters of interest	88
4.3 Application: Promoter activity and survival under acid stress in single cells	90
4.3.1 Bacterial stress response and cross-protection	90
4.3.2 Single cell survival can be predicted by activity of the acid stress promoter <i>P_{gadB}</i>	91
4.3.3 Experimental set-up	94
Bibliography	95

0 Introduction

Two discoveries in molecular biology and genetics during the middle of the last century were of fundamental importance for our modern understanding of how biological systems are organized [Alberts *et al.*, 2002]. First, this was the identification of the most important biochemical molecules and their roles, summarized in the so called central dogma of molecular biology (see Figure 1A) [Crick, 1970]. The second crucial discovery was the existence of a certain class of proteins: DNA-binding transcription factors (TFs) that determine which proteins are produced in which amounts and under which circumstances [Jacob and Monod, 1961]. While the central dogma describes the structure of biopolymers, gene expression research focuses on the regulation of their amount in a given condition. The reliability of cellular processes that determine the amount of biomolecules is the over-arching topic of this thesis.

Every step in the central dogma offers a possibility to tune the amount of protein that is made, most importantly transcription (DNA to mRNA) and translation (mRNA to protein). For example, the propensity of a ribosome to bind to an mRNA molecule and start translating will determine how many copies of the encoded protein will be made (on average) before a transcript is degraded. More important for this thesis, however, is the fact that some proteins bind to special regions of the DNA, termed regulatory regions, and influence the rate of transcription of a gene (see Figure 1B). This fact – that proteins can again interact with DNA and determine its utilization – closes the loop in the sense that the activity of one part of the DNA can influence the activity of another part. This potential to interact gives rise to gene networks (see Figure 1C).

Gene regulatory interactions and the networks arising from them are central in the modern understanding of several sub-fields of biology. Embryonic development, metabolism, cellular stress response, and even clinically relevant topics such as cancer and antibiotic resistance cannot be understood today without taking at least some aspects of gene regulation into account.

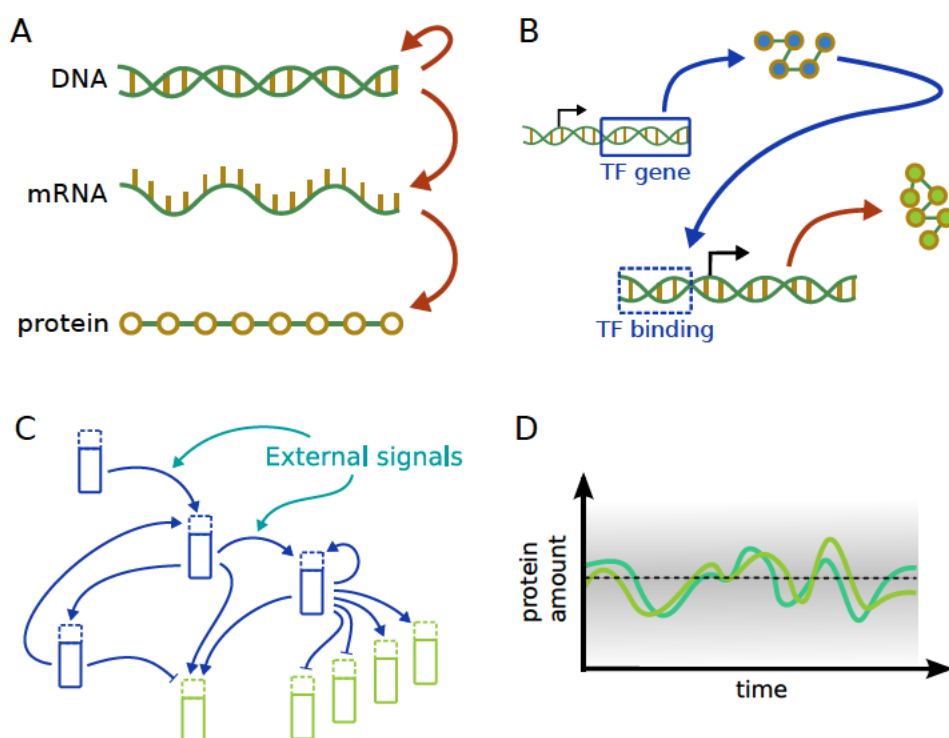


Figure 1: **(A)** The central dogma describes how information about the sequence of biopolymers is transferred (red arrows). The rate at which this information is used in the production of other molecules determines the amount of molecules that will be present at any given time point. **(B)** Some proteins, when they have folded into their functional form (blue protein), have the ability to bind to regions of the DNA (dashed blue box). This will then influence the amount of protein produced from a gene under the influence of this regulatory region (green protein). For activators (repressors), an increased amount of transcription factors will lead to an increased (decreased) production of the target gene and vice versa. Black arrows indicate the start of the protein coding region of a gene, and separate it from the regulatory region. **(C)** Many of these transcription factor genes (dashed box: regulatory elements, full box: protein coding DNA) are combined into gene regulatory networks. External signals (petroleum blue) enter the network at different times and nodes. This leads to the network adapting its state to the current environment. To achieve this, activating (pointy arrows) and repressing (blunt arrows) interactions are used. Blue genes are TFs and green genes various output genes of the network. **(D)** Two equivalent copies of a gene, exposed to a constant global environment will still differ in their activity to some degree (from each other and over time). They both fluctuate around the same mean (dashed black line) with some variance (gray shaded area). The two copies share some sources of stochasticity which causes their activity to be correlated – but they also differ in some aspects, pointing to an additional noise source.

One important aspect common to every step of gene expression and thus to every interaction in a gene network (TF production, diffusion, and binding, initiation of transcription and translation, degradation of mRNA and protein) is that they are performed by single molecules, sometimes present at very low copy numbers – most extremely exemplified by the fact that there is only a single gene coding for a particular type of protein [Babu *et al.*, 2004]. This discreteness and the susceptibility to thermally induced fluctuations in individual reaction events unavoidably lead to fluctuations in the whole reaction network [van Kampen, 1981] – in addition to other, naturally occurring sources of stochasticity [Raj and van Oudenaarden, 2008]. A classic tool used when studying the effects of diversity between clonal individuals, i.e. differences that cannot be traced to genetic differences, are twin studies [Raser and O’Shea, 2005; Burga *et al.*, 2011]. The underlying idea is that two identical copies of a system that are exposed to the same environment, will still display differences in their reaction or behavior (see Figure 1D). The first ‘twin study’ on the level individual genes, resp. the activities of two equivalent copies of it, is now usually taken (and cited) as the starting point for the investigation of stochasticity in gene expression [Elowitz *et al.*, 2002].

From the biophysical standpoint, the identification and quantification of noise sources – either just as a phenomenological model providing us with a quantitative language to talk about differences in variants, or with fundamental reasoning from first principles – is of central interest [Sanchez *et al.*, 2013]. Consequently, there is a large body of work characterizing different sources of noise and their genetic determinants [Swain *et al.*, 2002; Ozbudak *et al.*, 2002; Rosenfeld *et al.*, 2005; Pedraza and van Oudenaarden, 2005; Tkačik *et al.*, 2008c; Singh *et al.*, 2012]. To quantitatively describe different situations, a mathematical description was needed. The one most widely used today goes back to the random telegraph model (of transcription bursting) [Kepler and Elston, 2001; model for the statistical fluctuations of protein numbers in a microbial population, 1978; Cai *et al.*, 2006]. As experimental efforts got more focused, the data collected could distinguish more complicated models [Suter *et al.*, 2011; Sanchez *et al.*, 2011a] and dedicated inference schemes to learn their parameters were devised [Neuert *et al.*, 2013; Sherman and Cohen, 2014; Ruess *et al.*, 2015].

Questions with a more biological flavor revolve around the effects of noisy gene expression on various aspects of organismal behavior – and ultimately on their influence on evolution [Raser and O’Shea, 2005; Eldar and Elowitz, 2010; Wolf *et al.*, 2015]. From the engineering perspective, this means that noise can be taken into account when designing, building and optimizing synthetic gene networks [Bandiera *et al.*, 2016; Tsimring, 2014]. In developmental biology, the question about which architecture of regulatory circuits can provide an adequate bal-

ance between openness to innovation and yet deliver the necessary reliability for development continues to be a much researched topic [Houchmandzadeh *et al.*, 2002; Gregor *et al.*, 2007; Tkačik *et al.*, 2008b; Eldar *et al.*, 2009; Raj *et al.*, 2010; Dubuis *et al.*, 2013].

Stochastic effects have been studied widely in biological systems for a long time, also preceding the molecular era. Examples can be found in systems of very different scales ranging from single cell behavior and tissues, over neuronal networks to swarm behavior and whole ecosystems, often sharing tools for their mathematical analysis [MacKay, 2003; Rieke *et al.*, 1997; Bialek, 2013; Tkačik and Bialek, 2016]. In population genetics, the concept of finite population size leading to sampling noise in selection, and the existence of environmental fluctuations adding variance in phenotype to the genetic variance in a population are among the most prominent examples [Barton *et al.*, 2007; Lenormand *et al.*, 2009]. Recent progress in measurement techniques has made it possible to also investigate variations on the molecular level, and relate this to variations on higher levels of organismal organization – in the most extreme case, the difference between survival and death of single individuals.

As in other subfields of biology, noise in gene regulation has first been documented experimentally and studied mathematically, but the field has now moved on to understand the range of possible functional consequences: in which cases is noise only a ‘small correction’ to the average behavior and in which cases does it give rise to new phenomena?

In this thesis I present work relating to different aspects of stochastic gene expression mentioned above: the mathematical description of complex promoters that stochastically produce biomolecules (Chapter 1), fundamental limits to information processing the cell faces due to the interference from multiple fluctuating signals (Chapter 2), how the presence of gene expression noise influences the evolution of regulatory sequences (Chapter 3), and finally tools for the experimental study of origins and consequences of cell-cell heterogeneity, including an application to bacterial stress response systems (Chapter 4).

1 Noise and information transmission in complex promoters

The work presented in this chapter was conducted jointly with and Gašper Tkačik and has been published in the *Biophysical Journal* (see [Rieckh and Tkačik, 2014]) and is reproduced here with minimal changes.

1.1 Introduction

Gene regulation – the ability of cells to modulate the expression level of genes to match their current needs – is crucial for survival. One important determinant of this process is the wiring diagram of the regulatory network, specifying how environmental or internal signals are detected, propagated, and combined to orchestrate protein level changes [Levine and Davidson, 2005]. Beyond the wiring diagram, the capacity of the network to reliably transmit information about signal variations is determined also by the strength of the network interactions (the “numbers on the arrows” [Ronen *et al.*, 2002]), the dynamics of the response, and the noise inherent to chemical processes happening at low copy numbers [Elowitz *et al.*, 2002; Ozbudak *et al.*, 2002; Paulsson, 2004; Raj and van Oudenaarden, 2008].

How do these factors combine to set the regulatory power of the cell? Information theory can provide a general measure of the limits to which a cell can reliably control its gene expression levels. Especially in the context of developmental processes, where the precise establishment and readout of positional information has long been appreciated as crucial [Houchmandzadeh *et al.*, 2002], information theory can provide a quantitative proxy for the biological function of gene regulation [Tkačik and Walczak, 2011]. This has led to theoretical predictions of optimal networks that maximize transmitted information given biophysical constraints [Ziv *et al.*, 2007; Tkačik *et al.*, 2008a; Tkačik *et al.*, 2009; Walczak *et al.*, 2010; Tkačik and Walczak, 2011; Tkačik *et al.*, 2012], and hypotheses that certain biological networks might have evolved to maximize transmitted information [Tkačik *et al.*, 2008b]. Some evidence for these ideas has

been provided by recent high-precision measurements in the gap gene network of the fruit fly [Dubuis *et al.*, 2013]. In parallel to this line of research, information theory has been used as a general and quantitative way to compare signal processing motifs [Tostevin and ten Wolde, 2009; Tostevin and ten Wolde, 2010; Cheong *et al.*, 2011; de Ronde *et al.*, 2010; de Ronde *et al.*, 2012; de Ronde *et al.*, 2011; Tostevin *et al.*, 2012; Mugler *et al.*, 2013; Bowsher and Swain, 2012; Jost *et al.*, 2013; Hormoz, 2013; Levine *et al.*, 2007; Mancini *et al.*, 2013]. Further theoretical work has demonstrated a relationship between the information capacity of an organism's regulatory circuits and its evolutionary fitness [Taylor *et al.*, 2007; Rivoire and Leibler, 2011; Donaldson-Matasci *et al.*, 2010].

Previously, information theoretic investigations primarily examined the role of the regulatory network. Here we focus on the molecular level, i.e., on the events taking place at the regulatory regions of the DNA. Little is known about how the architecture of such microscopic events shapes information transfer in gene regulation. Yet it is precisely at these regulatory regions that the mapping from the “inputs” in the network wiring diagram into the corresponding “output” expression level is implemented by individual molecular interactions. In this bottleneck various physical sources of stochasticity – such as the binding and diffusion of molecules [Bialek and Setayeshgar, 2005; Bialek and Setayeshgar, 2008; Gregor *et al.*, 2007], and the discrete nature of chemical reactions [van Kampen, 1981] – must play an important role. In the simplest picture, gene expression is modulated through transcriptional regulation. This involves molecular events like the binding of transcription factors (TFs) to specific sites on the DNA, chemical events that facilitate or block TF binding (e.g., through chromatin modification), or events that are subsequently required to initiate transcription (e.g., the assembly and activation of the transcription machinery).

While the exact sequence of molecular events at the regulatory regions often remains elusive (especially in eukaryotes), quantitative measurements have highlighted factors that contribute to the fidelity by which TFs can affect the expression of their target genes. These findings have been succinctly summarized by the so-called “telegraph model” of transcriptional regulation [Peccoud and Ycart, 1995]: the two-state promoter switches stochastically between the states “ON” and “OFF”, with switching rates dependent on the concentration(s) of the regulatory factor(s). This dependence can either be biophysically motivated (e.g. by a thermodynamic model of TF binding to DNA), or it can be considered as purely phenomenological. The switching itself is independent of mRNA production, but determines the overall production rate. The production of mRNA molecules from one state is usually modeled as a Poisson process, with a first-order decay of messages; this is usually followed by a birth-death process

in which proteins are translated from the messages. This two-state model is well-studied theoretically [Peccoud and Ycart, 1995; Iyer-Biswas *et al.*, 2009; Shahrezaei and Swain, 2008; Dobrzynski and Bruggeman, 2009; Kepler and Elston, 2001; Hu *et al.*, 2011; Hu *et al.*, 2012] and has been used extensively to account for measurements of noise in gene expression [So *et al.*, 2011; Tkačik *et al.*, 2008c; Raser and O’Shea, 2004; Raj *et al.*, 2006]. An increasing amount of information about molecular details has motivated extensions to this model by introducing more than two states in specific systems [Sanchez and Kondev, 2008; Gutierrez *et al.*, 2009; Karmakar, 2010; Coulon *et al.*, 2010; Sanchez *et al.*, 2011a; Zhou and Zhang, 2012; Zhang *et al.*, 2012; Gutierrez *et al.*, 2012], and recent measurements of noise in gene expression provided some support for such complex regulatory schemes [Blake *et al.*, 2006; Blake *et al.*, 2003; Kandhavelu *et al.*, 2011; Suter *et al.*, 2011].

Here we address the general question of the functional effect of complex promoters with multiple internal states. How does the presence of multiple promoter states affect information transmission? Which promoter architectures transmit information more reliably when placed into a regulatory network? Under what conditions, if any, can multi-state promoters perform better than the two-state model? To address these questions, we consider a wide spectrum of generic promoter models that can be treated mathematically as state transition diagrams; many molecular “implementations” could thus share the same underlying model. When placed into a network, one must further specify which of the transitions are affected by concentrations of regulatory proteins, and which of the promoter states have nonzero expression rates. With this framework in hand, we derive the total noise in mRNA expression as a function of the induction level for all two- and three-state promoter models, and discuss how measurements of this function can be diagnostic of the underlying mechanism of regulation. To answer the main question of this paper – namely if additional complexity at the promoter can lead to an improvement in controlling the output level of a gene – we compute the information transmission from transcription factor concentrations to regulated protein expression levels through two- and three-state promoters. Finally, we analyze in detail three complex promoter architectures that outperform the two-state regulation.

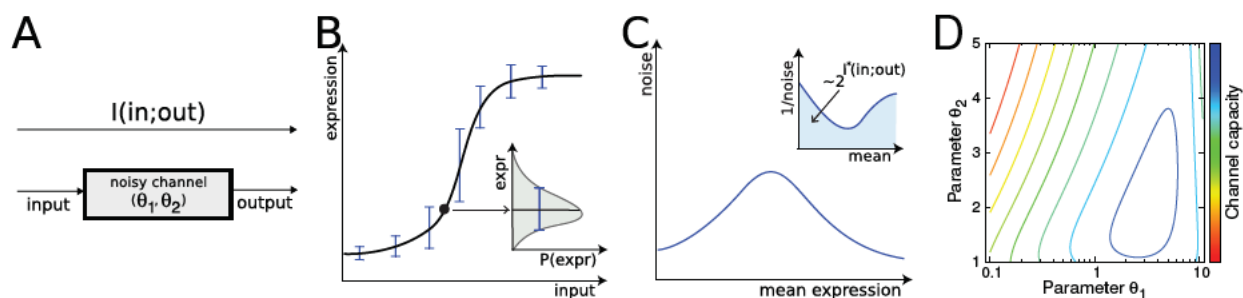


Figure 1.1: **A genetic regulatory element as an information channel.** **(A)** Mutual information I is a quantitative measure of the signaling fidelity with which a genetic regulatory element maps inputs (e.g. TF concentrations) into the regulated expression levels. In this schematic example, the properties of the element are fully specified by two parameters (θ_1, θ_2 ; e.g., the switching rates between promoter states). **(B)** In steady state, the input/output mapping can be summarized by the regulatory function $\bar{g}(k)$ (solid black line) for target protein expression (equivalently, $\bar{m}(k)$ for target mRNA expression, not shown); noise, $\sigma_g(k)$ (resp. $\sigma_m(k)$ for mRNA), induces fluctuations around this curve (inset and error bars on the regulatory function). **(C)** The “noise characteristic” (noise vs. mean expression) is usually experimentally accessible for mRNA using *in situ* hybridization methods and can reveal details about the promoter architecture. The maximal transmitted information (channel capacity I^* , see Eqs (1.2,1.3)) is calculated from the area under the inverse noise curve for the target protein, $\sigma_g^{-1}(\bar{g})$ (inset). **(D)** Channel capacity I^* is, in this example case, maximized for a specific choice of parameters θ_1, θ_2 (blue peak).

1.2 Channel capacity as a measure of regulatory power

We start by considering a genetic regulatory element – e.g., a promoter or an enhancer – as a communication channel, shown in Fig 1.1A. As the concentrations of the relevant inputs (for example, transcription factors) change, the regulatory element responds by varying the rate of target gene expression. In steady state, the relationship between input k and expression level of the regulated protein g is often thought of as a “regulatory function” [Setty *et al.*, 2003]. While attractive, the notion of a regulatory *function* in a mathematical sense is perhaps misleading: gene regulation is a noisy process, and so for a fixed value of the input we have not one, but a distribution of different possible output expression levels, $P(g|k)$ (see Fig 1.1B). When the noise is small, it is useful to think of a regulatory function as describing the *average* expression level, $\bar{g}(k) = \int dg g P(g|k)$, and of the noise as inducing some random fluctuation around that average. The variance of these fluctuations, $\sigma_g^2(k) = \int dg (g - \bar{g}(k))^2 P(g|k)$, is thus a measure of noise in the regulatory element; note that its magnitude depends on the input, k .

The presence of noise puts a bound on how precisely changes in the input can be mapped into resulting expression levels on the output side – or inversely, how much the cell can know about the input by observing the (noisy) outputs alone. In his seminal work on information theory [Shannon, 1948], Shannon introduced a way to quantify this intuition by means of *mutual information*, which is an assumption-free, positive scalar measure in bits, defined as

$$I(k; g) = \iint dk dg P(k)P(g|k) \log_2 \left[\frac{P(g|k)}{P(g)} \right] . \quad (1.1)$$

In Eq (1.1), $P(g|k)$ is a property of the regulatory element, which we will be computing below, while $P(k)$ is the distribution of inputs (e.g. TF concentrations) that regulate the expression; finally, $P(g) = \int dk P(g|k)P(k)$ is the resulting distribution of gene expression levels. With $P(g|k)$ set by the properties of the regulatory element and the biophysics of the gene expression machinery, there exists an optimal choice for the distribution of inputs, $P^*(k)$, that maximizes the transmitted information. This maximal value, $I^*(k; g) = \max_{P(k)} I(k; g)$, also known as the *channel capacity* [Cover and Thomas, 2006], summarizes in a single number the “regulatory power” intrinsic to the regulatory element [Tkačik *et al.*, 2008b; Tkačik *et al.*, 2008a; Tkačik *et al.*, 2009; Walczak *et al.*, 2010; Tkačik *et al.*, 2012].

Our goal is to compute the channel capacity between the (single) regulatory input and the target gene expression level for information flowing through various complex promoters. Under the assumption that noise is small and approximately gaussian for all levels of input, the complicated expression for the information transmission in Eq (1.1) simplifies, and the channel capacity $I^*(k; g)$ can be computed analytically from the regulatory function, $\bar{g}(k)$, and the noise,

$\sigma_g^2(k)$. The result is that [Tkačik *et al.*, 2008b; Tkačik *et al.*, 2008a; Tkačik *et al.*, 2009; Walczak *et al.*, 2010; Tkačik *et al.*, 2012]

$$I^*(k; g) = \log_2 \frac{Z}{\sqrt{2\pi e}}, \quad \text{with} \quad (1.2)$$

$$Z = \int_0^{k_{\max}} dk \frac{|d\bar{g}/dk|}{\sigma_g(k)} = \int_{\bar{g}_{\min}}^{\bar{g}_{\max}} \frac{d\bar{g}}{\sigma_g(\bar{g})}, \quad (1.3)$$

where in the last equality we changed the integration variables to express the result in terms of the average induction level, \bar{g} , using the regulatory function $\bar{g} = \bar{g}(k)$. This integral is graphically depicted in Fig 1.1C (inset). Finally, we will use this to explore the dependence of $I^*(k; g)$ on parameters that define the promoter architecture (see Fig 1.1D), looking for those arrangements that lead to large channel capacities and thus high regulatory power.

Information as a measure of regulatory power has a number of attractive mathematical properties (for review, see [Tkačik and Walczak, 2011]); interpretation-wise, the crucial property is that it roughly counts (the logarithm of) the number of distinguishable levels of expression that are accessible by varying the input – also taking into account the level of noise in the system. A capacity of 1 bit therefore suggests that the gene regulatory element could act as a binary switch with two distinguishable expression levels; capacities smaller than 1 bit correspond to (biased) stochastic switching, while capacities higher than 1 bit support graded regulation. An increase of information by 1 bit means that the number of tunable and distinguishable levels of gene expression has roughly doubled (!), implying that changes of less than a bit are meaningful. Careful analyses of gene expression data for single-input single-output transcriptional regulation suggest that real capacities can exceed 1 bit [Tkačik *et al.*, 2008b]. Increasing this number substantially beyond a few bits, however, necessitates very low levels of noise in gene regulation, requiring prohibitive numbers of signaling molecules [Tkačik *et al.*, 2008a].

1.3 Multi-state promoters as state transition graphs

To study information transmission, we must first introduce the noise model in gene regulation, which consists of two components: (i), the generalization of the random telegraph model to multiple states, and (ii), the model for input noise that captures fluctuations in the number of regulatory molecules. Starting with (i), we compute here the mean and variance for regulated mRNA levels, since these quantities are experimentally accessible when probing noise in gene expression. We assume that the system has reached steady state and that gene product degradation is the slowest timescale in the problem, i.e., that target mRNA or protein levels average over multiple state transitions of the promoter and that the resulting distributions of

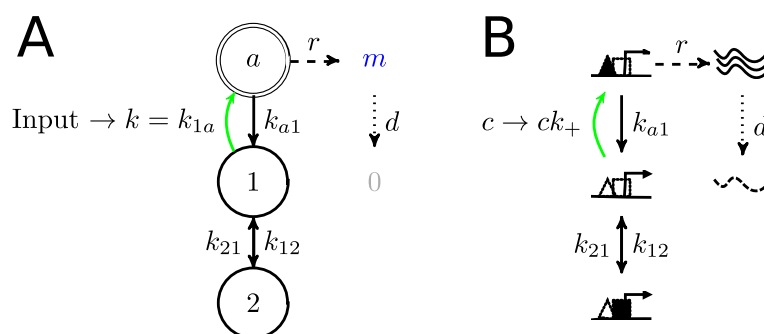


Figure 1.2: **Promoters as state transition graphs.** **(A)** A state transition graph for an example 3-state promoter. Active state a (double circle) expresses mRNA m at rate r , which are then degraded with rate d . Transition into a (green arrow) is affected by the input that modulates rate $k = k_{1a}$. Stochastic transitions between promoter states $\{a, 1, 2\}$ are an important contribution to the noise, $\sigma_m(k)$. **(B)** A possible mechanistic interpretation of the diagram in A: state 1 is an unoccupied promoter, state 2 is an inaccessible promoter (occupied by a nucleosome or repressor, black square). Transition to the active state (green arrow) is modulated by changing the concentration c of activators (filled triangles) which bind their cognate site (empty triangles) at the promoter with the rate ck_+ .

mRNA and protein are thus unimodal. While for protein levels these assumptions hold over a wide range of parameters and include many biologically relevant cases, there exist examples where promoter switching is very slow and the system would need to be treated with greater care (e.g., [Walczak *et al.*, 2005; Karmakar, 2010; Iyer-Biswas *et al.*, 2009]).

Let us represent the possible states of the promoter (and the transitions between them) by a state transition graph as in Fig 1.2A. Gene regulation occurs when an input signal modifies one (or more) of the rates at which the promoter switches between its states. To systematically analyze many promoter architectures, we choose not to endow from the start each graph with a mechanistic interpretation, which would map the abstract promoter states to various configurations of certain molecules on the regulatory regions of the DNA (as in Fig 1.2B). This is because there might be numerous molecular realizations of the same abstract scheme, which will yield identical noise characteristics and identical information transmission. In Fig 1.3, Fig 1.4, and Fig 1.5, we discuss known examples related to different promoter architectures.

Given a specific promoter architecture, we would like to compute the first two moments of the mRNA distribution under the above assumptions. Here, we only sketch the method for the promoter in Fig 1.2A; for a general description and details see section 1.9. We will denote the rate of mRNA production from the active state(s) by r and its degradation rate by d . Let further

p_i be the fractional occupancy of state $i \in \{a, 1, 2\}$ and k_{ij} the rate of transitioning from state i to j , $i \neq j$. Here, a is the active state, and 1, 2 are the non-expressing states. Equations (1.4) and (1.5) then describe the behavior of the state occupancy and mRNA level m :

$$\partial_t \mathbf{p} = \mathbf{K} \mathbf{p} + \boldsymbol{\xi}, \quad \partial_t m = r p_a - d m + \xi_m, \quad \text{with} \quad (1.4)$$

$$\mathbf{K} = \begin{bmatrix} -(k_{a1} + k_{a2}) & k_{1a} & k_{2a} \\ k_{a1} & -(k_{1a} + k_{12}) & k_{21} \\ k_{a2} & k_{12} & -(k_{2a} + k_{21}) \end{bmatrix}, \quad (1.5)$$

and $\mathbf{p} = (p_a, p_1, p_2)^T$; $\boldsymbol{\xi} = (\xi_a, \xi_1, \xi_2)^T$ and ξ_m are Langevin white-noise random forces [van Kampen, 1981; Gillespie, 2000] (see section 1.9). In this setup it is easy to compute the mean and the variance in expression levels given a set of chosen rate constants. Using the assumption of slow gene product degradation, $d \ll k_{ij}$, we can write the noise in a generic way:

$$\sigma_m^2 = \bar{m} \left[1 + \frac{r}{d} p_{\text{act}} \cdot \Delta \right], \quad (1.6)$$

where p_{act} is the occupancy of the active states (p_a or $p_a + p_b$), and the dimensionless expression for Δ depends on the promoter architecture and can be read out from Fig 1.3A for different promoter models. The expression for noise in Eq (1.6) has two contributions. The first, where the variance is equal to the mean ($\sigma_m^2 = \bar{m} + \dots$) is the ‘‘output noise’’ due to the birth-death production of single mRNA molecules (also called ‘‘shot noise’’ or ‘‘Poisson noise’’). The second contribution to the variance in Eq (1.6) is due to stochastic switching of the promoter between internal states, referred to as the ‘‘switching noise.’’ This term does depend on the promoter architecture and has a more complicated functional form than being simply proportional to the mean. A first glance at the expressions for noise seems to imply that going from two to three promoter states can only increase the noise (and by Eq (1.3) decrease information), since new, positive contributions appear in the expressions for σ_m^2 ; we will see that, nevertheless, transmitted information can increase for certain architectures.

1.4 Input noise

In addition to the noise sources internal to the regulatory mechanism, we also consider the propagation of fluctuations in the input, which will contribute to the observed variance in the gene expression level. Can we say anything general about the transmission of input fluctuations through the genetic regulatory element? Consider, for instance, the modulated rate k

that depends on the concentration c of some transcription factor, as in $k = k_+c$, where k_+ is the association rate to the TF's binding site. Since the TF itself is expressed in a stochastic process, we could expect that there will be (at least) Poisson-like fluctuations in c itself, such that $\sigma_c^2 \propto c$; this will lead to an effective variance in k that will be propagated to the output variance in proportion to the “susceptibility” of the regulatory element, $(\partial\bar{g}/\partial k)^2$. Extrinsic noise would affect the regulatory element in an analogous way, as suggested in [Swain *et al.*, 2002]. Independently of the noise origin, we can write

$$\sigma_g^2 = \dots + v \left(\frac{\partial\bar{g}}{\partial k} \right)^2 k, \quad (1.7)$$

where (...) indicate output and switching terms from Eq (1.8) and v is the proportionality constant ($\sigma_k^2 = vk$) that is related to the magnitude of the input fluctuations and, possibly, their subsequent time averaging [Paulsson, 2004].

Even if there were absolutely no fluctuations in the total concentration c of transcription factor molecules in the cell (or the nucleus), the sole fact that they need to find the regulatory site by diffusion puts a lower bound on the variance of the *local concentration* at the regulatory site. This diffusion limit, first formulated for the case of bacterial chemotaxis by Berg and Purcell [Berg and Purcell, 1977], has been subsequently derived for the general case of biochemical signaling [Bialek and Setayeshgar, 2005; Bialek and Setayeshgar, 2008]: the lower bound on the variance in local concentration obeys $\sigma_c^2 \propto cd'/D\ell$, where D is the diffusion constant of the TF molecules, ℓ is the linear size of the binding site, and $1/d'$ is the noise averaging time (here the lifetime of the gene product). Analyses of high-precision measurements in gene expression noise during early fruit fly development have shown that this diffusion noise represents a substantial contribution to the total [Gregor *et al.*, 2007; Tkačik *et al.*, 2008c]. Thus, for this biophysical limit set by diffusion, we find yet again that the variance in the input is proportional to the input itself. This, in sum, demonstrates that Eq (1.7) can be used as a very generic model for diverse kinds of input noise. To see which values the constant v can take, note that $\sigma_k^2 = k_+^2 \sigma_c^2 \propto k_+^2 cd'/D\ell = kk_+d'/D\ell$. As an example, consider diffusion-limited association, where $k_+ = 4\pi D\ell$ [Berg and von Hippel, 1985]. Depending on the accuracy and the geometry of the sensing mechanism we now get different values for $\tilde{v} = v/d'$, but in general \tilde{v} is expected to be of order unity. For example, the perfect absorbing sphere has $\sigma_c^2 = cd'/(4\pi D\ell)$ and therefore $\tilde{v} = 1$; the perfect monitoring sphere in the Berg–Purcell limit has $\sigma_c^2 = 3cd'/(5\pi D\ell)$ and therefore $\tilde{v} = 2.4$ [Endres and Wingreen, 2008; Berg and Purcell, 1977].

1.5 Experimentally accessible noise characteristics

Could complex promoter architectures be distinguished by their noise signatures, even in the easiest case where the input noise can be neglected (as is often assumed [Tkačik *et al.*, 2008c])? The expressions for the noise presented in Fig 1.3A hold independently of which transition rate the input is modulating. We can specialize these results by choosing the *modulation scheme*, that is, making one (or more) of the transition rates the regulated one. This allows us to construct the regulatory function (insets in Fig 1.3B). Additionally, we can also plot the noise (here shown as the Fano factor, σ_m^2/\bar{m}) as a function of the mean expression, \bar{m} , thus getting the *noise characteristic* of every modulation scheme. These curves, shown in Fig 1.3B, are often accessible from experiments [Carey *et al.*, 2013; So *et al.*, 2011], even when the identity of the expressing state or the mechanism of modulation are unknown. We systematically organize our results in Fig 1.3B (for the case when $k = k_{1a}$ is modulated), and provide a full version in Fig 1.4 and Fig 1.5; we also list four molecular schemes implementing these architectures in Fig 1.3C, while providing additional molecular implementations in Fig 1.4 and Fig 1.5. We emphasize that very different molecular mechanisms of regulation can be represented by the same architecture, resulting in the same mathematical analysis and information capacity.

Measured noise-vs-mean curves have been used to distinguish between various regulation models [de Ronde *et al.*, 2009; Carey *et al.*, 2013; So *et al.*, 2011]. For this, two conditions have to be met [Tkačik *et al.*, 2008c; Sanchez *et al.*, 2013]. First, it must be possible to access the full dynamic range of the gene expression in an experiment, and this sometimes seems hard to ensure. The second condition is that the input noise is not the dominant source of noise: input noise can mimic promoter switching noise and can, e.g., provide alternative explanations for noise measurements in [So *et al.*, 2011] that quantitatively fit the data (not shown).

Even if these conditions are met, it would be impossible to distinguish between certain promoter architectures (e.g., 2-a1 vs. 3E-a1) with this method, while some would require data of a very high quality to distinguish (e.g., activating 3E-1a vs. repressing 3E-12, see Fig 1.4 and Fig 1.5), at least in certain parameter regimes. On the other hand, there exist noise characteristics that can only be obtained with multiple states (e.g., 3M-1a).

One feature that can easily be extracted from the measured noise characteristics is the asymptotic induction: it can be equal to 1 (e.g., in 2-1a), or bounded away from 1 (e.g., in 3M-1a). While this distinction between architectures cannot be inferred from the shapes of the regulatory functions, the effect on the noise characteristics is unambiguous: in the case where the expressing state is never saturated, the Fano factor does not drop to the Poisson limit of 1

even at the highest expression levels (which seems to have been the case in [So *et al.*, 2011]).

Taken together, when the range of promoter architectures is extended beyond the two-state model, distinguishing between these architectures based on the noise characteristics seems possible only under restricted conditions, emphasizing the need for dynamical measurements that directly probe transition rates (e.g., [Suter *et al.*, 2011; Golding *et al.*, 2005]), or for the measurements of the full mRNA distribution (rather than only its second moment). We note that dynamic rates are often reported *assuming* the two-state model, as they are inferred from the steady state noise measurements (e.g., [So *et al.*, 2011; Raj *et al.*, 2006]), and only a few experiments probe the rates directly (e.g., [Geertz *et al.*, 2012]); for a brief review of the rates and their typical magnitudes see section 1.8.

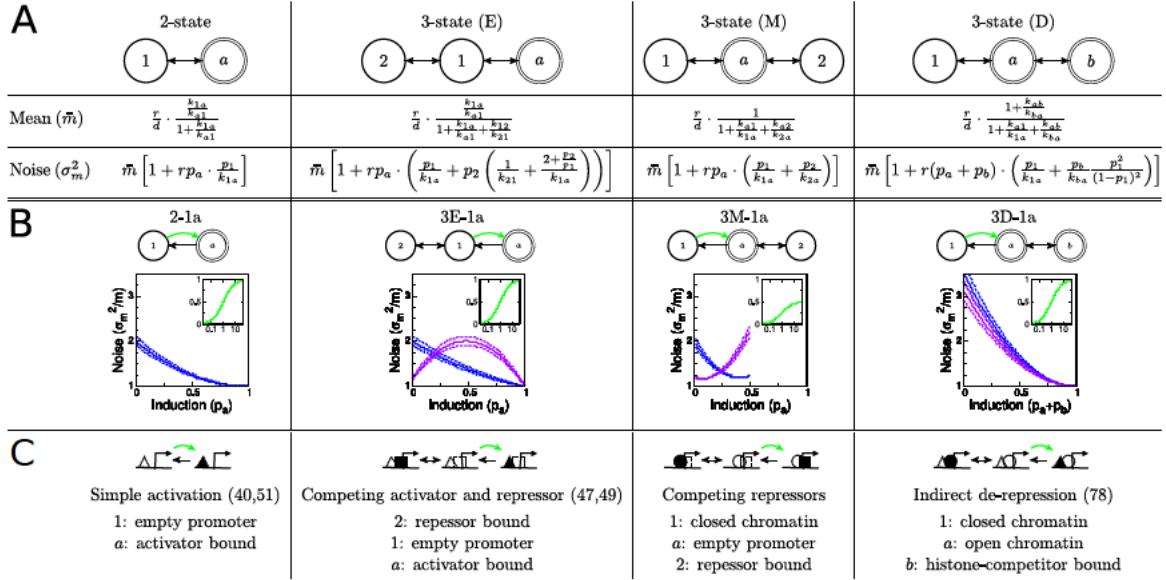


Figure 1.3: Mean expression level and noise for different promoter models. **(A)** Expressions for the mean (first row, \bar{m}) and the variance (second row, σ_m^2) of the mRNA distribution in steady state for different promoter architectures. In the limit of $k_{12} \rightarrow 0$ (resp. $k_{a2}, k_{ab} \rightarrow 0$), the expression for the two-state model is obtained from the models with three states. The names of the topologies indicate the position of the expressing state: E(nd), M(iddle), D(ouble). **(B)** Noise characteristics (\bar{m} on x -axis vs the Fano factor, σ_m^2/\bar{m} , on y -axis) for different promoter models. Here, in all models $k = k_{1a}$ is modulated (green arrow) to achieve different mean expression levels. For all rates (except k) equal to 1 (blue lines), the functional form of the noise characteristics is very similar. This remains true for a variation of k_{a1} of $\pm 10\%$ (blue dashed lines). Making the rates in/out of the third state (state 2 or b) slower by a factor of 5 (purple lines, dashed $\pm 10\%$) yields qualitatively different results. Insets show the induction curve, $p_{act}(k)$, where k is the modulated rate. A full table and possible molecular interpretations of different promoter schemes are given in Fig 1.4 and Fig 1.5 [Vinelas *et al.*, 2013; Sanchez *et al.*, 2011c; Garcia *et al.*, 2012; Yang and Ko, 2012; Earnest *et al.*, 2013; Mirny, 2010; Segal and Widom, 2009]. **(C)** Possible interpretations of the modulation schemes. Triangles represent activators, squares are repressors and circles are histones. The dotted shapes denote (empty) binding sites. Cited references use similar models.

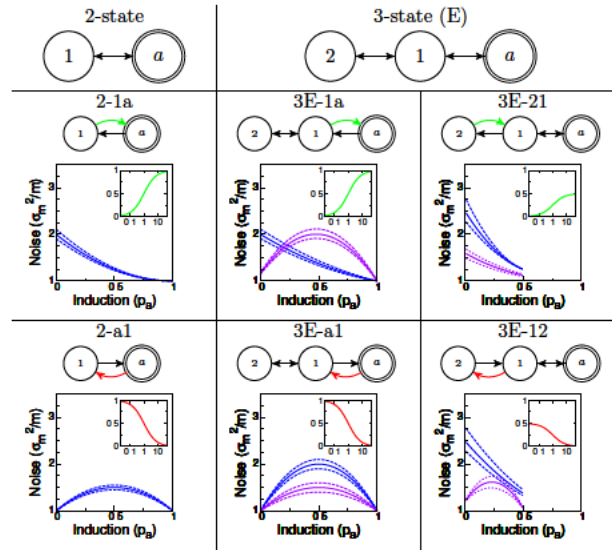


Figure 1.4: **Promoter architectures and interpretations.** **Scheme 2-1a:** (i) State a is the empty promoter (available for transcription), and state 1 corresponds to it being occupied by a repressing protein (e.g. a specific TF or a nucleosome). Mechanisms that change the rate of switching from state 1 into a (i.e. scheme 2-1a) are well documented for eukaryotic cells [Vinuelas *et al.*, 2013; Carey *et al.*, 2013]. (ii) Simple activation where state 1 is the empty promoter and state a has an activating TF bound to it; changing TF concentration modulates k_{1a} [Kepler and Elston, 2001; Sanchez *et al.*, 2011a]. **Scheme 2-a1:** The bacterial *lac*-promoter, where the binding of a specific TF represses expression. A change in the concentration of TFs now corresponds to modulating the rate k_{a1} and the rate k_{1a} is determined by the interaction energy between TF and its binding site [Sanchez *et al.*, 2011c; Garcia *et al.*, 2012]. **Scheme 3E:** (i) A promoter that has overlapping binding sites for both, an activator and a repressor (dual regulation) [Karmakar, 2010; Sanchez and Kondev, 2008; Yang and Ko, 2012]. Changing the concentration of the activator (repressor) leads to a change in rate k_{1a} (k_{12}), which means it can be modeled as a 3E-1a (3E-12) scheme. (ii) Eukaryotic promoters with a TATA box can be modeled as 3E-21 [Sanchez *et al.*, 2013]. Here, state 2 is the DNA in a state not available for transcription (closed chromatin), state 1 is the conformation where the promoter (and the TATA box) is exposed and state a corresponds to the active configuration of the DNA with the pre-initiation complex assembled. Changing the concentration of chromatin remodelers now influences the rate k_{21} (similar to the k_{1a} -modulation in the scheme 2-1a mentioned above), which yields scheme 3E-21. (iii) A coarse-grained model of DNA-looping in the *lac*-operon [Earnest *et al.*, 2013]. Cited references use similar models.

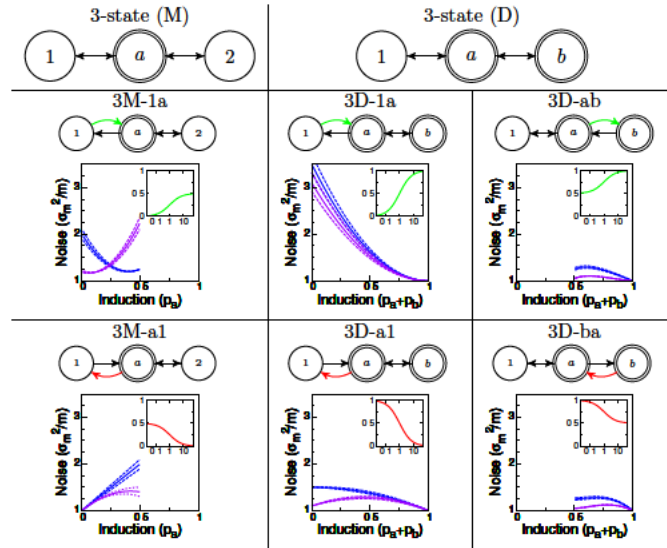


Figure 1.5: **Promoter architectures and interpretations (Continued).** **Scheme 3M:** A nucleosome and a specific, repressing TF compete for a promoter; changes in the input TF concentration correspond to 3M-a1, while changes in factors decreasing nucleosome occupancy correspond to 3M-1a. **Scheme 3D:** (i) State 1 is the closed chromatin formation, state a is the empty promoter and in state b a TF is bound to the DNA in such a way that it prohibits the closing of the chromatin (but still permits transcription) [Mirny, 2010; Segal and Widom, 2009]. In this way, even though the input molecule does not necessarily interact with the RNA polymerase directly, it can act as an activator (or rather as a de-repressor), yielding scheme 3D-ab. (ii) A scheme used to describe any promoter where the basal expression does not follow a Poisson process (optionally with different rates of expression from a and b). (iii) Promoter with a TATA box and a competing nucleosome (cf. 3E-21) if there is a significant amount of expression from the basal state a . (iv) A scheme used as a phenomenological model with unidentified third state to explain universal noise behavior in bacterial gene expression [Sanchez *et al.*, 2013; So *et al.*, 2011]. Cited references use similar models.

1.6 Information transmission in simple gene regulatory elements

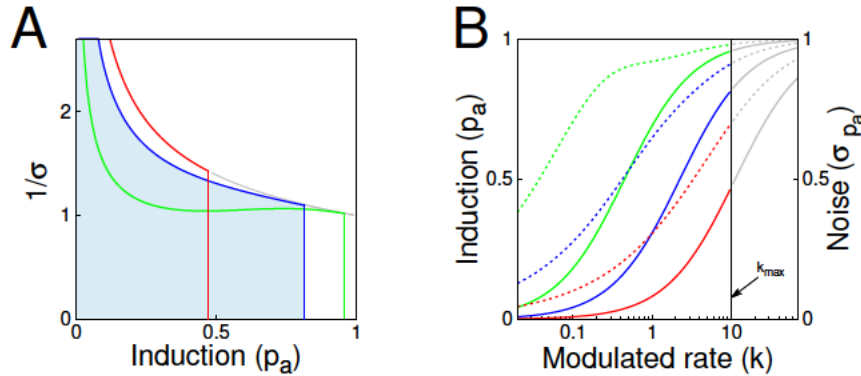


Figure 1.6: **Finding k_- that optimizes information transmission in a two-state promoter.**

The strength of the input noise is fixed at $\nu = 2$ and the input dynamic range for k is from 0 to $k_{\max} = 10$. **(A)** The integrand of Eq (1.10) is shown for the optimal choice of k_- (blue), and for two alternative k_- values: a factor of 5 larger (red) or smaller (green) than the optimum. While increasing k_- lowers the noise, it also decreases the integration limit, and vice versa for decreasing k_- . **(B)** The effect on the regulatory function (solid, left axis) and the noise (dashed, right axis), of choosing different k_- values. Optimal k_- (blue curves) from (A) leads to a balance between the dynamic range in the mean response (the maximal achievable induction), and the magnitude of the noise. Higher k_- (red curves), in contrast, lead to smaller noise, but fail to make use of the full dynamic range of the response. The gray part of the regulation curves cannot be accessed, since the input only ranges over $k \in [0, k_{\max}]$.

Protein noise. In most cases the functional output of a genetic regulatory element is not the mRNA, but the translated protein. Incorporating stochastic protein production into the noise model does not affect the functional form of the noise, but only rescales the magnitude of the noise terms. To see this, we let proteins be produced from mRNA at a rate r' and degraded or diluted at d' , such that $d' \ll d$ is the slowest timescale in the problem. Then the mean protein expression level is $\bar{g}(k) = (r'/d')\bar{m}(k)$. The output and promoter switching noise contributions are affected differently, so that the protein level noise can be written as [Tkačik *et al.*, 2008c]:

$$\sigma_g^2 = \bar{g} \left[(1 + \nu) + \frac{r'}{d'} p_{\text{act}} \cdot \Delta' \right] + (\text{input noise}), \quad (1.8)$$

where $\nu = r'/d$ is the *burst size* (the average number of proteins translated from one mRNA molecule), $\Delta' = \frac{r}{d} \frac{d'}{d} \Delta$ and the other quantities are as defined in Eq (1.6).

Information transmission in the two-state model. To establish the baseline against which to compare complex promoters, we look first at the two-state promoter (2-1a). Here the transition into the active state is modulated by TF concentration c via $k = k_{1a} = k_+c$, as it would be in the simple case of a single TF molecule binding to an activator site to turn on transcription. Adding together the noise contributions of Eqs (1.8,1.7), we obtain our model for the total noise:

$$\sigma_g^2 = \bar{g} \left[(1 + \nu) + \frac{rr'}{dk_{a1}}(1 - p_a)^2 + \frac{v}{k_{a1}} \frac{rr'}{dd'}(1 - p_a)^3 \right]. \quad (1.9)$$

To compute the corresponding channel capacity, we use Eq (1.3) with the noise given by Eq (1.9):

$$Z = \int_{\bar{g}_{\min}}^{\bar{g}_{\max}} \frac{d\bar{g}}{\sigma_g(\bar{g})} = \sqrt{N_{\max}} \int_0^{p_a^{\max}} dp_a \times \\ \times p_a^{-1/2} \left[1 + \frac{1}{\tilde{k}_-}(1 - p_a)^2 + \frac{\tilde{v}}{\tilde{k}_-}(1 - p_a)^3 \right]^{-1/2} \quad (1.10)$$

$$= \sqrt{N_{\max}} Z_0. \quad (1.11)$$

Here, $N_{\max} = (rr')/[(dd')(1 + \nu)]$ and $\tilde{k}_- = k_{a1}/(d'N_{\max})$ is the dimensionless combination of parameters related to the off-rate for the TF dissociation from the binding site. N_{\max} can be interpreted as the number of *independently* produced output molecules when the promoter is fully induced [Tkačik *et al.*, 2009; Tkačik *et al.*, 2012; Walczak *et al.*, 2010]. In the case where mRNA transcription is the limiting step for protein synthesis, N_{\max} corresponds to the maximal average number of mRNA synthesized during a protein lifetime: $N_{\max} = r/d' \cdot \nu/(1 + \nu) \approx r/d'$. With this choice of parameters, N_{\max} affects Z multiplicatively and thus simply adds a constant offset to the channel capacity I^* [see Eq (1.2)] without affecting the parameter values that maximize capacity. In what follows we therefore disregard this additive offset, and examine in detail only $I^* = \log_2 Z_0$. We also only use dimensionless quantities (as above, e.g., the rates are expressed in units of d'), but leave out the tilde symbols for clarity.

Optimizing information transmission. What parameters maximize the capacity of the two-state promoter 2-1a given by Eq (1.10)? Given that the dynamic range of input (e.g., TF concentration) is limited [Tkačik *et al.*, 2008a; Tkačik *et al.*, 2009; Tkačik *et al.*, 2012; Walczak *et al.*, 2010], $k \in [0, k_{\max}]$, and given a choice of v that determines the type and magnitude of input noise, the channel capacity I^* for the two-state promoter only depends non-trivially on the choice of a single parameter, k_- . Figure 1.6 shows the tradeoff that leads to the emergence of a well-defined optimal value for k_-^* : at a fixed dynamic range for the input, $k \in [0, k_{\max}]$, the information-maximizing solution chooses k_-^* that balances the strength of binding (such that the dynamic range of expression is large), while simultaneously keeping the noise as low as possible. If this abstract promoter model were interpreted in mechanistic terms where a TF

binds to activate gene expression, then choosing the optimal k_- would amount to choosing the optimal value for the dissociation constant of our TF; importantly, the existence of such a nontrivial optimum indicates that, at least in an information-theoretic sense, the best binding is not the tightest one [Li *et al.*, 2009; Grönlund *et al.*, 2013; Tkačik *et al.*, 2008a; Tkačik *et al.*, 2009; Walczak *et al.*, 2010; Tkačik *et al.*, 2012; Levine *et al.*, 2007]. This tradeoff between noise and dynamic range of outputs (also called “plasticity”) has also been noticed in other contexts [Lehner, 2010; Bajic and Poyatos, 2012].

1.7 Improving information transmission with multi-state promoters

We would like to know if complex promoter architectures can outperform the two-state model in terms of channel capacity. To this end, we have examined the full range of three-state promoters, summarized in Fig 1.4 and Fig 1.5, and found that generally – as long as only one transition is modulated and only one state is active – extra promoter states lead to a decrease in the channel capacity relative to two-state regulation. However, by relaxing these assumptions, architectures that outperform two-state promoters can be found.

1.7.1 Cooperative regulation

The first such pair of architectures is illustrated in Fig 1.7A and B: three-state promoters with one (or two) expressing states, where *two* transitions into the expressing states are simultaneously modulated by the input. A possible molecular interpretation of these promoter state diagrams is an AND-architecture cooperative binding for the model with one expressing state, and an OR-architecture cooperative activation for the model with two expressing states. In case of an AND-architecture, a TF molecule hops onto the empty promoter (state 2) with rate $2k$ (since there are two empty binding site), while a second molecule can hop on with rate ρk (called “recruitment” if $\rho > 1$), bringing the promoter into the active state. The first of two bound TF molecules falls off with rate $2\gamma^{-1}k_-$ (called “cooperativity” if $\gamma > 1$), bringing the promoter back to state 1, and ultimately, the last TF molecule can fall off with rate k_- . The dynamics are now described (cf. Eq (1.5)) by the matrix

$$\mathbf{K} = \begin{bmatrix} -(2\gamma^{-1}k_-) & \rho k & 0 \\ 2\gamma^{-1}k_- & -(k_- + \rho k) & 2k \\ 0 & k_- & -(2k) \end{bmatrix}, \quad (1.12)$$

and $\mathbf{p} = (p_a, p_1, p_2)^T$, resp. $\mathbf{p} = (p_b, p_a, p_1)^T$. To compute the noise, we can use the solutions for the generic three-state model 3E from Fig 1.3A by making the following substitutions: $k_{a1} = 2\gamma^{-1}k_-$, $k_{1a} = \rho k$, $k_{12} = k_-$, $k_{21} = 2k$.

To simplify our exploration of the parameter space, we choose $\rho = 1$ (i.e., no recruitment), but keep k_- (unbinding rate) and γ (cooperativity) as free parameters; the modulated rate k is proportional to the concentration of TF molecules and is allowed to range from $k \in [0, k_{\max}]$. For every choice of (k_-, γ) , we computed the regulatory function and the noise, and used these to compute the capacity, $I^*(k; g)$, using Eqs (1.2, 1.3). This information is shown in Fig 1.7C and D for the AND- and OR-architecture, respectively.

In the case of an AND-architecture, where both molecules of the TF have to bind for the promoter to express, there is a ridge of optimal solutions: as we move along the ridge in the direction of increasing information, cooperativity is increased and thus the doubly-occupied state is stabilized, while the unbinding rate increases as well. This means that the occupancy of state 1 becomes negligible, and the regulation function becomes ever steeper, as is clear from Fig 1.7E, while maintaining the same effective dissociation constant (the input $k = k_{1/2}$ at which the promoter is half induced, i.e., $p_a(k_{1/2}) = 0.5$). In this limit, the shape of the regulation function must approach a Hill function with the Hill coefficient of 2, $p_a(k) = k^2/(k^2 + k_{1/2}^2)$. Surprisingly, information maximization favors weak affinity of *individual* TF molecules to the DNA, accompanied by strong cooperativity between these molecules. The OR-architecture portrays a different picture: here, the maximum of information is well-defined for a particular combination of parameters (k_-, γ) , as shown in Fig 1.7D. As $\gamma \rightarrow 0$ (increasing destabilization for $\gamma < 1$), the second active state (b) is never occupied, and the model reverts to a two-state model.

For both architectures we can assess the advantage of the three-state model relative to the optimal two-state promoter. Figures 1.7E and F show the information of the optimal solutions as a function of the input noise magnitude as well as the input range, k_{\max} . As expected, the information increases as a function of k_{\max} since the influence of input and switching noise can be made smaller with more input molecules. This increase saturates at high k_{\max} because output noise becomes limiting to the information transmission – this is why the capacity curves converge to the same maximum, the $v = 0$ curve that lacks the input noise altogether. The advantage (increase in capacity) of the three-state models relative to the two-state promoter is positive for any combination of parameters k_{\max} and v . It is interesting to note that increasing k_{\max} and decreasing v have very similar effects on channel capacity, since both drive the system to a regime where the limiting factor is the output noise.

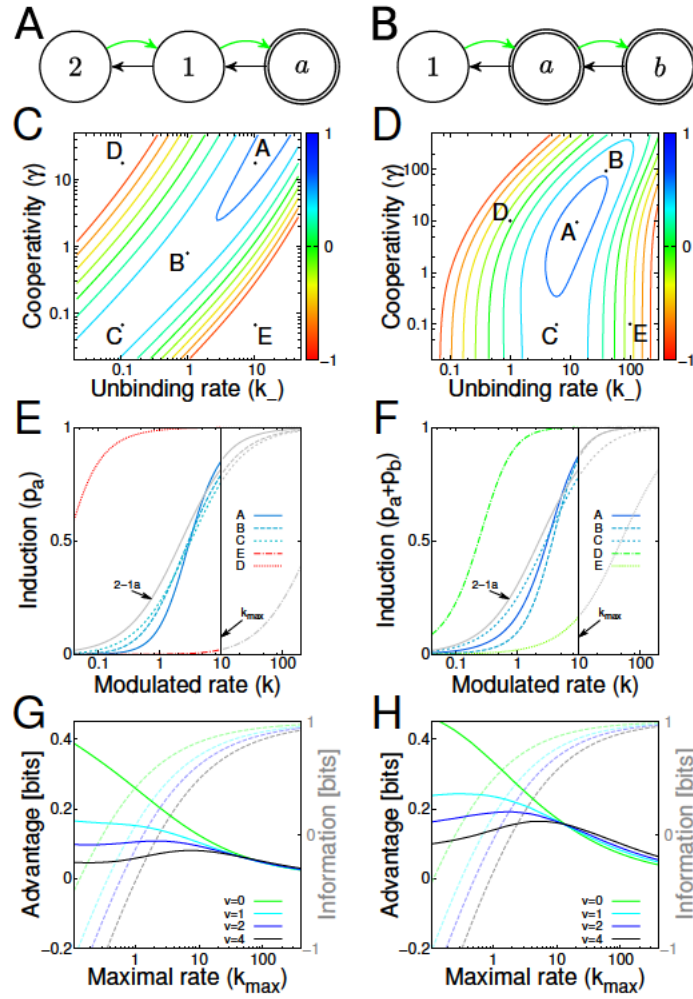


Figure 1.7: **Improving information transmission via cooperativity.** (A,B) The state transition diagram for the AND-architecture three-state promoter (at left, one active state) and OR-architecture (at right, two active states). (C, D) The information planes showing the channel capacity (color code) for various combinations of k_- and γ at a fixed maximum allowed $k_{max} = 10$, for the AND- and OR-architectures and $v = 2$. (E, F) The regulatory functions of various models selected from parameters denoted by dots in the information planes in (B, C). Solutions that maximize the information denoted with a solid blue line. Colors indicate the channel capacities. The gray part of the regulation curves cannot be accessed, since the input only ranges over $k \in [0, k_{max}]$. For comparison, we also plot the regulation curve of the 2-1a scheme (with its optimal k_-^*). (G, H) Channel capacity I_3^* (dashed lines, axes at right) and the advantage of the multi-state scheme over the best two-state promoter, $I_3^* - I_2^*$ (solid lines, axes at left), of the AND and OR models, as a function of the maximal input range, k_{max} , and the strength of the input noise (v , color). Since there is no globally optimal choice for γ for the AND-architecture, we fix $\gamma = 10$ in (G), and optimize only over k_- values.

1.7.2 Regulation with dual-role transcription factors

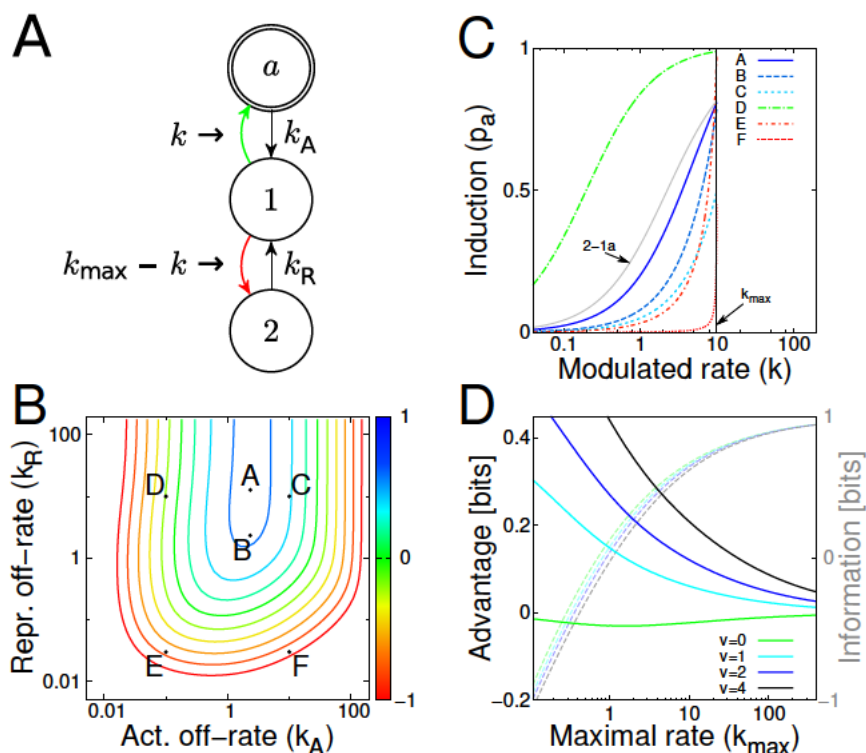


Figure 1.8: **Improving information transmission via dual-role regulation.** **(A)** The signal k increases the concentration of TFs in the activator role that favor the transition (green) to the expressing state, while simultaneously decreasing the rate of switching (red) into the inactive state with the repressor bound. **(B)** Channel capacity (color) as a function of off-rates (k_A, k_R) shows a peak at A. $k_{\max} = 10$ and $v = 4$. **(C)** The regulatory functions for the optimal solution A (solid blue line) and other example points (B-E) from the information plane, show that this architecture can access a rich range of response steepnesses and induction thresholds. For comparison, we also plot the regulation curve of the scheme 2-1a (with its optimal k^*). **(D)** The channel capacity (dashed line) and information advantage over the optimal two-state architecture (solid line), as a function of k_{\max} .

In the second architecture that we consider a transcription factor can switch its role from repressor to activator, depending on the covalent TF modification state or formation of a complex with specific co-factors. A well-studied example is in Hedgehog (Hh) signaling, where the TF Gli acts as a repressor when Hh is low, or as an activator when Hh is high [Parker *et al.*, 2011; White *et al.*, 2012; Müller and Basler, 2000]. Figure 1.8A shows a possible dual-role signaling scheme where the total concentration of dual-role TFs is fixed (at k_{\max}), but the signal modulates the fraction of these TFs that play the activator role (k) and the remaining fraction that

act as repressors ($k_{\max} - k$), which compete for the same binding site. The channel capacity of this motif is depicted in Fig 1.8B as a function of promoter parameters k_A and k_R , showing that a globally optimal setting (denoted “A”) exists for these parameters; with these parameters, the input/output function, shown in Fig 1.8C, is much steeper than what could be achieved with the best two-state promoter, and that is true despite the fact that the molecular implementation of this architecture uses only a single binding site. The ability to access such steep regulatory curves allows this architecture to position the mid-point of induction at higher inputs k , thus escaping the detrimental effects of the input noise at low k , while still being able to induce almost completely (i.e., make use of the full dynamic range of outputs) as the input varies from 0 to k_{\max} . This is how the dual-role regulation can escape the tradeoff faced by the two-state model 2-1a (shown in Fig 1.6). Sharper transition at higher input would lead us to expect that the advantage of this architecture over the two-state model is most pronounced when input noise is dominant (small k_{\max} , large v), which is indeed the case, as shown in Fig 1.8D.

1.7.3 Promotor cycling

In the last architecture considered here, promoters “cycle” through a sequence of states in a way that does not obey detailed balance, e.g., when state transitions involve expenditure of energy during irreversible reaction steps. In the scheme shown in Fig 1.9A, the regulated transition puts the promoter into an active state a ; before decaying to an inactive state, the promoter must transition through another active state b . Effectively, this scheme is similar to the two-state model in which the decay from the active state is not first-order with exponentially distributed transition times, but rather with transition times that have a sharper peak. The benefits of this architecture are maximized when the transition rates from both active states are equal. While it always outperforms the optimal two-state model, the largest advantage is achievable for small k_{\max} . At large k_{\max} the advantage tends to zero: this is because the optimal off-rates are high, causing the dwell times in the expressing states to be short. In this regime the gamma distribution of dwell times (in a three-state model) differs little from the exponential distribution (in a two-state model). Note that this model would not yield any information advantage if the state transitions were reversible.

Figures 1.9C, D show that irreversible transitions alone do not generate an information advantage: a promoter that needs to transition between two inactive states (1, 2) to reach a single expressing state a from which it exits in a first-order transition, is always at a loss compared to a two-state promoter. This is because here the effective transition rate to the active state in the equivalent two-state model is lower (since an intermediate state must be traversed to induce),

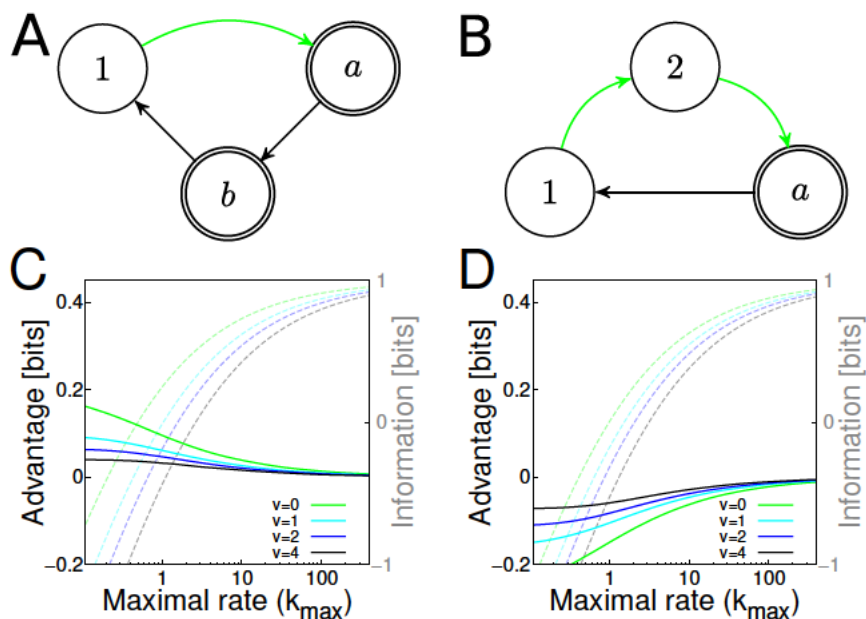


Figure 1.9: **Improving information transmission via cycling.** **(A)** Promoter cycles through two active states (a, b) expressing at identical production rates before returning to inactive state (1), from which the transition rate back into the active state (green) is modulated. For each value of k_{\max} , we look for the optimal choice of $k_{ab} = k_{b1}$. The information and advantage relative to the optimal two-state model is shown in **(C)**. **(B,D)** A similar architecture where turning the gene on is a multi-step regulated process. This architecture always underperforms the optimal two-state model, indicated by a negative value of the advantage for all choices of v and k_{\max} .

necessitating the use of a lower off-rate k_{a1} , which in turn leads to higher switching noise.

It is interesting to note that recent experimental data on eukaryotic transcription seem to favor models in which the distribution of exit times from the expressing state is exponential, while the distribution of times from the inactive into the active state is not [Suter *et al.*, 2011], pointing to the seemingly underperforming architecture of Fig 1.9B. The three-state promoter suggested here is probably an oversimplified model of reality, yet it nevertheless makes sense to ask why irreversible transitions through multiple states are needed to switch on the transcription of eukaryotic genes [Larson, 2011; Yosef and Regev, 2011] and how this observation can be reconciled with the lower regulatory power of such architectures. This is the topic of our ongoing research.

1.8 Experimentally measured promoter switching rates

Direct measurements of switching rates are rare since they require live imaging. Examples include the relative measurements of on-, off- and mRNA-production rates in *E. coli* [So *et al.*, 2011] using the MS2-GFP system [Golding *et al.*, 2005], reporting 2 – 10 fold higher on- than off-rates, and mRNA production rates an order of magnitude higher than the on-rates; original bursting reported in [Golding *et al.*, 2005] finds the on-time duration to be roughly 6 and the off time 37 minutes in a synthetic *E. coli* reporter system. Recently, on-rates of $\sim 3 \cdot 10^{-2} \text{ min}^{-1}$, roughly ten-fold higher off-rates, and mRNA production rates ranging from 0 – 5 min^{-1} have been reported in mammalian cells using the luciferase reporter system [Suter *et al.*, 2011]. Using new high-throughput microfluidic methods, it is now possible to measure TF binding and unbinding times directly: [Geertz *et al.*, 2012] reports mouse and yeast in vitro transcription factor dissociation rates between $\sim 10 \text{ s}^{-1}$ and 10^{-2} s^{-1} , as well as the range of the corresponding association rates; it is, however, less clear if these can be unambiguously identified with switching rates in functional models.

A larger body of work extracts the rates of the two-state model from the noise characteristics (which are the primary measurement), *assuming* the two-state model without diffusion noise is applicable. The reported Fano factors for mRNA counts vary, but are of the order of 1 – 10. The typical values for kinetic parameters extracted for a range of *E. coli* promoters are $10^{-3} - 10^{-2} \text{ s}^{-1}$ for the on-rate, $10^{-1} - 1 \text{ s}^{-1}$ for the mRNA production rate when induced, and a variable off-rate that depends strongly on the induction level [So *et al.*, 2011]. Using a similar technique in mammalian cells, [Raj *et al.*, 2006] extracted two-state parameters and found the on-rate normalized by mRNA decay time to be roughly of order unity, while the ratio of mRNA production rate to the off-rate varied from $\sim 10 - 400$, depending on the system and the induction level.

1.9 Langevin method and promoters as state-transition graphs

In this section we describe the general method used to derive the behavior of noise and mean for different promoter architectures, followed by a calculation for one example architecture.

1.9.1 Translating a state transition diagram into dynamic equations

Let $\{a, b, \dots M\}$ denote the states of the promoter that produce mRNA at a fixed rate r and $\{1, 2, \dots N\}$ denote states without production. For $\mathcal{S} = \{a, b, \dots M, 1, 2, \dots N\} \ni i, j$, let $k_{ij} \geq$

$0, i \neq j$ be the rate with which the promoter switches from state i to state j , d be the rate of mRNA-degradation, and p_i be the fractional occupancy of state i . For simplicity, we will only treat the case $M = 1$ here.

Deterministic equations. The list of (non-zero) rates fully defines the state-transition graph, i.e. the promoter model. This directly translates into a linear system of equations that describes the dynamics of the system:

$$\partial_t \mathbf{p} = \mathbf{K} \mathbf{p} \quad , \quad \text{with} \quad (1.13)$$

$$\mathbf{K} = \begin{bmatrix} -\sum_{j \in \mathcal{S}} k_{aj} & k_{1a} & \cdots & k_{Na} \\ k_{a1} & -\sum_{j \in \mathcal{S}} k_{1j} & \cdots & k_{N1} \\ \vdots & \vdots & \ddots & \vdots \\ k_{aN} & k_{1N} & \cdots & -\sum_{j \in \mathcal{S}} k_{Nj} \end{bmatrix} , \quad (1.14)$$

$\mathbf{p} = [p_a, p_1, \dots, p_N]^T$, subject to the normalization constraint $\sum_{i \in \mathcal{S}} p_i = 1$.

The dynamics of mRNA are described by linking them to the activity of the promoter:

$$\partial_t m = r p_a - d m \quad . \quad (1.15)$$

To compute the average amount of mRNA \bar{m} in steady state, we set the time derivatives to 0 and solve the linear set of equations

$$\mathbf{K} \bar{\mathbf{p}} = 0 \quad , \quad (1.16)$$

$$\bar{m} = \frac{r}{d} p_a \quad . \quad (1.17)$$

As the occupancy of the active state p_a is a function of the rates in \mathbf{K} , we can obtain the dependence of \bar{m} on any rate of interest, i.e. we can obtain the regulation function.

Langevin approach to calculate noise behavior. For the noise behavior, we linearize Eqs (4,5) of the main text around the mean:

$$\mathbf{p}(t) = \bar{\mathbf{p}} + \delta \mathbf{p}(t), \quad (1.18)$$

$$m(t) = \bar{m} + \delta m(t) \quad (1.19)$$

and introduce the Fourier-transformed variables

$$\delta p_i(t) = (2\pi)^{-1} \int d\omega \delta \hat{p}_i(\omega) \exp(-i\omega t) \quad , \quad (1.20)$$

$$\delta m(t) = (2\pi)^{-1} \int d\omega \delta \hat{m}(\omega) \exp(-i\omega t) \quad , \quad (1.21)$$

so that we get the linear response to random fluctuations:

$$(-i\omega)\delta\hat{p} = \mathbf{K}\delta\hat{p} + \hat{\xi} \quad , \quad (1.22)$$

$$(-i\omega)\delta\hat{m} = r\delta\hat{p}_a - d\delta\hat{m} + \hat{\xi}_m \quad . \quad (1.23)$$

The statistics of the Langevin forces are given by:

$$\langle \hat{\xi}_i^* \hat{\xi}_j \rangle = -(\hat{p}_i K_{ij} + \hat{p}_j K_{ji}) \quad , \quad (1.24)$$

$$\langle \hat{\xi}_m^* \hat{\xi}_m \rangle = 2d\bar{m} \quad ; \quad (1.25)$$

to see this for the variances, consider $\langle \xi_i^* \xi_i \rangle = -2\hat{p}_i K_{ii} = 2\hat{p}_i \sum_j k_{ij}$, since all entries in the diagonal of \mathbf{K} are negative. This is two times the rate of leaving state i . Similarly, for $\langle \xi_m^* \xi_m \rangle$ the variance is two times the rate of degrading a molecule. The factor of two comes from the fact that we consider a system at steady state, so the rates of entering and leaving a state (or creating an destroying a molecule) must be equal. For the covariances $\langle \xi_i^* \xi_j \rangle$ ($i \neq j$), the two Langevin forces are anti-correlated, since leaving one state means entering another. The rate of changing between the two states is the probability of being in state i (p_i) times the rate of transition from that state into the other ($k_{ij} = K_{ij} - K_{ji}$) – and this holds for both directions between the pair of states. Also, since we assume that production of mRNA and promoter switching are independent, $\langle \xi_i^* \xi_m \rangle = 0$ for all states i .

To get the variance in mRNA, we compute $\sigma_m^2 = (2\pi)^{-1} \int d\omega |\delta\hat{m}(\omega)|^2$, where $\delta\hat{m}(\omega)$ is obtained from Eq (1.23) as

$$\langle \delta\hat{m}^* \delta\hat{m} \rangle = \frac{2d\bar{m}}{d^2 + \omega^2} + \frac{r^2}{d^2 + \omega^2} \langle \delta\hat{p}_a^* \delta\hat{p}_a \rangle \quad , \quad (1.26)$$

where $\langle \delta\hat{p}_a^* \delta\hat{p}_a \rangle$ is calculated by solving Eq (1.22) and using the Langevin noise magnitudes from Eqs (1.24,1.25).

With the assumption $d \ll k_{ij}$, Eq (1.23) becomes

$$0 = \mathbf{K}(\delta\hat{p}) + \hat{\xi} \quad , \quad (1.27)$$

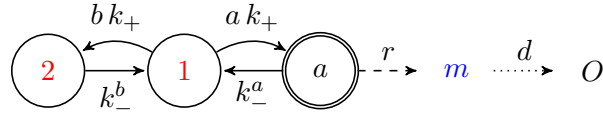
$$\sum_i \delta\hat{p}_i = 0 \quad , \quad (1.28)$$

which simplifies the expressions for the $\delta\hat{p}_i$. This is because the terms with $(-i\omega)$ in the denominator (as seen in the next section in Eqs (1.48,1.49)) would give an additional, multiplicative term of the form $1/(k_{ij}^2 + \omega^2)$ in Eq (1.26). The ω -dependence of these terms can be neglected for the integration, since for $d \ll k_{ij}$ we have

$$\frac{1}{k_{ij}^2 + \omega^2} \frac{1}{d^2 + \omega^2} \approx \frac{1}{k_{ij}^2} \frac{1}{d^2 + \omega^2} \quad . \quad (1.29)$$

1.9.2 Example: Dual regulation (3E-1a)

We are interested in a system where the promoter of a gene can either be occupied by an activator (present at concentration a) or a repressor (present at concentration b). If it is in the active state, it produces mRNA at a constant rate r , which is later degraded at rate d .



Deterministic equations. Following the setup from the last section, we translate the state transition diagram into a matrix that describes the dynamics at the promoter:

$$\begin{bmatrix} \partial_t p_a \\ \partial_t p_1 \\ \partial_t p_2 \end{bmatrix} = \begin{bmatrix} -k_-^a & ak_+ & 0 \\ k_-^a & -(ak_+ + bk_+) & k_-^b \\ 0 & bk_+ & -k_-^b \end{bmatrix} \cdot \begin{bmatrix} p_a \\ p_1 \\ p_2 \end{bmatrix} . \quad (1.30)$$

This is then the basis for a description of the dynamics of the output (here mRNA):

$$\partial_t m = rp_a - dm \quad , \quad (1.31)$$

$$\partial_t p_a = ak_+p_1 - k_-^a p_a \quad , \quad (1.32)$$

$$\partial_t p_2 = bk_+p_1 - k_-^b p_2 \quad , \quad (1.33)$$

$$p_a + p_1 + p_2 = 1 \quad . \quad (1.34)$$

With the definitions $A = \frac{ak_+}{k_-^a}$, $B = \frac{bk_+}{k_-^b}$, $S = 1 + A + B$ and $R = \frac{r}{d}$ we get for the steady state:

$$\bar{p}_a = A/S, \quad \bar{p}_1 = 1/S, \quad \bar{p}_2 = B/S \quad , \quad (1.35)$$

$$\bar{m} = R\bar{p}_a = RA/S \quad . \quad (1.36)$$

Langevin approach. To see how the dynamics of the promoter influence the statistics of mRNA we perturb the systems with Langevin forces (while still keeping the gene copy number constant):

$$\partial_t m = rp_a - dm + \xi_m \quad , \quad (1.37)$$

$$\partial_t p_a = ak_+p_1 - k_-^a p_a + \xi_a \quad , \quad (1.38)$$

$$\partial_t p_2 = bk_+p_1 - k_-^b p_2 + \xi_2 \quad , \quad (1.39)$$

$$p_a + p_1 + p_2 = 1 \quad . \quad (1.40)$$

The mean of the Langevin forces is zero ($\langle \xi_i(t) \rangle = 0$) and they are uncorrelated in time ($\langle \xi_i(t) \xi_i(t') \rangle \propto \delta(t - t')$).

We linearize around the mean, where deviations from the mean are denoted by δ :

$$m(t) = \bar{m} + \delta m(t) \quad , \quad (1.41)$$

$$p_a(t) = \bar{p}_a + \delta p_a(t) \quad , \quad (1.42)$$

$$p_2(t) = \bar{p}_2 + \delta p_2(t) \quad , \quad (1.43)$$

$$\delta p_1 = -\delta p_a - \delta p_2 \quad . \quad (1.44)$$

After inserting the linearized equations into the Langevin approach we perform a Fourier transform:

$$-i\omega \delta \hat{m} = r \delta \hat{p}_a - d \delta \hat{m} + \hat{\xi}_m \quad , \quad (1.45)$$

$$-i\omega \delta \hat{p}_a(\omega) = ak_+(-\delta \hat{p}_2 - \delta \hat{p}_a) - k_-^a \delta \hat{p}_a + \hat{\xi}_a \quad , \quad (1.46)$$

$$-i\omega \delta \hat{p}_2(\omega) = bk_+(-\delta \hat{p}_a - \delta \hat{p}_2) - k_-^b \delta \hat{p}_2 + \hat{\xi}_2 \quad . \quad (1.47)$$

Starting with the equations for the occupancies, we rewrite Eqs (1.46,1.47) and use the approximation that d is significantly slower than the other rates to get:

$$\delta \hat{p}_a(\omega) = \frac{ak_+ \delta \hat{p}_1 + \hat{\xi}_a}{k_-^a - i\omega} \approx A \delta \hat{p}_1 + \frac{\hat{\xi}_a}{k_-^a} \quad , \quad (1.48)$$

$$\delta \hat{p}_2(\omega) = \frac{bk_+ \delta \hat{p}_1 + \hat{\xi}_2}{k_-^b - i\omega} \approx B \delta \hat{p}_1 + \frac{\hat{\xi}_2}{k_-^b} \quad , \quad \text{or} \quad (1.49)$$

$$\delta \hat{p}_a = -\delta \hat{p}_2 \frac{A}{(1+A)} + \frac{\hat{\xi}_a}{k_-^a} \frac{1}{(1+A)} \quad , \quad (1.50)$$

$$\delta \hat{p}_2 = -\delta \hat{p}_a \frac{B}{(1+B)} + \frac{\hat{\xi}_2}{k_-^b} \frac{1}{(1+B)} \quad . \quad (1.51)$$

Solving this system yields:

$$\delta \hat{p}_a = -\frac{\hat{\xi}_2}{k_-^b} \bar{p}_a + \frac{\hat{\xi}_a}{k_-^a} (\bar{p}_1 + \bar{p}_2) \quad . \quad (1.52)$$

The variances of the Langevin forces are:

$$\langle \hat{\xi}_a^* \hat{\xi}_a \rangle = 2k_-^a \bar{p}_a \quad , \quad (1.53)$$

$$\langle \hat{\xi}_2^* \hat{\xi}_2 \rangle = 2k_-^b \bar{p}_2 \quad , \quad (1.54)$$

$$\langle \hat{\xi}_m^* \hat{\xi}_m \rangle = 2d\bar{m} \quad , \quad (1.55)$$

and their covariances vanish, since the direct transition from state a to state 2 is not allowed.

From Eqs (1.45,1.52) we get:

$$\langle \delta \hat{p}_a^* \delta \hat{p}_a \rangle = 2 \frac{\bar{p}_2}{k_-^b} \bar{p}_a^2 + 2 \frac{\bar{p}_a}{k_-^a} (\bar{p}_1 + \bar{p}_2)^2 \quad , \quad (1.56)$$

$$\langle \delta \hat{m}^* \delta \hat{m} \rangle = \frac{2d\bar{m}}{d^2 + \omega^2} + \frac{r^2}{d^2 + \omega^2} \langle \delta \hat{p}_a^* \delta \hat{p}_a \rangle \quad . \quad (1.57)$$

Finally, with $\frac{1}{2\pi} \int_{-\infty}^{\infty} 2 \frac{1}{x^2 + \omega^2} d\omega = \frac{1}{x}$ we get:

$$\begin{aligned} \sigma_m^2 &= \frac{d\bar{m}}{d} + \frac{r^2}{d} \left(\frac{p_2}{k_-^b} p_a^2 + \frac{p_a}{k_-^a} (p_1 + p_2)^2 \right) = \\ &= \bar{m} \left[1 + r \left(\frac{p_2}{k_-^b} p_a + \frac{1}{k_-^a} (1 - p_a)^2 \right) \right] . \end{aligned} \quad (1.58)$$

This is one description of noise in the 3E architecture. To get the noise characteristics for modulation scheme 3E-1a, we need to express p_2 in terms of p_a (not shown). From Eq (1.58) we can see that in the absence of repressors ($p_2 = 0$) and also for very fast unbinding of the repressors ($k_-^b \rightarrow \infty$) the noise shows the quadratic dependence on the occupation of the promoter that we see in the corresponding two-state model 2-1a.

1.9.3 Comparison to other methods

The results obtained with the Langevin approach were compared against two other methods: (i) the exact numerical solution of the chemical master equation and (ii) results from stochastic simulation using the Gillespie algorithm. Two kinds of comparisons are relevant: first, how well the gaussian distribution approximates the true distribution of mRNA levels; and second, how the Langevin-derived expressions for the noise characteristics compare to the exact values.

Fig 1.10A compares the distribution of mRNA levels obtained from the numerical solution of the chemical master equation to the gaussian approximation for the dual regulation architecture discussed in the last section.

The stochastic simulation algorithm is time consuming and offers no special benefit for the simple systems studied here, but we have nevertheless checked a few example architectures against simulation results. The results for dual regulation are shown in Fig 1.10B. Values for ak_+ and k_-^a were chosen from a grid. This makes it possible to show the agreement with the Langevin-derived noise characteristics in two different modulation schemes (cf. inset in Fig 1.10B).

Another way to obtain analytical expressions for the mean and variance of the mRNA-distributions is the method of partial moments (e.g., [Sanchez *et al.*, 2011a; Sanchez and Kondev, 2008]). While this method can also be used to derive higher moments, a minor advantage of the Langevin method for the purposes here is that the approximation $d \ll k_{ij}$ can be used earlier in the derivations, leading to simpler expressions.

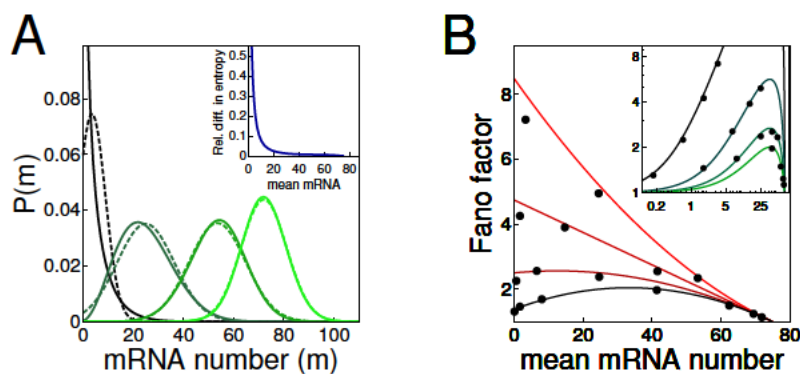


Figure 1.10: **Comparison of Langevin approach to other methods.** **(A)** Probability distribution of the model 3E for different values of ak_+ (thus treating it as 3E-1a). Solid lines are the numerical solution of the master equation and dashed lines are a gaussian approximation using the analytical expressions for mean and variance described in the main text. Different colors are for different values of ak_+ . Parameters for this plot: $ak_+ = [1, 10, 50, 500]$, $k_-^a = 10$, $bk_+ = 10$, $k_-^b = 10$, $r = 75$ and $d = 1$. Inset: The relative difference in entropy between the full distribution and the gaussian approximation used in this study (i.e. using the results for mean and variance from the Langevin method). The error drops to $< 5\%$ very fast, with the main difference at slow rates stemming mainly from non-gaussianity at low expression levels. Importantly, for information calculations this non-gaussianity is mitigated by the protein averaging; even without the averaging, the effect on the comparison between architectures is minor. **(B)** Comparison of Fano factors for different mean expression levels for the 3E-1a model (black data points = Gillespie simulation, solid lines = Langevin method). Different colors correspond to different values for k_-^a . For small values of k_-^a and ak_+ we start to see deviations since the approximation $d \ll k_{ij}$ no longer holds. Parameters for the Gillespie simulations: $k_-^a = [10, 20, 50, 200]$, $ak_+ = [1, 10, 50, 500]$, $bk_+ = 10$, $k_-^b = 10$, $r = 75$ and $d = 1$. The Fano factor was calculated from 10000 runs. Error bars from resampling are smaller than the symbols. Inset: The same results from the Gillespie simulations replotted (on a log-log scale) and compared to the noise characteristics of the 3E-a1 scheme, i.e. the rate k_-^a is modulated to obtain different mean mRNA levels. The solid lines are the noise characteristics calculated with the Langevin methods for different (fixed) values of $ak_+ = [1, 10, 50, 500]$.

1.10 Discussion

When studying noise in gene regulation one is usually restricted to the use of phenomenological models, rather than a fully detailed biochemical reaction scheme. Simpler models, such as those studied in this paper, also allow us to decouple questions of mechanistic interpretation

from the questions of functional consequences. Here, we extended a well-known functional two-state model of gene expression to multiple internal states. We introduced state transition graphs to model the “decision logic” by which changes in the concentrations of regulatory proteins drive the switching of our genes between various states of expression. This abstract language allowed us to systematically organize and explore non-equivalent three-state promoters. The advantage of this approach is that many microscopically distinct regulatory schemes can be collapsed into equivalent classes sharing identical state transition graphs and identical information transmission properties.

The functional description of multi-state promoters confers two separate benefits. First, it is able to generate measurable predictions, such as the noise vs mean induction curve. Existing experimental and theoretical work using the two-state model has demonstrated how the measurements of noise constrain the space of promoter models [So *et al.*, 2011], how the theory establishes the “vocabulary” by which various measured promoters can be classified and compared to each other [Zenklusen *et al.*, 2008], and how useful a baseline mathematical model can be in establishing quantitative signatures of deviation which, when observed, must lead to minimal model revisions able to accommodate new data [Suter *et al.*, 2011]. Alternative complex promoters presented here could explain existing data better either because of the inclusion of additional states (c.f. [Sanchez *et al.*, 2011a]), or because we also included and analyzed the effects of input (diffusive) noise, which can mimic the effects of promoter switching noise but is often neglected [Tkačik *et al.*, 2008c]. As a caveat, it appears that in many cases discriminating between promoter architectures based on the noise characteristics alone would be very difficult, and thus dynamical measurements would be necessary.

The second benefit of our approach is to provide a convenient framework for assessing the functional impact of noise in gene regulation, as measured by the mutual information between the inputs and the gene expression level. We were interested in the question whether multi-state promoters can, at least in principle, perform better than the simple ON/OFF two-state model. We find that generically, i.e., for all three-state models where one state is expressing and only one transition is modulated by the input, the multi-state promoters underperform the two state model. Higher information transmission can be achieved when these conditions are violated, and biological examples for such violations can be found. For example, we find that a multi-state promoter with cooperativity has a higher channel capacity than the best comparable two-state promoter, even when promoter switching noise is taken into account (c.f. [Tkačik *et al.*, 2009]). Dual-regulation yields surprisingly high benefits, which are largest when input noise is high. In the context of metazoan development where the concentrations of the mor-

phogen molecules can be in the nanomolar range and the input noise is therefore high [Gregor *et al.*, 2007], the need to establish sharp spatial domains of downstream gene expression (as observed, [Dessaud *et al.*, 2008]) might have favored such dual-role promoter architectures. Lastly, we considered the simplest ideas for a promoter with irreversible transitions and have shown that they can lead to an increase in information transmission by sharpening the distribution of exit times from the expressing state [Pedraza and Paulsson, 2008].

The main conclusion of this article – namely that channel capacity can be increased by particular complex promoters – is testable in dedicated experiments. One could start with a simple regulatory scheme in a synthetic system and then by careful manipulation gradually introduce the possibility of additional states (e.g., by introducing more binding sites), using promoter sequences which show weaker binding for individual molecules yet allow for stronger cooperative interaction. In both the simple and complex system one could then measure the noise behavior for various input levels. Information theoretic analysis of the resulting data could be used to judge if the design of higher complexity, while perhaps noisier by some other measure, is capable of transmitting more information, as predicted.

The list of multi-state promoters that can outperform the two-state regulation and for which examples in nature could be found is potentially much longer and could include combinations of features described in this article. Rather than trying to find more examples, we should perhaps ask about the fundamentally different mechanisms and constraints that our analysis did not consider. In all cases that we analyzed the largest difference between the two- and three-state models was at low k_{\max} . This makes sense: at high k_{\max} the dominant source of noise is the (bursty) Poisson production of gene products, which is the same regardless of the promoter architecture, while at low inputs, the input fluctuations filter through the promoter in ways that depend on its architecture. What other tricks could biology use to cope with input noise? By expending energy to keep the system out of equilibrium, one could design robust reaction schemes where, for example, the binding of a regulatory protein leads (almost) deterministically to some tightly controlled response cycle, perhaps evading the diffusion noise limit [Aquino and Endres, 2010] and increasing information transmission. At the same time, cells might be confronted by sources of stochasticity we did not discuss here, for example, due to cross-talk from spurious binding of non-cognate regulators. Finally, cells need to not only *transmit* information through their regulatory elements, but actually perform *computations*, that is, combine various inputs into a single output, thereby potentially discarding information. A challenging question for the future is thus about extending the information-theoretic framework to these other cases of interest.

2 The broadcasting cell

The work presented in this chapter was performed in collaboration with Sarah Cepeda-Humerez and Gašper Tkačik and was partly published in Physical Review Letters (see [Cepeda-Humerez *et al.*, 2015]). This part is reproduced here with small changes in section 2.3. The software to calculate the data for Figures 2.4, 2.5, and 2.6 was written by Sarah Cepeda-Humerez.

2.1 Introduction

The mathematical machinery of modeling promoters as state transition diagrams used in the previous chapter can also be used to address a range of fundamental questions in cell signaling: Given the largely non-compartmentalized nature of the cell, we can ask how the presence of many different signals influences signaling through one particular channel under consideration. Is there a limit to how much information can be transmitted through a channel, given that many other signals are broadcast at the same time?

For a single channel in isolation, more reliable signaling can generally be achieved by using more signaling molecules [Tkačik *et al.*, 2009]. This, however, can carry a metabolic burden, or slow down the adaptation of a system to a new condition. In this section, we show that these costs are not only the ones that might be relevant in shaping cellular signaling networks. Rather, we argue that there is a cost intrinsic to the presence of many signaling molecules in the same compartment. Given that these signals are broadcast and have only a finite specificity for their target [von Hippel *et al.*, 1974; Bird, 1995; Gerland *et al.*, 2002; Johnson *et al.*, 2005; Maerkl and Quake, 2007; Wunderlich and Mirny, 2009; Rockel *et al.*, 2013; Todeschini *et al.*, 2014] , we show that this interference effect sets a scale for useful amounts of signaling molecules. This is therefore a global effect, not specific to a single gene – rather something the whole cell as a system is constrained by.

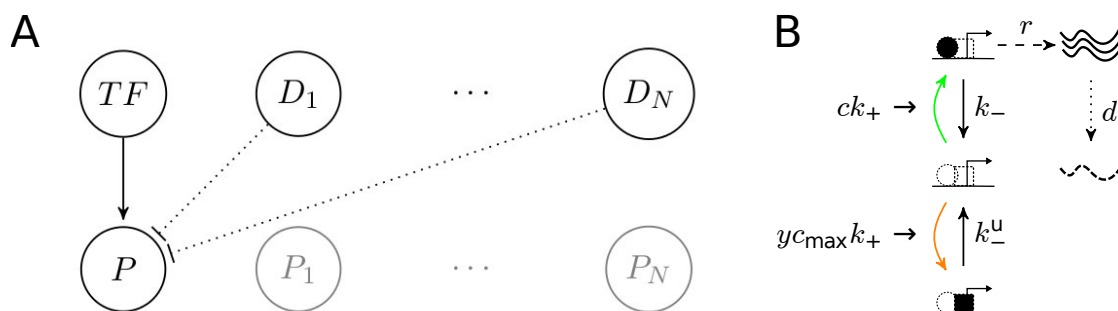


Figure 2.1: **(A)** The presence of many different signals (TF, D_1, \dots, D_N with cognate signal recipients P, P_1, \dots, P_N) in the same compartment poses a challenge to filter the cognate signal for a certain channel under consideration ($TF \rightarrow P$) if there is some residual activity in non-cognate channels ($D_1 \rightarrow P, \dots, D_N \rightarrow P$). **(B)** The 3-state model that can be used to represent the situation from (A). The specific input (TF) has a restricted range $c \in [0, c_{\max}]$ and an unbinding rate k_- . Other channels (D_1, \dots, D_N) are restricted by the same c_{\max} , where their influence $y = N\bar{c}$ is a combination of the number of channels N and their typical concentrations \bar{c} . Non-specific binding is characterized by an off-rate k_-^u .

To build an intuition, we first remark that the direct effect of spuriously binding proteins (termed decoys in the remainder) is not important for noise considerations. We can use the simple 3-state model investigated in section 1.9.2 from the previous chapter. In this model, we interpret state 2 as the promoter being occupied by a non-specific decoy molecule. From this, we can conclude that if an individual decoy molecule binds significantly shorter than the specific transcription factors, the second term in equation (1.58) does not contribute much to the total noise since $1/k_{off}^b \ll 1/k_{off}^a$. This is true when comparing the noise for the same mean expression level. A promoter, however, can not achieve the same mean expression in the presence of decoys given the same binding rate for cognate molecules. This causes an indirect effect of decoys on the switching noise, since we have to decrease the off-rate of the specific TF to get to the same mean, i.e. retain a useful dynamic range for signaling.

Since it is not possible to increase the number of transcription factor molecules indefinitely, the promoter occupancy of the decoys has to be counteracted through stronger binding of the (specific) activators, which increases the switching noise (see also section 1.3 in the previous chapter). This leads to the following model assumption: we assume that signaling molecules that act as decoys for one promoter are specific for some other promoter; thus, increasing the number of all transcription factors – since all are cognate to some channel – leaves the fraction of one particular species the same (see Figure 2.1).

2.2 A simple model demonstrates the detrimental effects of cross-talk on signaling fidelity

We can take the model 3E-1a from the previous chapter, as depicted in Figure 2.1B, and its noise model (still in dimensionless units):

$$\sigma_m^2 = \bar{m} \left[1 + r \frac{1}{k_{a1}} (1 - p_a) \left[1 - p_a \left(1 - \frac{k_{a1}}{k_{21}} p_2^0 \right) \right] + \frac{r}{d} \frac{1}{D} \frac{1}{k_{a1}} p_1^0 (1 - p_a)^3 \right] . \quad (2.1)$$

To adapt it to our purposes, we identify the rates with the ones from Figure 2.1B: $ck_+ = k$, $k_{a1} = k_-$, $k_{21} = k_-^u$. Most importantly we add the following relationship between the rates to represent the presence of the decoys: $k_{12} = yk_{\max}$, where y is a factor summarizing affinity and copy number of decoy molecules from other channels. Note that their influence will scale with the maximally allowed number of signaling molecules for the specific channel k_{\max} . Setting $p_2^0 = \frac{yk_{\max}}{k_-^u + yk_{\max}}$ and $p_1^0 = 1 - p_2^0$ completes the noise model. The upper limit for the integration to calculate the mutual information is $p_a^{\max} = k_{\max} / \left[k_- + k_{\max} \left(1 + y \frac{k_-}{k_-^u} \right) \right]$.

With this noise model at hand, we can now ask how increasing the number of signaling molecules improves information transmission capacity for a given number of interfering channels. Figure 2.2 shows that the presence of additional signals can have a significant influence on the information capacity of a channel. Specifically, Figure 2.2B suggests that there is a limit up to which increasing the number of signaling molecules can help increase information capacity. Viewed from the perspective of the whole cell, this then suggests that there is an upper limit for the number of signaling molecules that can usefully be employed in a typical channel. This upper limit is not caused by factors external to the signaling apparatus, such as the metabolic cost of producing the molecules, but rather by an intrinsic property of the system.

2.3 Kinetic proofreading alleviates deleterious effects of cross-talk

Faced with the problem of interfering channels, improved specificity of the signals to their cognate channels can alleviate the problem of cross-talk. In this section, we describe a way to improve the insulation between channels by increasing the ability of a regulatory sequence to distinguish between cognately and non-cognately binding transcription factors. This increased distinguishability does not necessarily have to come from a larger difference in binding energies between cognate and non-cognate molecules if we allow for irreversible transitions in the state-transition-graph that models the process of gene expression. The results here therefore are

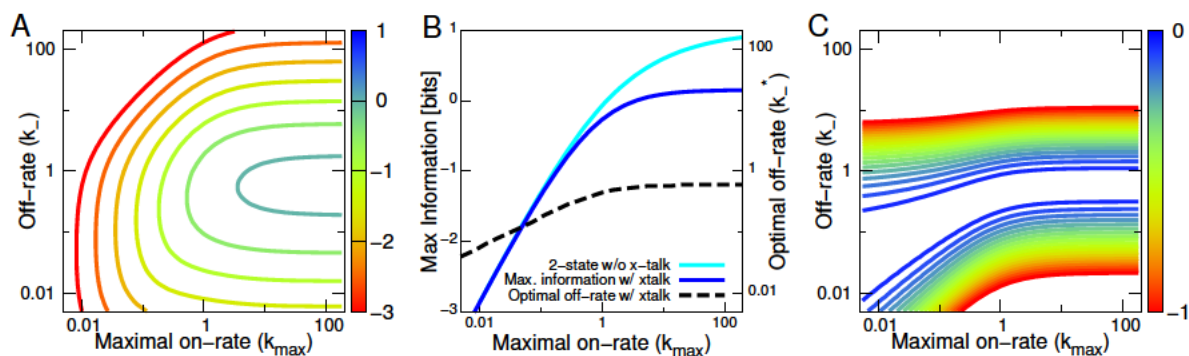


Figure 2.2: **(A)** Channel capacity calculated with the small noise approximation in the broadcasting cell as a function of the on- and off-rates for the specific TFs. Parameters: $k_{-}^u = 100$, $y = 100$, $v = 1$. **(B)** Loss in information capacity due to decoy binding sites. *Left axis:* Optimal channel capacity for the broadcasting cell (blue) and for the 2-1a model representing a signaling channel with no cross-talk (cyan). Around $k_{\max} = 1$ interference from other signals is starting to incur an information cost. *Right axis:* The optimal off-rate for specific binding k_{-}^* (dashed black). **(C)** Sensitivity of information capacity to the choice of the specific unbinding rate k_{-} , expressed as the difference to the maximal information for a fixed k_{\max} .

another example for how out-of-equilibrium schemes, such as the cycling promoter in section 1.7.3 of the previous chapter, can outperform equilibrium schemes.¹

In prokaryotes, transcription factors recognize and bind specific DNA sequences $L = 10 - 20$ basepairs (bp) in length, usually located in promoter regions upstream of the regulated genes [Ptashne and Gann, 2002]. Regulation by a single TF, or a small number of TFs interacting cooperatively, is sufficient to quantitatively account for the experimental measurements of gene expression [Kuhlman *et al.*, 2007b], as well as to explain how any gene can be individually “addressed” and regulated only by its cognate TFs [Wunderlich and Mirny, 2009], without much danger of regulatory crosstalk. In eukaryotes, however, TFs seem to be much less specific ($L = 5 - 10$ bp; but the total genome size is larger than in prokaryotes by $\sim 10^3$) [Wunderlich and Mirny, 2009; Sandelin *et al.*, 2004], binding promiscuously to many genomic locations [Li *et al.*, 2008], including to their non-cognate binding sites [Rockel *et al.*, 2013]. What are the implications of this reduced specificity for the precision of gene regulation?

Thermodynamic models of regulation postulate that the rate of target gene expression is given by the equilibrium occupancy of various TFs on the regulatory sequence [Shea and Ack-

¹From here on, [Cepeda-Humerez *et al.*, 2015] is reproduced with minimal changes, and we change notation to stay consistent with the published version.

ers, 1985a; Bintu *et al.*, 2005], and the success of this framework in prokaryotes [Kinney *et al.*, 2010] has prompted its application to eukaryotic, in particular, metazoan, enhancers [Janssens *et al.*, 2006; He *et al.*, 2010; Fakhouri *et al.*, 2010]. To illustrate the crosstalk problem in this setting, consider the ratio σ of the dissociation constants to a nonspecific and a specific site for an eukaryotic TF; typically, $\sigma \sim 10^3$ (corresponding to a difference in binding energy of $\sim 7 k_B T$) [Maerkl and Quake, 2007; Rockel *et al.*, 2013]. Because there are $\nu \sim 10^2 - 10^3$ of different TF species in a cell, TFs nonspecific to a given site will greatly outnumber the specific ones. For an isolated binding site, this would imply roughly equal occupancy by cognate and noncognate TFs, suggesting that crosstalk could be acute. For multiple sites, cooperative binding is known for its role in facilitating sharp and strong gene activation even with cognate TFs of intermediate specificity – but could the same mechanism also alleviate crosstalk? First, note that there exist well-studied TFs which do not bind cooperatively (e.g. [Giorgetti, 2010]). Second, while many proposed regulation schemes give rise to cooperativity (e.g., nucleosome-mediated cooperativity [Mirny, 2010], or synergistic activation [Todeschini *et al.*, 2014]) they will not suppress crosstalk; for the latter, cooperativity needs to be strong and specific, stabilizing only the binding of *cognate* TFs. Third, even when cooperative interactions are specific, crosstalk can pose a serious constraint. Regulating a gene implies varying the cognate TF concentration throughout its dynamic range, and when this concentration is low and the target gene should be uninduced, cooperativity cannot prevent the erroneous induction by noncognate TFs. For that, the cell could either keep the genes inactive by binding of specific repressors, or by making the whole gene unavailable for transcription. The first strategy seems widely used in bacteria but less so in eukaryotes; the second strategy (“gene silencing”) is widespread in eukaryotes, but only happens at a slow timescale and involves a complex series of nonequilibrium steps.

Here we propose a plausible and fast molecular mechanism which alleviates the effects of crosstalk; a detailed account of when crosstalk poses a severe constraint for gene regulation is presented in [Friedlander *et al.*, 2015]. The proposed mechanism is consistent with the known tight control over which genes are expressed in different conditions or tissues (e.g., during development [McGinnis and Krumlauf, 1992]) on the one hand, and on the other, explains the high levels of measured noise in transcription initiation of active genes [Raj *et al.*, 2006; Little *et al.*, 2013].

The simplest proofreading architecture for transcriptional gene activation that can cope with erroneous binding is presented in Fig 2.3A,B, motivated by a scheme first proposed by Hopfield [Hopfield, 1974]. Specificity is only conveyed by differential rates of TF unbinding (“off-rates” k_-^c, k_-^{nc} , with $\sigma = k_-^{nc}/k_-^c$). There are ν noncognate TF species whose typical concentration we

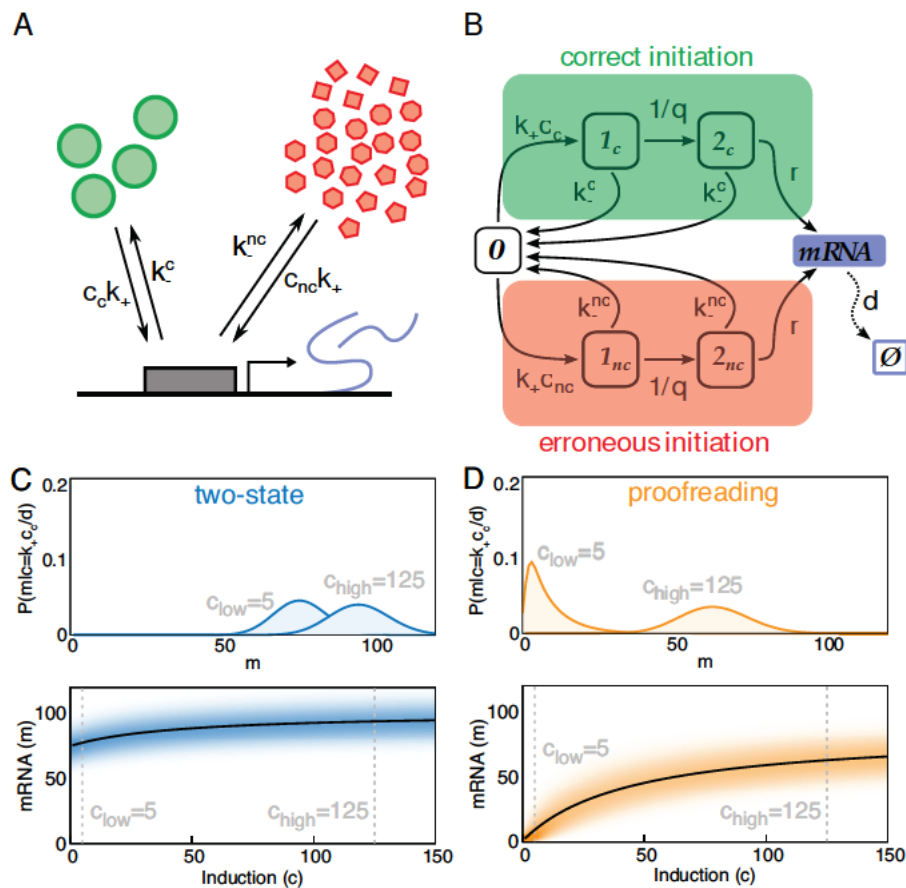


Figure 2.3: **(A)** A schematic of cognate (green circles) and ν kinds of noncognate (various red shapes) TFs binding to a gene regulatory element on the DNA (gray box), to control the mRNA expression level. **(B)** Transition state diagram for the proofreading gene regulation. The regulatory element can cycle between an empty state (0), state occupied by either cognate (1_c) or noncognate (1_{nc}) TF; to initiate gene expression, a further non-equilibrium transition into “2” states (with rate $1/q$) is required, driven by, e.g., hydrolysis of ATP. mRNA is expressed at rate r and degraded with rate d , the slowest process that sets our unit for time. In this figure we use $r/d = 100$, $k_-^{nc}/d = 2500$, $\sigma = 500$, $\nu = 50$, $\Lambda = \nu/\sigma = 0.1$; dimensionless concentration is $c = k_+c_c/d$. **(C,D)** Steady-state mRNA distributions for low and high concentrations of the cognate TF, c . As $qd \rightarrow 0$ (C), the proofreading model reduces to the two-state model of gene expression [Rieckh and Tkačik, 2014]; here, noncognate TFs initiate transcription at a high rate even when c is low, causing overlapping output distributions (blue; top) and small dynamic range (black line = $\langle m(c) \rangle$, blue shade = $\sigma_m(c)$; bottom). Proofreading (D) suppresses erroneous initiation, leading to separable output distributions (orange; top) and higher dynamic range (bottom).

take to be $c_{nc} = \frac{1}{2}\nu C$, and C is the maximal concentration for the cognate TFs $c_c, c_c \in [0, C]$. The ratio $\Lambda = \nu/\sigma$ determines the severity of crosstalk, which is weak for $\Lambda \ll 1$ and strong for $\Lambda \gg 1$. The response of the promoter to the dimensionless input concentration c ($= k_+c_c/d$, see Fig 2.3B) of cognate TFs is captured by the steady state distribution of mRNA, $P(m|c)$; the spread of this distribution is due to the stochasticity in gene expression, which includes random switching between promoter states and the birth-death process of mRNA expression [Peccoud and Ycart, 1995]. If the reaction rates are known, $P(m|c)$ is computable from the chemical Master equation corresponding to the transition diagram in Fig 2.3B; using finite-state truncation, this becomes a linear problem that is numerically tractable.

Figures 2.3C and D each compare the steady state distributions of mRNA at low and high concentration of cognate TF, c . The behavior crucially depends on the out-of-equilibrium rate qd . When $qd \rightarrow 0$, the scheme of Fig 2.3B becomes a normal two-state promoter as the states 1_c and 2_c (likewise 1_{nc} and 2_{nc}) fuse into a single state. In this limit, the effect of crosstalk is highly detrimental already at $\Lambda = 0.1$ used in this example: at low c , the promoter repeatedly cycles through erroneous initiation and the gene is highly expressed both at low c as well as at high c (where most of the expression is indeed due to correct initiation); as a result, the distributions $P(m|c)$ show substantial overlap in the two input conditions shown in Fig 2.3C. In contrast, for a non-trivial choice of q ($k_-^c \ll 1/q \simeq k_-^{nc}$), the model can exhibit proofreading. Even at low cognate concentration c , the slow irreversible transition ensures that noncognate TFs unbind from the promoter and that erroneous initiation is consequently rare, which is manifested as a sharp peak of $P(m|c_{low})$ at small m in Fig 2.3D. The proofreading architecture generates a larger output dynamic range and consequently makes the responses distinguishable.

What are the costs to the cell of the proposed proofreading mechanism? First, the mechanism requires an energy source, e.g., ATP, to break detailed balance. Whether such a metabolic cost is a burden to the cell is unclear: a few molecules of ATP paid per initiation should be negligible compared to the processive cost of transcription and translation. Second, however, is an indirect cost in terms of gene expression noise. While proofreading decreases erroneous induction, it takes longer to traverse the state transition diagram from empty state 0 to expressing state 2, and since the promoter can perform aborted erroneous initiation cycles, the fluctuations in the time-to-induction will also increase [Bel *et al.*, 2009]. This will result in additional variance in the mRNA copy number at steady state compared to the two-state ($qd \rightarrow 0$) scheme. While the speed/specificity tradeoff in protein synthesis has been examined before using deterministic chemical kinetics [Savir and Tlusty, 2013], this stochastic formulation of proofreading has, to our knowledge, remained unexplored. Proofreading in gene regulation is thus expected to

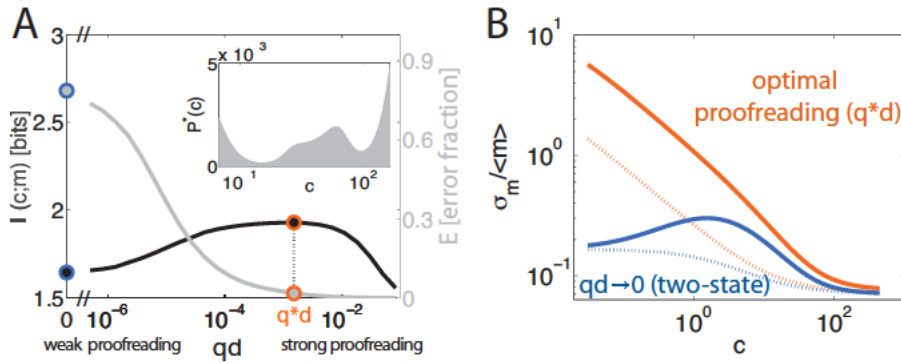


Figure 2.4: **(A)** Maximal information transmission (left axis, black) and the error fraction (right axis, gray) as a function of the inverse irreversible reaction rate, qd . Increasing qd suppresses the error fraction, but only at the cost of increasing the gene expression noise, leading to a tradeoff and an information-maximizing value of q^*d (orange). This maximum is reached robustly with input distributions that are close to optimal (inset). **(B)** Noise in gene expression, $\sigma_m / \langle m \rangle$, computed from the moments of $P(m|c)$, as a function of the dimensionless input concentration c , for the optimal proofreading (orange) and the two-state (blue) architectures. Dotted lines show the Poisson limit, $\sigma_m^2 = \langle m \rangle$, for comparison. In both cases, the average number of mRNA expressed is fixed to $\bar{m} = 100$.

increase the output dynamic range, which is favorable for signaling, but also to increase the noise, which is detrimental.

How can we formalize the tradeoff between noise and dynamic range for gene regulatory schemes and find when proofreading is beneficial? In existing analyses of proofreading the erroneous incorporation of the substrate leads to an error product that is *different* from the correct one [Hopfield, 1974; Savir and Tlusty, 2013]; in contrast, here the gene always expresses the *same* mRNA. What is important for signal transduction, however, is how well this expression correlates with the input signal, c . To quantify the regulatory power of the proofreading architecture, we computed the mutual information, $I(c; m)$ [Shannon and Weaver, 1949], between the signal c and the mRNA expression level m , following previous applications of information theory to gene regulation [Tkačik and Walczak, 2011; Rieckh and Tkačik, 2014]. The information depends not only on $P(m|c)$, which we compute from the Master equation, but also on the *a priori* unknown distribution of input concentrations, $P(c)$; we therefore determined the input distribution $P^*(c)$ that maximizes information transmission, subject to a constraint on the average number of expressed mRNA, $\bar{m} = \int dc P(c) \sum_m m P(m|c)$. This constraint on average number of mRNA was imposed to compare different regulatory architectures; otherwise, higher average expression could yield higher information transmission for trivial reasons. Such con-

strained information (capacity) maximization is a well-known problem in information theory that can be solved using the Blahut-Arimoto algorithm [Blahut, 1972].

Figure 2.4A shows how the information transmission $I(m; c)$ through the promoter depends on the (inverse) reaction rate qd . We start by looking at the classic measure of proofreading performance, the “error fraction,” i.e., the ratio of the mRNA expressed from state 2_{nc} due to noncognate TFs, vs mRNA expressed from state 2_c due to cognate TFs. As qd is increased, the error fraction drops, with no clear optimum. In contrast, there exists an optimal q^*d at which the information is maximized – this is the point where proofreading is most effective, optimally trading off erroneous induction (here, suppressed by a factor of ~ 30 relative to no proofreading), noise in gene expression, and dynamic range at the output. In Fig 2.4B we plot the noise in gene expression, as a function of the input concentration c for the optimal proofreading architecture and the non-proofreading limit. In both cases the noise has super-Poisson components due to the switching between promoter states, but this excess is substantially higher in the proofreading architecture, as expected.

While attractive, these results still depend on the particular rates chosen for the model in Fig 2.3B. Surprisingly, if we choose to compare the *optimal* proofreading scenario with the *optimal* non-proofreading one, the problem simplifies further. Given that the input TF concentration c varies over some limited dynamic range, $c \in [0, C_{\max} = k_+C/d]$, there should exist also an optimal setting for k_-^c : set too high, the cognate TFs will be extremely unlikely to occupy the promoter for any significant fraction of the time and induce the gene; set too low, the switching contribution to noise in gene expression will blow up. With k_-^c and q in the “correct initiation” pathway of Fig 2.3B set by optimization, the remaining rates in the “erroneous initiation” pathway are fixed by the choice of crosstalk severity Λ . The remaining parameters regulating mRNA expression – the average mRNA count \bar{m} and the rate r – do not change the results qualitatively. The mRNA expression rate r simply sets the maximal number of mRNA molecules at full expression in steady state (r/d); this influences the Poisson noise at the output, but does so equally for any regulatory architecture, proofreading or not. As long as r is large enough so that the average mRNA constraint \bar{m} is achievable, the precise choice of these values is not crucial (we use $r/d = 200$, $\bar{m} = 100$, plausible for eukaryotic expression). In sum, we can compare how well the optimal proofreading architecture does compared to optimal non-proofreading architecture in terms of information transmission, as a function of two key parameters: the crosstalk severity, Λ , and the input dynamic range, C_{\max} .

Figure 2.5A shows the advantage, in bits, of the optimal proofreading architecture relative to the optimal non-proofreading one. This “information plane,” $I_{q^*}(m; c) - I_{q=0}(m; c)$, is plotted

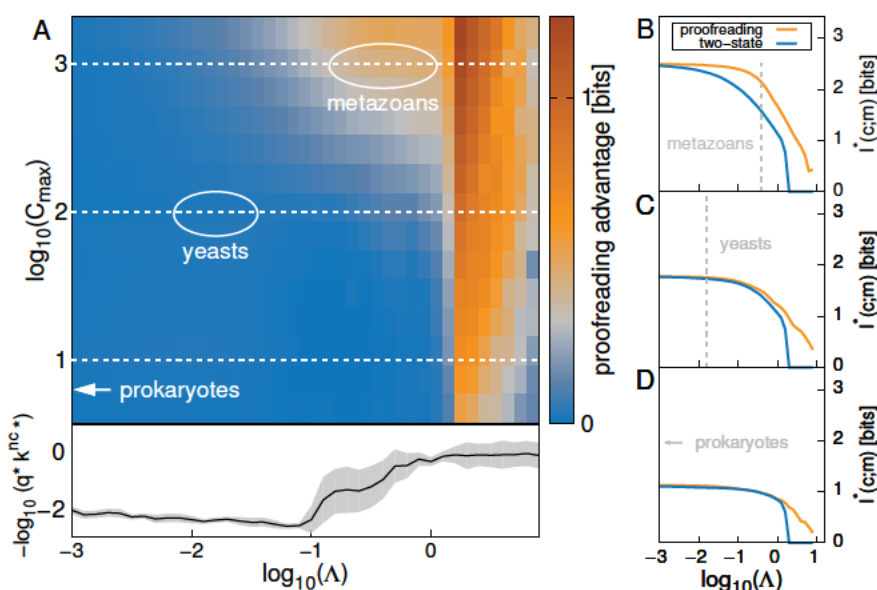


Figure 2.5: **(A)** Information advantage (in bits, color scale) of optimal proofreading over optimal two-state architectures, as a function of crosstalk severity Λ and dynamic range of input TF concentration, C_{\max} . Typical values for prokaryotes, yeast, and metazoans are marked in white. Lower inset: optimal rates, $q^* k_{-}^{\text{nc},*}$ (black line = average over C_{\max} , gray shade = std), indicate a switch to the proofreading strategy. **(B, C, D)** Cuts through the information plane in (A) along white dashed lines showing the collapse of two-state performance as $\log_{10}(\Lambda) \rightarrow 0$ and a clear proofreading advantage for metazoan regulation.

as a function of Λ and C_{\max} . In the limit $\Lambda \rightarrow 0$, the difference in performance goes to zero: there, optimization drives $q^*k_-^{\text{nc},*} \gg 1$, but proofreading offers vanishing advantage over the optimal two-state promoter architecture when noncognate binding is negligible. As Λ increases, proofreading becomes beneficial over the two-state architecture, and more so for higher values of C_{\max} . Higher input concentrations $c \in [0, C_{\max}]$ permit faster on-rates, resulting in faster optimal off rates $k_-^{c,*}$ and faster optimal $1/q^*$. Generally, faster switching of promoter states in Fig 2.3B means that promoter switching noise will be lower and thus information higher (at fixed mean mRNA expression \bar{m}); in particular, optimization tends to minimize promoter switching noise by selecting the fastest $1/q$ that still admits error rejection, i.e., $q^*k_-^{\text{nc},*} \sim 1$. At $\Lambda = \nu/\sigma \simeq 1$, the signaling capacity of the non-proofreading architecture collapses completely, with $I_{q=0}(c; m) \approx 0$ ². At this point optimal proofreading architectures are affected, but still generally maintain at least half of the capacity seen at $\Lambda = 0$; proofreading extends the performance of the gene regulation well into the $\Lambda > 0$ region, before finally succumbing to crosstalk.

Where do different organisms lie in the information plane? Prokaryotes have on the order of $\nu \sim 100$ types of transcription factors, whose binding site motifs typically contain around 23 bits of sequence information [Wunderlich and Mirny, 2009], corresponding to the binding energy difference of $16 k_B T$ between cognate and noncognate sites [Gerland *et al.*, 2002], and thus a specificity of roughly $\sigma \sim 10^7$. This corresponds to a small value of crosstalk severity, $\Lambda \sim 10^{-5}$. For yeast, the typical sequence information is 14 bits ($10 k_B T$) [Wunderlich and Mirny, 2009], which gives $\Lambda \sim 0.01$ (for $\nu \sim 200$ [Jothi *et al.*, 2009]). For multicellular eukaryotes, the typical sequence information is 12 bits ($8 k_B T$), and the number of TF species varies between $\nu \approx 10^3$ (*C. elegans*) to $\nu \approx 2 \cdot 10^3$ (human) [Milo *et al.*, 2010], putting Λ between 0.1 and 1. We can also estimate the dimensionless parameter $C_{\max} = k_+ C/d$. Assuming diffusion-limited binding of TFs to their binding sites, $k_+ C/d \approx 3DaN/R^3d$, where $D \sim 1 \mu\text{m}^2/\text{s}$ is the typical TF diffusion constant [Milo *et al.*, 2010], $a \sim 3 \text{ nm}$ is the binding site size, $R = 3 \mu\text{m}$ ($1 \mu\text{m}$) is the radius of an eukaryotic nucleus (prokaryotic cell), and N is the typical copy number of TFs per nucleus ($N \sim 10$ for prokaryotes, 10^3 for yeast, $10^3 - 10^5$ for eukaryotes). Typical mRNA lifetimes are 5 – 10 min in prokaryotes, 20 – 30 min in yeast, and > 1 hour in metazoans. This yields C_{\max} of order 10 for prokaryotes, 10^2 for yeast cells, and $> 10^3$ for multicellular eukaryote cells. While these are very rough estimates, different kinds of cells clearly differ substantially in their location on the information plane of Fig 2.5A.

²This is independent of whether one modulates Λ by changing ν , as for Fig 2.5A, or by changing σ ; although the optimal rates may take on different values, the information plane is essentially unchanged irrespective of how Λ is modulated.

Taken together, these values suggest that crosstalk is acute for metazoans and that proofreading in gene regulation could provide a vast improvement over equilibrium regulation schemes, as in Fig 2.5B. In contrast, our proposal offers no advantage for prokaryotes, and remains agnostic about yeast (Figs 2.5C, D). While much remains unknown about the molecular machinery of eukaryotic gene regulation, it has been experimentally shown that transcriptional initiation (not just elongation) involves a series of out-of-equilibrium steps. Amongst those, perhaps the most intriguing are the covalent modifications on the eukaryotic RNA polymerase II CTD tail [Egloff and Murphy, 2008]. The tail contains tandem repeats of short peptides (from 26 repeats in yeast to 52 in mammals), which need to get phosphorylated in order to initiate transcription and subsequently cleared after completed transcription in order to reuse the polymerase; genetic interference with this tail seems to be lethal. One can contemplate a scenario where a sequence of such phosphorylation steps corresponds to the out-of-equilibrium reaction q of our simple proofreading scheme, “ticking away” time until the polymerase commits to initiation, with every tick giving the machinery another opportunity to check if cognate TFs are still bound and, if not, abort transcription. The existence of any such (or similar) proofreading scheme would be interesting, but is currently purely hypothetical.

Why would eukaryotes employ a method of gene regulation so qualitatively different from prokaryotes, instead of simply using longer, specific binding sites that would drive crosstalk severity Λ towards zero? While beyond the scope of this work, one possible hypothesis is that such longer sites are not easily evolvable and, additionally, that the complexity of regulation calls for combinatorial control of single genes by many TFs of different species, each of which could have weak specificity. Such cooperative or combinatorial control could indeed address a specific target gene uniquely, as proposed (e.g., [Todeschini *et al.*, 2014; Wunderlich and Mirny, 2009]); what has largely been neglected in previous discussions is that it would be difficult to *prevent* the target gene from being erroneously induced by crosstalk. Here we advanced a possible hypothetical mechanism, proofreading-based transcriptional regulation, to mitigate this problem. It is interesting to note that, unlike most biophysical problems where we clearly appreciate their out-of-equilibrium nature, transcriptional regulation has remained a textbook example of a non-trivial *equilibrium* molecular recognition process, likely due to the success of the equilibrium assumption in prokaryotes. Perhaps constraints imposed by crosstalk will motivate us to re-examine this assumption in eukaryotic regulation more closely.

2.3.1 Robustness to non-optimal input distributions

Capacities for both proofreading and two-state models are computed by finding the optimal input distributions, $P^*(c)$, for each model. One could be concerned that the information advantage of the proofreading mechanism requires a very finely tuned input distribution, and that small deviations from such a distribution would erase the information advantage. In Fig 2.6 we show that this is not the case.

Specifically, we pick a proofreading model representative of metazoans (from the corresponding region in the phase diagram in Fig 2.5A), and the matched two-state model (with $q \rightarrow 0$). The optimal distribution for this model looks biologically plausible, with two peaks, at the high and at the low level of the TF concentration, c . We then perturb the optimal distribution of the proofreading model, holding other parameters fixed, and compute how this impacts the information transmission. The perturbations are done by first finding a parametric approximation to the optimal distribution by fitting a mixture-of-Gaussians (two mixture components). Within this parametric model class, we can now generate new (perturbed) distributions by sampling their parameters using a Monte Carlo scheme: the weight of the two mixture components is drawn randomly from a uniform distribution on $[0, 1]$, while the means and variances of the two mixture components are drawn from Gaussian distributions centered on the fitted values and of increasing variance. We ensure proper discretization, truncation to the original domain, and normalization; using rejection sampling we only retain distributions that obey the mean mRNA expression constraint at $\langle m \rangle = 100 \pm 5$, as in the previous section.

Figure 2.6 shows that we can easily generate distributions that are significantly different from the optimal one by visual inspection, and as quantified by the Kullback-Leibler divergence measure D_{KL} ; qualitatively speaking, as long as the approximate bimodal shape is maintained with roughly appropriate weights in the two broad peaks (quantitatively, $D_{KL} < 0.5$ bits), the distribution details do not matter much and the proofreading will outperform the best possible two-state model. For a similar result on the robustness of mutual information in a transcriptional channel see [Tkačik *et al.*, 2008a].

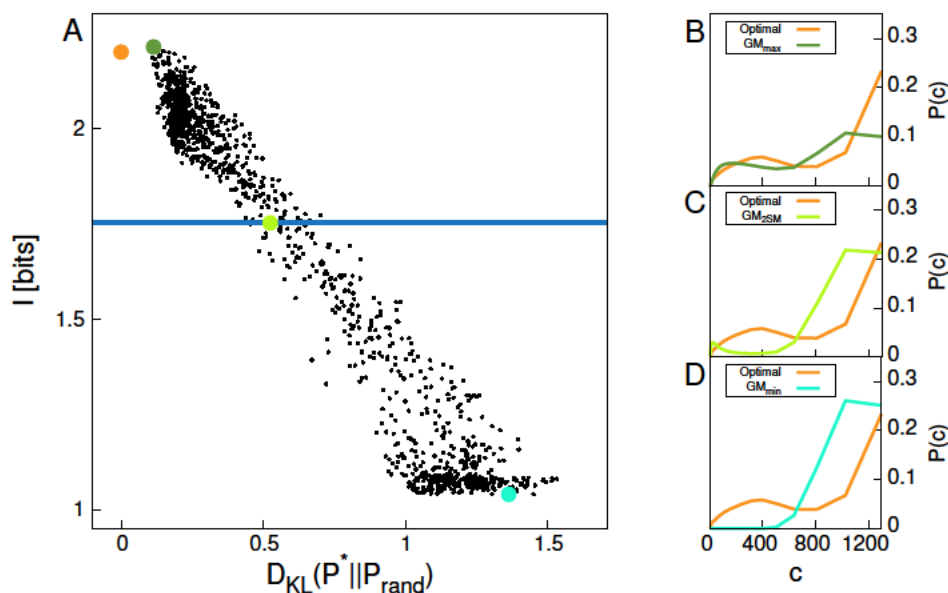


Figure 2.6: **Robustness of the proofreading model to non-optimal input distributions.** (A) When changing the optimal input distribution P^* , the mutual information I of the channel will be lower than the channel capacity. We have created trial input distributions by randomly perturbing Gaussian-mixture-distributions P_{rand} . The Kullback-Leibler divergence (D_{KL}) on the horizontal axis quantifies the difference between the optimal input distribution, $P^*(c)$, and various trial distributions. The blue line shows the channel capacity of the two-state-model. (B-D) Plots of the input distributions used to calculate (A) with the optimal distribution P^* as a reference in orange. Specifically, these are the Gaussian-mixture-distribution within the sample with (B) the highest mutual information, GM_{max} , (C) the mutual information closest to the channel capacity of the two-state-model, GM_{2SM} , (D) the lowest mutual information, GM_{min} . The colored dots in (A) correspond to the distributions of same color in (B-D). The mutual information for the input distribution GM_{max} is slightly higher than the channel capacity for the proofreading model, because of the small permitted deviation from the mean mRNA number constraint in our Monte Carlo rejection sampling (which is strictly enforced for the optimal solution).

2.3.2 Possible molecular implementations of proofreading in transcriptional regulation

We discuss two suggestions for a biologically plausible implementation of the abstract proofreading scheme studied in the previous section. The main difference between the two proposed mechanisms is which molecular complex carries the “mark” of the high-energy state (histone or polymerase). The first mechanism works also for “naked” DNA, whereas the second proposed mechanism involves epigenetic factors beyond the DNA and the basic transcription machinery.

Independent of the model under consideration, as described in [Hopfield *et al.*, 1976], a set of two experiments can determine if a given non-equilibrium scheme contributes to proofreading (i.e. to improved specificity). These experiments look at the energy consumption per product molecule under different circumstances. If the high-energy intermediate (the product of the “irreversible” driving step) always yields the correct product, the energy consuming step will be used exactly once per product molecule. If, however, a proofreading mechanism is in place, the energy consumed per product molecule will be higher. In the case where only the non-cognate “leg” is possible, the energy consuming-step will be used many more times per product molecule than in the case where only the cognate “leg” is possible. This is the key experimental signature of proofreading. For transcriptional proofreading, these experiments would consist of measuring the ATP consumption per transcript initiation event. Proofreading is indicated by an increased expenditure of ATP when initiation starts using noncognate TFs.

RNAPol-based mechanism

One possible mechanism involving modifications of the RNA polymerase (RNAPol) is schematized in Fig 2.7. A transcription factor (TF) binds and unbinds the DNA in an equilibrium process. When the TF is bound to DNA together with the RNAPol it can catalyze (or help catalyze) modifications on the RNAPol that are necessary for transcription; expenditure of energy is needed to perform such a modification. Even after RNAPol is modified, the TF can still dissociate, which will induce the dissociation of RNAPol as well and prevent transcriptional initiation. This constitutes the second specificity-conveying dissociation step of the abstract scheme proposed in the previous section. Transcription initiation can only commit from a complex of TF, DNA, and modified RNAPol. Importantly, for this scheme to work, modified RNAPol should not be allowed to exist in free form and bind directly to DNA (as this would constitute skipping the first specificity-conveying TF-DNA interaction step).

What could these modifications to the RNAPol be? An intriguing case are the modifications

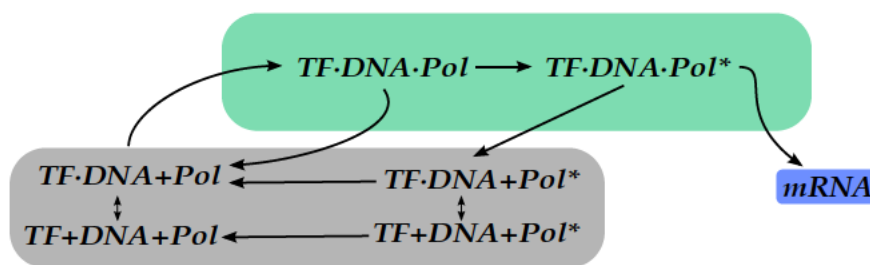


Figure 2.7: **A model of transcriptional proofreading based on RNAPol-modifications.** The grey box corresponds to state 0 in the abstract model in the main text. Only the correct initiation branch is shown. Importantly, the polymerase can only acquire the mark necessary for initiation ($Pol \rightarrow Pol^*$) with the TF present and can not rebind, once unbound from the DNA. Also, the mark necessary for initiation can only be removed while not bound to DNA. For the purpose of analysis in the previous section, we ignore details of how the first complex arises (the Michaelis step), and therefore treat the grey box as one effective state.

of RNAPol carboxy-terminal domain (CTD) in eukaryotes. This “tail” of the largest subunit of the RNAPol II consists of several repeats of the amino-acid sequence $Tyr^1 - Ser^2 - Pro^3 - Thr^4 - Ser^5 - Pro^6 - Ser^7$ (heptad repeats), and is unique to this polymerase. The phosphorylation state of these different repeats (heptads) is central in promoter escape (\sim transcription initiation) [Buratowski, 2009]. A phosphorylated tail (Ser^5P -repeats) is necessary to successfully initiate transcription of mRNA [Egloff *et al.*, 2012], whereas dephosphorylation of the CDT stimulates recruitment to the promoter, consistent with the requirement that RNAPol needs to be “reset” (i.e., have the marks cleared) before it can rebind to DNA [Prelich, 2002; Cho *et al.*, 1999].

The evolution of the CTD tail and its modifications [Stiller and Hall, 2002] seems to show a compelling association with the overall complexity of the organism, shown in Fig 2.8. Even though speculative, we note that a longer CTD tail provides the cell with the possibility to perform multiple reaction steps on the path towards initiation: this can either serve to fine-tune the rate $1/q$ in our model, or, by expending additional energy at every step and allowing TF dissociation after every step, could be used to boost the specificity beyond what is possible with a single proofreading step, a possibility mentioned already by Hopfield in his original paper. Interestingly, the evolutionary appearance of the CDT precedes all complex life and may therefore be regarded as a pre-requisite for eukaryotic regulation; all eukaryotes seem to trace back to a single “CDT-ancestor.”

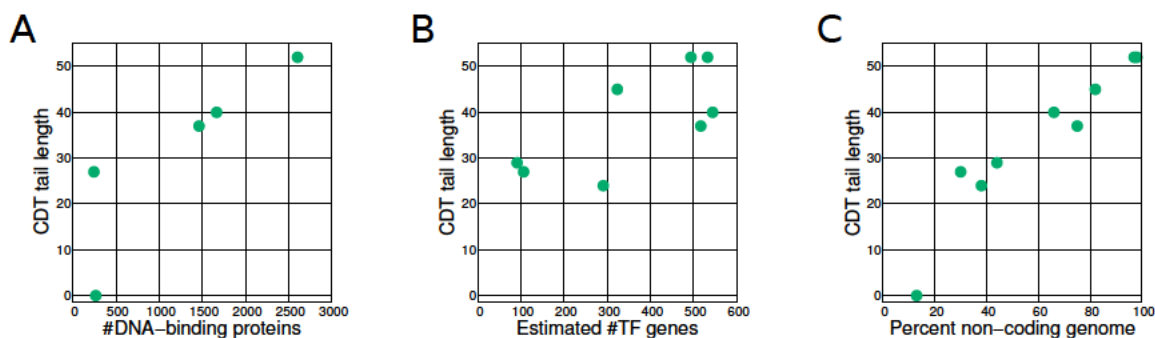


Figure 2.8: **Various measures of organismal complexity are associated with the length of the CTD tail on RNAPol.** (A) Number of DNA-binding proteins. (B) Estimated number of transcription factor genes. (C) Relative amount of non-coding DNA in the genome. The plots were created by combining data from [Prelich, 2002; Taft *et al.*, 2007; van Nimwegen, 2003; Babu *et al.*, 2004].

Histone-based mechanism

Histone-mediated models share an obvious similarity to the RNAPol-based models on the abstract level. The main difference, important for experimental validation, is that these mechanisms might be hard to test *in vitro* (on naked DNA); *in vivo*, histone manipulation and modification processes are complex and consume energy even when cognate initiation takes place, making traditional signatures of proofreading harder to detect. Also, it could be difficult *in vivo* to produce two necessary environments for a direct experimental test, one depleted in cognate TFs and rich in noncognate TFs, and one depleted in noncognate TFs and rich in cognate TFs.

The key idea is that a regulatory protein (for convenience still abbreviated TF) modifies the histone ($\text{HIS} \rightarrow \text{HIS}^*$) and this modification cannot happen without the TF being bound. Furthermore, the modification can only be removed when the TF is not present. Transcription initiation is only possible while a complex involving the TF is present *and* the histones have been modified to facilitate promoter escape, as shown in Fig 2.9. Beyond these conditions, the detailed dynamics of complex assembly are not important for the existence of proofreading.

Possible molecular candidates for the central modification of the proposed scheme concern the first nucleosome after the promoter (the +1 nucleosome). Certain +1 nucleosome histone variants have been shown to be more permissive to transcription than others, such as the H3.3 and the H2A.Z variant [Venkatesh and Workman, 2015; Weber *et al.*, 2014]. The proposed mechanism would involve the exchange of histones in the nucleosome at the +1 position. A TF could bind to an enhancer and, together with the mediator complex, form a complex that enables the necessary modifications to the +1 nucleosome. Successful initiation is subsequently

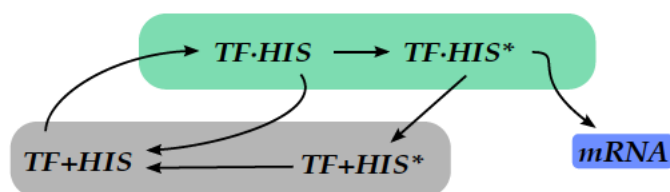


Figure 2.9: **A model of transcriptional proofreading based on histone-modifications.** Transcription is only possible while a complex involving the TF is still present and histones have been modified to facilitate promoter escape. Resetting the histone is only possible when the TF is no longer bound.

much more likely as long as the TF-mediator complex association is in place [Nock *et al.*, 2012; Liu *et al.*, 2011]. In this model it is crucial that the TF is present at its binding site (in the enhancer) during initiation.

Indirect experimental evidence in support of proofreading via such a mechanism could be gained by genetically interfering with various factors responsible for histone modifications and specifically testing whether the resulting phenotypes show increased rate of spontaneous activation (as a proxy for erroneous activation in our model) in the absence of cognate TFs. Such a test, for example, would be possible in the developmental context, where the maternal gradients of transcriptional activators provide a convenient way to probe the response of genetic regulatory elements across a natural range of cognate TF concentrations.

An intricate interaction of the CDT of RNAPol II and the (three-way) interactions with the mediator complex and the downstream nucleosome could make up a more complicated scheme than the two simplest scenarios outlined here. Understanding such an elaborate proofreading setup would seem to require knowledge of the kinetic steps involved that is beyond our current understanding.

3 Biophysical aspects of noisy gene expression evolution

This chapter reports on joint work with Tiago Paixão, Nicholas Barton, and Gašper Tkačik.

3.1 Introduction

In the two previous chapters I have described aspects of gene expression noise as a topic of biophysical interest by discussing different sources of noise and how those affect the signaling challenges a cell faces. In this chapter I turn to the question of how the unavoidable presence of noise in gene expression influences the evolution of gene expression.

Selection on gene expression is paramount [Fay and Wittkopp, 2007; Zheng *et al.*, 2011; Romero *et al.*, 2012; Hoekstra and Coyne, 2007]. In recent years, evidence has accumulated that also the width of the expression distribution (i.e. gene expression noise) is subject to natural selection. Cases for selection towards reduced [Lehner, 2008; Metzger *et al.*, 2015], but also elevated [Blake *et al.*, 2006] noise levels have been documented. Examples range from stress-, persistence- and infection related [Blake *et al.*, 2006; Arnoldini *et al.*, 2014], over metabolic [Wang and Zhang, 2011], to developmental genes [Raj *et al.*, 2010; Eldar *et al.*, 2009].

Variability in stress-related genes has often been discussed in the light of so-called bet-hedging strategies. These strategies help organisms to survive in a variable environment with limited or no sensing of the current conditions by preparing a part of the population for a range of possible challenges. It has received both theoretical and experimental attention in recent years [Kussell and Leibler, 2005; Acar *et al.*, 2008; Davidson and Surette, 2008; Beaumont *et al.*, 2009; Rotem *et al.*, 2010; Balaban, 2011; Koh and Dunlop, 2012; Müller *et al.*, 2013]. Here we will use microbial stress response pathways as an example to understand fitness effects of variability in gene expression. On the one hand, this is because these genes have

been characterized as having high noise [Newman *et al.*, 2006; Silander *et al.*, 2012]; on the other hand, with the ability to survive under stressful conditions, a measure of fitness is readily available. A concrete example for this will be discussed in the next chapter.

A very different context where the question of precision in gene regulation is important, is the case of embryonic development [Houchmandzadeh *et al.*, 2002; Gregor *et al.*, 2007; Raj *et al.*, 2010; Dubuis *et al.*, 2013]. Even though it is clear that there must be fitness effects of (im)proper cell-fate decision making, it is much harder to quantify these for several genetic variants than in the microbial case.

Here we will study the evolution of gene regulation while taking noise in gene expression into account explicitly. Research in evolution has identified several factors central to the understanding of how populations adapt to various challenges over time [Barton *et al.*, 2007]. Most important for the purposes of this chapter, these are the mutational and the adaptive landscapes, non-heritable phenotypic variability, and (finite) population size [Iwasa, 1988; Sella and Hirsh, 2005; Barton and Coe, 2009; Berg *et al.*, 2004]. For the interactions of the latter two some work has been done [Ito *et al.*, 2009; Sato *et al.*, 2003; Sakata *et al.*, 2009; Wang and Zhang, 2011; Wolf *et al.*, 2015] but an integrated model of all three components is still missing. Generally, the evolution of a population will be constrained by the combined action of neutral and selective forces. To explore these trade-offs, this chapter will deal with interactions of the genotype-phenotype map, finite populations size and the dynamics of adaptation. The work in this chapter also builds on recent advances in the biophysical understanding of how regulatory sequences give rise to a noisy genotype-phenotype map to derive quantities central to evolutionary dynamics from first principles [Rieckh and Tkačik, 2014; Sanchez *et al.*, 2011b; Jones *et al.*, 2014; Raser and O'Shea, 2004; Garcia and Phillips, 2011; Kinney *et al.*, 2010; Brewster *et al.*, 2012].

In this chapter, we will first address how to understand the fitness effects of noise in section 3.2. In section 3.3 we will then try to understand how the noise and mean phenotypes are encoded in the genome, thus providing variants on which selection can act. Finally, we will explain how noise influences the composition of an adapted population (sections 3.4 and 3.5) and also their trajectories during adaptation (section 3.6).

3.2 Noise changes the fitness landscape

3.2.1 Effective fitness

We will first show how noise changes the fitness landscape under which gene regulation evolves. We consider a gene of interest that is central for the survival or proper development of an organism (see Fig 3.1A). Depending on the expression level g (e.g. protein copy number) at a specific point in time, the organism has a certain probability $f(g)$ of surviving (or developing correctly). This probability is therefore directly related to the fitness of the genotype, which is why we will call $f(g)$ the molecular fitness function. Given a large number of clonal offspring, the genotype under consideration produces a range of different expression levels distributed according to $P(g; \phi)$, where ϕ are parameters of the expression distribution. Typically, these parameters will be the mean μ and the standard deviation σ , thus providing a way to model noise and incorporate the results from Chapter 1, e.g. the dependence of noise magnitude on the mean ($\sigma(\mu)$, noise characteristic).

The two consecutive probabilistic steps of gene expression and survival given a certain expression level can be combined into a joint measure of *effective fitness* \hat{f} :

$$\hat{f}(\phi) = \hat{f}(\mu, \sigma) = \int f(g) P(g; \mu, \sigma) dg \quad . \quad (3.1)$$

This quantity summarizes the interactions of expression distribution and molecular fitness function. Similar frameworks have been considered recently [Zhuravel *et al.*, 2010; Charlebois *et al.*, 2011; Charlebois, 2015; Wolf *et al.*, 2015]. From equation (3.1), we can see that the (molecular) fitness of the mean (i.e. $f(\mu)$) is not necessarily the effective fitness \hat{f} (also called mean fitness). Additionally, as illustrated in Fig 3.1B, a wider distribution neither generally increases, nor generally decreases effective fitness – rather, the effect of increased noise will depend also on the mean level and the relation to the fitness function. In this section, we will mainly work with the effective fitness of the mean $\hat{f}(\mu)$ and only later (sections 3.5 and 3.6) consider noise to also be an evolvable quantity, independent of the mean.

The assumption that only the instantaneous level of gene expression is relevant for fitness might be too strong for some systems. In general, effective fitness could depend in complicated ways on the whole trajectory of gene expression $g(t)$ through some stages of the life cycle of an organism, in which case it can be written as

$$\hat{f}(\phi) = \int f(g, t) P(g|t; \phi) dg dt \quad (3.2)$$

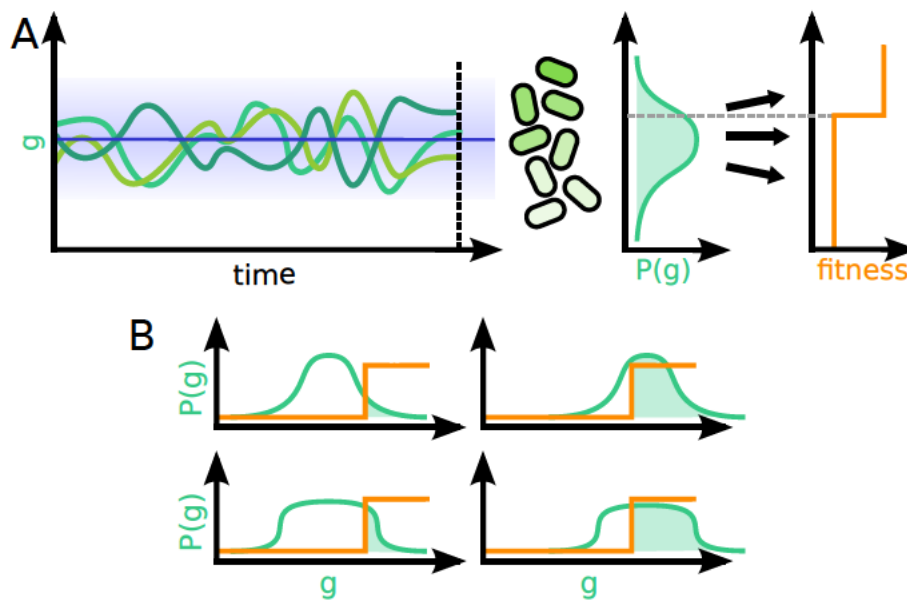


Figure 3.1: **A simple model to explicitly account for the influence of gene expression noise on fitness.** (A) The expression level of a fitness relevant gene g fluctuates over time around its mean (blue line) within certain boundaries (shaded blue region). A collection of clonal cells carrying this gene will generate a distribution of expression levels $P(g)$. At a certain point in time (black dashed line), fitness is evaluated according to the current expression level and the molecular fitness function (orange). Only a subpopulation survives this step, the relative size of which determines the effective fitness. (B) Mean and variability only jointly determine fitness. Shown are four examples with high/low mean and high/low variability. Effective fitness is indicated by the size of the green shaded area. In this example, increased noise is only beneficial if the center of the distribution is below the selection threshold, whereas for higher-mean distributions, a low variation leads to higher effective fitness.

where $f(g, t)$ is the fitness of a certain time trace and $P(g|t; \phi)$ is the probability of an expression trajectory given some parameters ϕ of the stochastic process. For general fitness functions $f(\cdot)$ and stochastic processes $P(\cdot)$, this problem can become hard to solve. A more modest generalization of equation 3.1 is the following: gene expression is a stationary stochastic process with some mean, variance, and timescale, and the dependence of fitness on the expression trajectory is simple time averaging. This case will be treated in section 3.6.3.

3.2.2 Gene expression noise is not generally equivalent to a change in population size

A well known stochastic effect in evolution is the sampling noise stemming from the finite number of individuals in each generation of an evolving population [Barton *et al.*, 2007]. It is therefore interesting to see if noise in gene expression can – in its mathematical treatment – simply be equated to a change in (effective) population size.

For this, we look at a comparatively simple situation: an ensemble of monomorphic, haploid populations that has reached its steady state after all the effects of the dynamics of adaptation have decayed [Sella and Hirsh, 2005; Barton and Coe, 2009]. In this setting, the probability of observing a monomorphic population with a certain genotype s , or a certain phenotype ϕ is given by

$$p(s) = \frac{1}{Z_s} \hat{f}(s)^N \quad , \quad (3.3)$$

$$p(\phi) = \frac{1}{Z_\phi} g(\phi) \cdot \hat{f}(\phi)^N \quad , \quad (3.4)$$

where $\hat{f}(s)$ and $\hat{f}(\phi)$ are the effective fitnesses of a genotype s or a phenotype ϕ , N is the effective population size, $g(\phi)$ is the number of genotypes for a given phenotype ϕ (the neutral distribution, or density of states) and Z_s and Z_ϕ are normalizing factors. In this chapter, the phenotype ϕ to be considered will be the parameters of the expression distribution (e.g. mean and variance of a Gaussian). For more on $g(\phi)$ see the next section.

From the form of the above equations one can see that the answer to the question whether noise and population size are equivalent, will depend on how exactly the width of the expression distribution enters into the expression for effective fitness $\hat{f}(\phi)$ and $\hat{f}^N(\phi)$, respectively.

If we assume a noise model where the width of the expression distribution is constant (and therefore independent of the mean), selection can under certain conditions be shown to be equivalent to the noiseless case with a different effective population size. This, however, does not hold generally when more realistic models of gene expression noise are considered.

To see this, we look at a Gaussian molecular fitness function with mean μ_f and variance σ_f^2 and a Gaussian expression distribution with mean μ and variance $\sigma(\mu)^2$, as the variance will generally depend on the mean (also see Chapter 1). For the effective fitness of a certain mean

μ we then get

$$\hat{f}(\mu) = \hat{f}(\mu, \sigma(\mu)) = \int \mathcal{G}(g; \mu_f, \sigma_f) \mathcal{G}(g; \mu, \sigma(\mu)) dg \quad (3.5)$$

$$\mu' = \frac{\sigma_f^{-2} \mu + \sigma(\mu)^{-2} \mu_f}{\sigma_f^{-2} + \sigma(\mu)^{-2}}, \quad \sigma'^2 = \frac{\sigma_f^2 \sigma(\mu)^2}{\sigma_f^2 + \sigma(\mu)^2} \quad (3.6)$$

$$\hat{f}(\mu) = \int \mathcal{G}(\mu; \mu_f, \sqrt{\sigma(\mu)^2 + \sigma_f^2}) \mathcal{G}(m; \mu', \sigma') dm = \quad (3.7)$$

$$= \mathcal{G}(\mu; \mu_f, \sqrt{\sigma(\mu)^2 + \sigma_f^2}) \cdot 1, \quad (3.8)$$

where \mathcal{G} is the Gaussian function.

Fixed noise: Simple equivalence between phenotypic noise and effective population size

We start with the fixed-noise case, i.e. $\sigma(\mu) = \sigma$, and get:

$$\begin{aligned} \hat{f}^N(\mu, \sigma) &\propto \frac{1}{(\sigma^2 + \sigma_f^2)^{N/2}} \cdot \exp \left[-N \frac{(\mu - \mu_f)^2}{\sigma^2 + \sigma_f^2} \right] \propto \\ &\propto \exp \left[-N \frac{(\mu - \mu_f)^2}{\sigma^2 + \sigma_f^2} \right]. \end{aligned} \quad (3.9)$$

From this we see that as far as the shape of the fitness function is concerned, we can exchange N and σ in the following way:

$$\frac{N_1}{N_2} = \frac{\sigma_1^2 + \sigma_f^2}{\sigma_2^2 + \sigma_f^2}, \quad (3.10)$$

and if we take $\sigma_1 = 0$ as a comparison to the noise-less case, we get

$$N_2 = \left(\frac{\sigma_2^2}{\sigma_f^2} + 1 \right) N_1. \quad (3.11)$$

Thus a change in population size can analytically be treated as a change in the selection width (σ_f) or in the width of the expression distribution (σ), as was reported previously [Wang and Zhang, 2011].

While this noise model ($\sigma(\mu) = \sigma$) may hold for abstract phenotypic noise, it is in contradiction with our biophysical understanding of gene expression.

Biophysical noise: Non-equivalence of expression noise and effective population size

To see the effect of a biophysically motivated noise model (and break-down of the equivalence) we look at $\sigma^2(\mu) = \nu\mu$. The effective fitness term now is

$$\begin{aligned} \hat{f}^N(\mu, \sigma(\mu)) &= \hat{f}^N(\mu, \nu) \propto \\ &\propto \frac{1}{(\nu\mu + \sigma_f^2)^{N/2}} \cdot \exp \left[-N \frac{(\mu - \mu_f)^2}{\nu\mu + \sigma_f^2} \right]. \end{aligned} \quad (3.12)$$

Figure 3.2 shows how effective fitness calculated with this Poisson-like noise model ($\sigma^2(\mu) = \nu\mu$) can not be replaced by the molecular fitness function (in this case a Gaussian) with changed parameters.

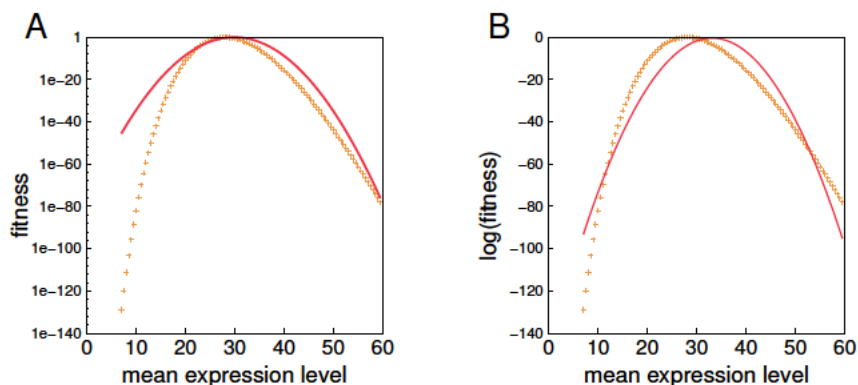


Figure 3.2: Non-equivalence of noise and finite population size: We can try to approximate the effective fitness term $\hat{f}^N(\mu, \sigma(\mu))$ calculated with a biophysical noise model $\sigma^2(\mu) = \nu\mu$ (orange dots) by a reparametrized version of the molecular fitness function (red lines). In the left panel, the best mean square fit is performed in the space of probabilities, while in the right panel it is performed in the space of log-probabilities. The details of these approximations will depend on sampling density and sampling range, but we can see that this approximation can not satisfy position and curvature of the effective fitness term over the whole range of phenotypes simultaneously. Parameters for the plots are: $N = 100$, $\nu = 8$, $\sigma_f = 10$, $\mu_f = 30$.

3.2.3 Noise smoothens the effective fitness function and can shift the maximally fit phenotype

In addition to the Gaussian cases above, it is instructive to also consider other types of molecular fitness functions and families for expression distributions.

Generally, a wider distribution in equation 3.1 will smoothen the molecular fitness function. This will affect, depending on noise characteristics, the position and slope around the maximally fit phenotypes (here: mean expression values).

Figure 3.3 shows the two prominent effects that the Poisson-like noise characteristic can have: in the case of a step-like molecular fitness function (truncating selection), gene expression noise smoothens the step. As noise increases, the selective advantage for higher mean levels starts already at low mean levels and the selection for a ‘margin of safety’ between the expected expression value and the selection cut-off becomes more and more stringent. For the

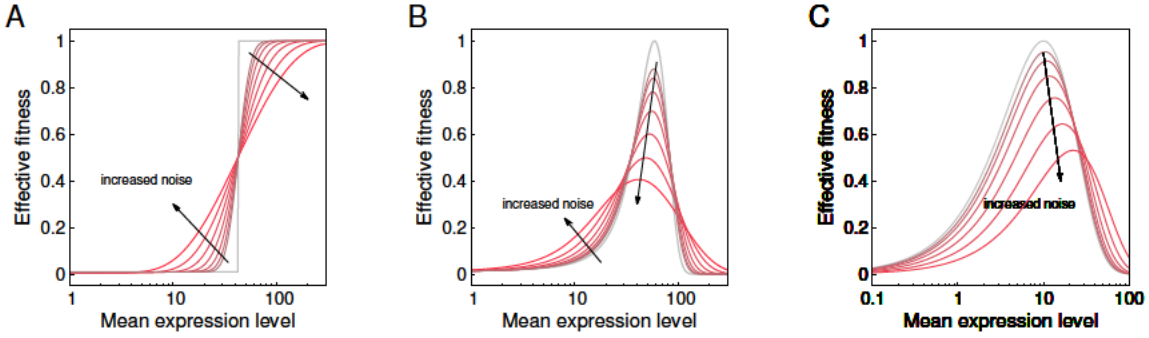


Figure 3.3: Effects of increased Poisson-like noise on effective fitness. **(A)** Effective fitness is generally smoother than the molecular fitness function it is derived from. For truncating selection, this leads to optimal phenotypes being pushed to higher mean expression levels, away from the selection threshold. For low means, however, increased noise leads to higher effective fitness. The molecular fitness function is a step function located at $g = 43$. **(B)** For stabilizing selection, increased noise renders selection for the optimal phenotype less stringent. Additionally, in the Gaussian case, the fittest phenotype is located at lower expression levels than for the noise-less case. Notice that the left flank is qualitatively similar to the case in **(A)**. The molecular fitness function is a Gaussian with mean of 60 and standard deviation of 20. **(C)** For Gamma distributed gene expression and molecular fitness ($\alpha_f = 2$, $\beta_f = 0.1$), we can see a similar widening effect on the selection window as in **(B)**, but the attainable fitness maximum is shifted to higher mean levels. For all panels, the molecular fitness function is depicted in gray and effective fitness calculated with $\sigma^2 = \nu\mu$ and $\nu = 1, 2, 4, 8, 16$, and 32 are shown. Black arrows indicate the effect of increased noise.

case of stabilizing selection (here: a Gaussian molecular fitness function), the window of selected phenotypes gets wider and its maximum shifts to lower mean expression levels. Finally, the case of Gamma distributions for the molecular fitness function and expression distribution with the Poisson-like noise characteristic shows that increased noise can also lead to the maximum of effective fitness shifting to higher means as noise increases.

Gamma distributions and Poisson-like noise characteristics

For a Gamma distribution, we have $y(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp[-\beta x]$, with its mode at $(\alpha - 1)/\beta$, the mean α/β and the variance α/β^2 . Thus we take as the molecular fitness function

$$f(g) = g^{\alpha_f-1} \exp[-\beta_f g] \quad , \quad (3.13)$$

and for the expression distribution

$$P(g) = \frac{\beta^\alpha}{\Gamma(\alpha)} g^{\alpha-1} \exp[-\beta g] \quad , \quad (3.14)$$

$$\mu = \alpha\beta^{-1}, \quad \sigma^2 = \alpha\beta^{-2} \quad , \quad (3.15)$$

and the biophysical (Poisson-like) noise model

$$\sigma^2 = \mu\nu = \alpha\beta^{-2} = \mu\beta^{-1} \quad , \quad (3.16)$$

$$\beta = \nu^{-1}, \quad \alpha = \mu\nu^{-1} \quad . \quad (3.17)$$

To get the effective fitness, we calculate

$$\hat{f}(\alpha, \beta) = \int f(m) \frac{\beta^\alpha}{\Gamma(\alpha)} m^{\alpha-1} \exp[-\beta m] dm = \quad (3.18)$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \int m^{\alpha_f-1} m^{\alpha-1} \exp[-\beta m] \exp[-\beta_f m] dm = \quad (3.19)$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \int m^{(\alpha+\alpha_f-1)-1} \exp[-(\beta + \beta_f)m] dm = \quad (3.20)$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \alpha_f - 1)}{(\beta + \beta_f)^{\alpha+\alpha_f-1}} \quad . \quad (3.21)$$

Using the properties of Gamma function, for integer α_f we get:

$$\hat{f}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \alpha_f - 1)}{(\beta + \beta_f)^{\alpha+\alpha_f-1}} = \quad (3.22)$$

$$= \frac{(\alpha + \alpha_f - 2)\Gamma(\alpha + \alpha_f - 2)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta + \beta_f)^{\alpha+\alpha_f-1}} = \quad (3.23)$$

$$= \frac{(\alpha + \alpha_f - 2)(\alpha + \alpha_f - 3) \cdots (\alpha + \alpha_f - \alpha_f)\Gamma(\alpha)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta + \beta_f)^{\alpha+\alpha_f-1}} = \quad (3.24)$$

$$= \alpha^{\overline{\alpha_f-1}} \cdot \frac{\beta^\alpha}{(\beta + \beta_f)^{\alpha+\alpha_f-1}} \quad , \quad (3.25)$$

with $x^{\overline{n}} = \underbrace{x(x+1) \cdots (x+n-1)}_{n \text{ factors}}$. This further simplifies to

$$\hat{f}(\alpha, \beta) = \alpha^{\overline{\alpha_f-1}} \cdot \beta^{-(\alpha_f-1)} \cdot (1 + \beta_f/\beta)^{-\alpha-(\alpha_f-1)} \quad . \quad (3.26)$$

Inserting the biophysical noise model, we get:

$$\hat{f}(\alpha, \beta) = (\mu/\nu)^{\overline{\alpha_f-1}} \cdot \nu^{\alpha_f-1} \cdot [(1 + \nu\beta_f)]^{-\mu/\nu-(\alpha_f-1)} \quad (3.27)$$

$$= \underbrace{\mu(\mu + \nu)(\mu + 2\nu) \cdots (\mu + (\alpha_f - 2)\nu)}_{\alpha_f-1 \text{ factors}} \cdot [(1 + \nu\beta_f)]^{-\mu/\nu-(\alpha_f-1)} \quad . \quad (3.28)$$

For $\alpha_f = 2$, this simplifies to:

$$\hat{f}(\alpha, \beta) = \frac{\alpha\beta^\alpha}{(\beta + \beta_f)^{\alpha+1}} \quad , \quad (3.29)$$

$$\hat{f}(\mu, \nu) = \mu(1 + \nu\beta_f)^{-(1+\mu/\nu)} \quad . \quad (3.30)$$

This, in turn, can be shown to attain its maximum at

$$\mu^*(\nu) = \frac{\nu}{\log[1 + \beta_f \nu]} = \frac{1}{\beta_f} + \frac{\nu}{2} - \frac{\beta_f}{12} \nu^2 + \frac{\beta_f^2}{24} \nu^3 + \dots, \quad (3.31)$$

where we note that $1/\beta_f$ is the mode of the molecular fitness function (cf. $(\alpha_f - 1)/\beta_f$ with $\alpha_f = 2$). This shows the shift in the position of the fittest mean to higher levels as noise increases.

3.3 Genotype-phenotype maps based on thermodynamic models of gene expression

To understand the evolution of a population, in addition to the fitness landscape discussed in the previous section, one also needs to understand the corresponding genotypic landscape constraining evolutionary trajectories. Thermodynamic models of gene expression are widely used to describe how properties of regulatory sequences map into gene expression [Bintu *et al.*, 2005; Shea and Ackers, 1985b; Berg and von Hippel, 1987; Kuhlman *et al.*, 2007a]. This class of models aims at describing gene regulation by calculating the equilibrium probability of a promoter to be in a certain state, which is then associated with an expression activity. For the description of the mean expression level, these models are therefore equivalent to the ones discussed in Chapter 1. They can be combined with models of how the sequence gives rise to the protein-DNA binding energy, which will then determine the occupancy of a site [Stormo and Fields, 1998; Stormo and Zhao, 2010]. One of the simplest models of how the binding energy arises from the sequences uses an energy matrix \mathcal{E} , the entries of which represent the independent contributions from every nucleotide at every position. The binding energy E_b is then computed as the sum of the individual contributions:

$$E_b = \sum_{i=1}^L \mathcal{E}_{i,j}, \quad (3.32)$$

where $\mathcal{E}_{i,j}$ is the contribution of nucleotide j (with $j \in \{A, C, G, T\}$) at position i in the binding site and L is the length of the binding site. These energy contributions have been measured in vitro and in vivo, and typically range from 1 to 3 in units of $k_B T$ [Maerkl and Quake, 2007; Kinney *et al.*, 2010; Brewster *et al.*, 2012; Gerland *et al.*, 2002]. The typical length of the binding site is 5 to 10 nucleotides for eukaryotes and 10 to 20 in prokaryotes [Wunderlich and Mirny, 2009; Ptashne and Gann, 2002]. These models predict the strength of binding well for a few mutations away from a strongest binding sequence [Brewster *et al.*, 2012]. A simplification of the energy matrix model is the so-called mismatch model, which assumes the existence of

a preferred or consensus binding sequence [von Hippel and Berg, 1986; Gerland and Hwa, 2002]. The number of nucleotides in the sequence of a binding site that are not the same as in the consensus sequence indicate the level of mismatch, and, when multiplied by an energy penalty per nucleotide, give the respective binding energy.

In this section, we will use these models to illustrate important properties of biophysically motivated genotype-phenotype maps for gene expression. As we have seen already in equation 3.4, the number of genotypes $g(\phi)$ displaying the same phenotype ϕ is an important quantity for evolutionary questions. In statistical physics it is called the density of states (DoS), whereas in population genetics it is called the neutral distribution (of phenotypes).

3.3.1 Simple form of neutral distribution of mean expression phenotypes

We will now show that the neutral distribution (DoS) for the mean expression level in a thermodynamic model, where the binding energies are derived from a mismatch model, takes a simple form. Furthermore this distribution can be well-approximated by a power-law.

We use a mismatch model to calculate binding energies from sequences, with k being the number of mismatches and E_x the energy contribution from a single mismatch:

$$E_b = E_x \cdot k \quad . \quad (3.33)$$

Consequently, $E_b = 0$ is the maximal binding strength and k takes values in $[0, L]$, where L is the length of the binding site.

To calculate the DoS of binding energies, we can view the sequence as a binomial process with L iterations and a probability of success (i.e. finding a mismatch) of $3/4$. We can then use the mean and variance of this process to approximate the DoS with a Gaussian:

$$p(E_b) = A \exp[-d_w(E_b - E_0)^2] \quad , \quad (3.34)$$

$$E_0 = \frac{3}{4} E_x L \quad , \quad (3.35)$$

$$1/d_w = 2 \cdot E_x^2 \frac{1}{4} \frac{3}{4} \cdot L = \frac{3}{8} E_x^2 L \quad . \quad (3.36)$$

The simplest energy-to-expression model (which has been tested experimentally [Brewster *et al.*, 2012]) is of the form

$$\mu = \exp[-E_b] \quad , \quad (3.37)$$

with μ signifying the (normalized) mean expression level. Furthermore, $E_b = -\log[\mu]$ and $\partial_\mu E_b = -\partial_\mu \log[\mu] = -1/\mu$.

With $\mu = f_\mu(E_b)$, or $E_b = f_\mu^{-1}(\mu)$, and $p_\mu(\mu) = p_{E_b} [f_\mu^{-1}(\mu)] \cdot |\partial_\mu f_\mu^{-1}(\mu)|$ we therefore get for the DoS of the means:

$$p_\mu(\mu) = A \exp \left[-d_w \underbrace{(-\log [\mu] - E_0)}_{=E_b} \right] \cdot \mu^{-1} \quad , \quad (3.38)$$

$$p_\mu(\mu) = A \exp [-d_w (\log [\mu] + E_0)^2] \cdot \mu^{-1} \quad . \quad (3.39)$$

We now need to argue why this distribution of mean expression levels can be developed as $p_\mu \sim \mu^{-\alpha}$. If we consider

$$\log [p(\mu)] = \log [A] - d_w (\log [\mu] + E_0)^2 - \log [\mu] \quad (3.40)$$

$$= \log [A] - d_w (\log^2 [\mu] + 2 \log [\mu] E_0 + E_0^2) - \log [\mu] \quad (3.41)$$

$$= a_E - (1 + 2E_0 d_w) \log [\mu] - d_w \log^2 [\mu] \quad (3.42)$$

with $E_0 = \frac{3}{4} E_x L$ and $d_w = \frac{8}{3} E_x^{-2} L^{-1}$ we get $2E_0 d_w = 2 \frac{3}{4} E_x L \frac{8}{3} E_x^{-2} L^{-1} = 4E_x^{-1}$. We can see that the linear term $(1 + 2E_0 d_w)$ is independent of L and larger than d_w , since E_0 and d_w are positive. Therefore, $\log [p(\mu)]$ is dominated by $\log [\mu]$ and $p(\mu)$ can be approximated by a power law. Note that the quadratic term will diminish as L gets longer and the exponent of the power law only depends on E_x .

Figure 3.4A shows that a similar argument using measured position weight matrices [Kinney *et al.*, 2010; Brewster *et al.*, 2012] and the central limit theorem to get the Gaussian distribution of binding energies can be made to approximate the DoS of mean expression levels with a power law (see Figure 3.4B). Figure 3.4C uses the more general energy-to-expression model

$$\mu = \mu_0 \frac{1}{1 + \exp [E_b - E_c]} \quad , \quad (3.43)$$

which also lends itself to being approximated by a power law. The 8 most informative positions in the measured position weight matrix were used to calculate the DoSs. The free parameters were chosen such that the average binding energy of the *E. coli* genome is zero, the wild-type *lac* promoter has an energy of $-5.35 k_B T$ [Brewster *et al.*, 2012], and corresponds to an expression level of 20 mRNA.

3.3.2 Joint densities of states for mean and noise in thermodynamic models

For a given network architecture, for example an upstream TF that can bind to a downstream promoter and influence its expression activity, we can ask how many sequences encoding the network give rise to a certain mean and noise level of the downstream gene. This will extend the DoS studied in the previous section to a joint density of states of mean and noise levels.

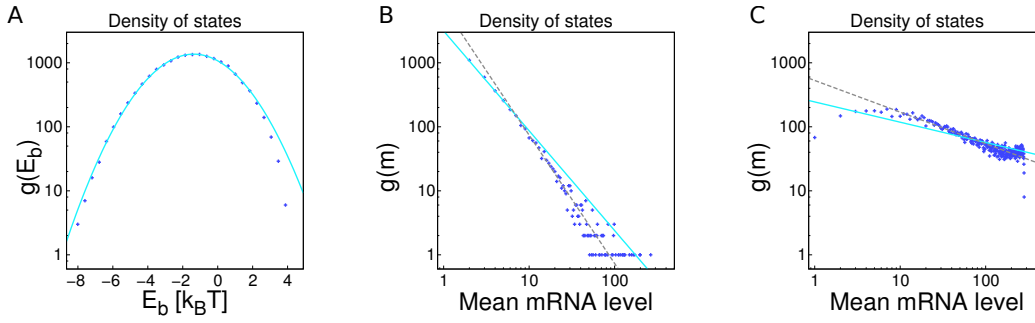


Figure 3.4: **(A)** The densities of states of binding energies and **(B)** in the space of mean expression levels for the model $\mu \propto \exp[-E_b]$ (dark blue dots). The DoS in energy space can be approximated by a Gaussian and the DoS of mean expression levels follows a power-law distribution (light blue lines). Also shown is a power-law fit for expression levels above 10 (gray dashed line). **(C)** DoS for the mean expression levels calculated with the expression function $\mu = \mu_s / (1 + \exp[E_b - E_c])$, $E_c = -2.0$ (dark blue dots), also showing a power-law approximation between minimal and maximal expression level (light blue line). The tail of the distribution above expression level 10 offers a better fit to the power-law (gray dashed line).

Here, we use the promoter models of the type studied in Chapter 1 to understand the joint DoS for mean and noise of the expression distribution for two simple regulatory architectures.

To relate switching rates to binding energies, we can look at thermodynamic models such as in equation 3.43 and develop the mathematical equivalence (cf. Figure 1.3 for the mean of the two state model):

$$\mu = \frac{\exp[E_c]}{\exp[E_c] + \exp[E_b]} = \frac{k_{ON}}{k_{ON} + k_{OFF}} \quad , \quad (3.44)$$

which, together with the argument about exponential transformations (cf. equation 3.37 and the Arrhenius equation [Bialek, 2013]) justifies the use of power laws also for the DoS of switching rates.

We can now use the promoter models from Chapter 1 (cf. section 1.9 and figure 1.4) to compute the joint DoS for mean and noise values, by combining them with the DoS of the switching and production rates. As a simple example, we pick the two-state model and use power-law DoS for the rates. In these models, there are two kinds of rates: one is determined by the concentration of a TF and the other is calculated from the affinity of a TF molecule to a sequence. The difference between activator and repressor schemes is in the sign of the exponent for ON- and OFF-rates. For example, in the activator scheme, the ON-rate is set by the concentration of an activating TF – for which we can assume that sequences encoding high concentrations are rare. Thus this exponent is negative. For the OFF-rate in the same scheme,

however, the exponent is positive. This is because fast rates here corresponds to weak binding of the TF and many sequences give rise to weak TF binding.

Figure 3.5 shows the joint densities of states for mean and noise for a simple activator and a simple repressor scheme. Generally, we clearly see the Poisson limit in both cases, prohibiting solutions for a high mean and a low noise at the same time. Above the Poisson limit, one can see a concentration of sequences for low super-Poissonian noise for a given mean. Also, this pair of models for activator and repressor suggests that for a given mean it should be easier to find a high-noise sequence in the activator than in the repressor setting. For simplicity, the used models do not account for the potential overlap of TF binding sites and the RNA polymerase. These interactions can, however, add some interesting features to the genotypic landscape [Paixao and Bauer, 2015]. In section 3.5 we will see that the details of the joint map can be important for the outcome of evolution.

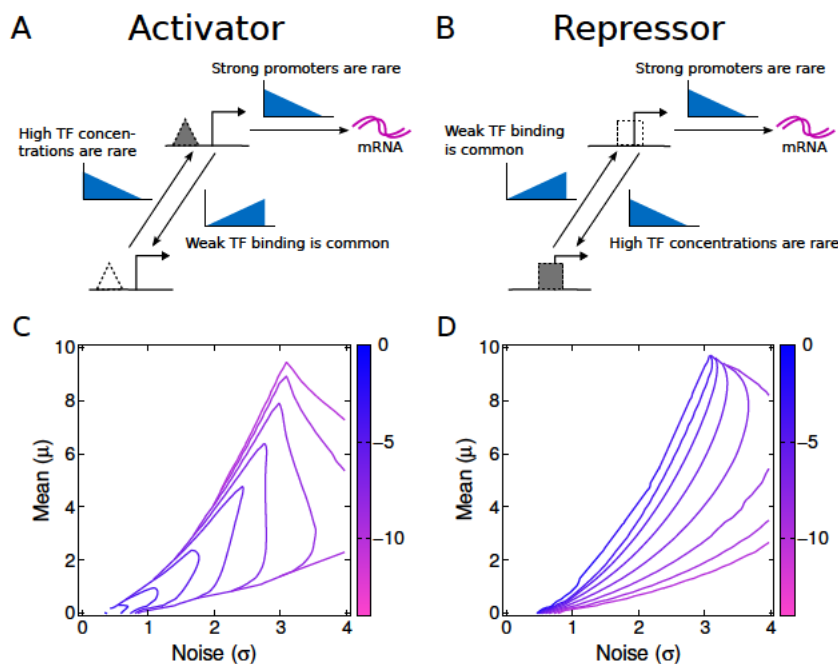


Figure 3.5: **(A,B)** Two-state models of gene regulation where the promoter switches between an inactive state (lower state, OFF-state) and an active state (upper state, ON-state), in which mRNA is produced. Switching and production rates are derived from regulatory sequences, resulting in the densities of states (blue triangles) depicted next to all rates (black arrows). The rationale for both schemes is that in genotype space, fast TF unbinding-rates are common (weak binding of the TF) and high TF concentrations are rare (high expression of the TF). Similarly, high production rates from the active state are not supported by many sequences. **(C,D)** The gene expression model together with the densities of states gives rise to a joint density for mean expression levels and variation around it. For a fixed mean, there are generally more sequences with low noise levels in these models, as long as the noise is above the Poisson limit. For all densities of switching rates, the power-law distribution with exponents -2 or $+2$ (as indicated in (A,B)) with rates from the interval $[0.1, 10]$ were used. The contour lines represent \log_{10} -densities normalized to the maximum density.

3.4 Combined effects of expression noise, density of states, and finite population size on genetic diversity

One might expect that since noise in gene expression can generate phenotypic diversity, it would also lead to higher genotypic diversity in a population. In this section, we therefore ask how phenotypic diversity caused by noise in gene expression affects genotypic diversity. As a measure for the latter we choose the expected number of genotypes in an ensemble of monomorphic population at steady state. This can be defined in a straight-forward manner from the description of a monomorphic population in steady state used in section 3.2.2. Genotypic diversity \mathcal{N} is then defined using the distribution of phenotypes and multiplicity of each phenotype (DoS) as

$$\mathcal{N} = \int g(\phi)p(\phi) d\phi \quad . \quad (3.45)$$

So for the simple case of the evolving mean expression level ($\phi = \mu$), e.g. calculated with some binding energies and a thermodynamic model of gene expression, and by using equation 3.4 we obtain:

$$\mathcal{N} = \int g(\mu)p(\mu) d\mu = \frac{\int g^2(\mu)\hat{f}(\mu)^N d\mu}{\int g(\mu)\hat{f}(\mu)^N d\mu} \quad . \quad (3.46)$$

From the partial equivalence of gene expression noise to a change in effective population size (cf. section 3.2.2), one might expect that an increase in noise will always lead to less stringent selection, thus leaving more genotypes compatible with the requirements of selection. However, Fig. 3.6 shows that this is generally not the case.

How can this be understood mathematically? For large population sizes N , we can evaluate the integrals using the Laplace method. For this, we observe that for two functions $k(x)$ and $h(x)$ the following holds:

$$\int k(x) \exp [N \cdot h(x)] dx \approx k(x^*) \exp [N \cdot h(x^*)] \sqrt{\frac{2\pi}{N|h''(x^*)|}} \quad , \quad (3.47)$$

with $h'(x^*) = 0$, where x^* is the position of the maximum of $h(x)$.

With the Laplace method with $h(x) = \log f(\mu)$ and $k(x) = g(\mu)$, we can now approximate integrals of the type occurring in the definition of \mathcal{N} resulting in

$$\mathcal{N} \propto g(\mu^*)\hat{f}^N(\mu^*) \sqrt{\frac{\hat{f}(\mu^*)}{N|\hat{f}''(\mu^*)|}} \quad , \quad (3.48)$$

where μ^* is the mean expression value that maximizes fitness. This expression emphasizes how the DoS together with the effective fitness determines the effect of increased noise. To

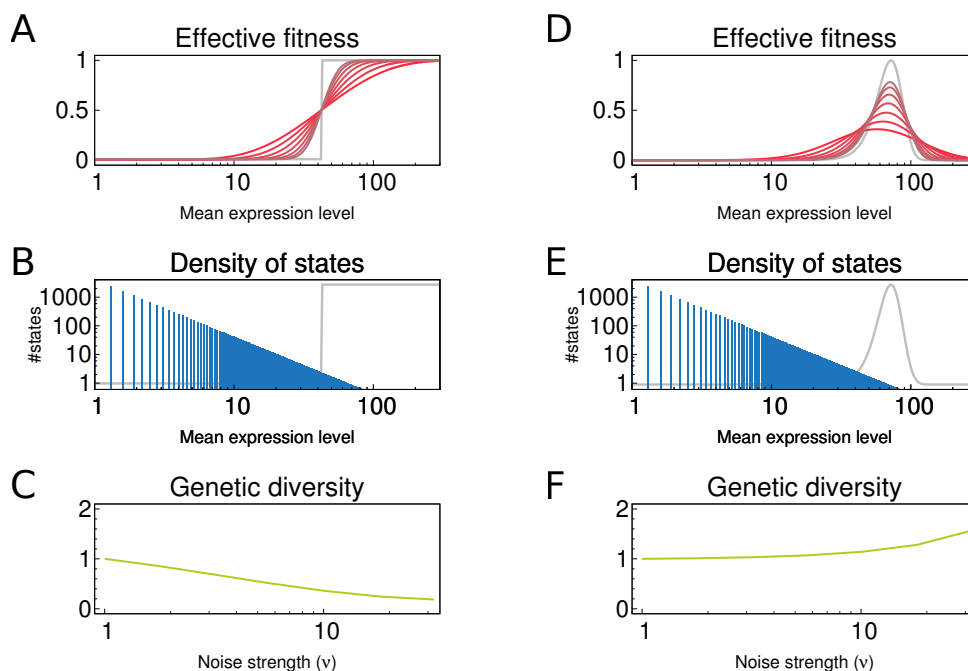


Figure 3.6: **Noise can influence genetic diversity in different ways.** Two cases, based on the molecular fitness functions in Figure 3.3, where increased gene expression noise has opposing effects on genetic diversity. In case of the truncating selection (left column, **(A-C)**), increasing noise will shift the optimal expression mean to a region of lower sequence abundance, as seen in (B). For stabilizing selection (right column, **(D-F)**), on the other hand, the same increase in noise leads to a shift of the maximum of the effective fitness to lower mean expression levels. For these low-mean phenotypes, a larger number of genotypes are available, as can be seen in (E). Genetic diversity in (C,F) are normalized to diversity at $\nu = 1$. Shown are exact numerical results for genetic diversity \mathcal{N} , and the power-law shaped DoS in (B,E) are motivated by thermodynamic models of gene expression, as discussed in section 3.3.1.

understand the effect of increased noise, the most important features to track are the position of the maximal fitness (μ^*) and the curvature of effective fitness around this maximum ($\hat{f}''(\mu^*)$). From equation 3.48 we can also see that since a change in effective population size can never systematically account for a shifted position of the maximal fitness, the equivalence of population size can only hold in special cases (cf. section 3.2.2). However, the curvature of the effective fitness function around its maximum can be traded against a change in population size, as can be seen in the denominator of the expression under the square-root, $N|\hat{f}''(\mu^*)|$, which accounts for the intuition that increased noise leads to greater diversity.

3.5 Effects of a joint genotype-phenotype map for noise and mean expression

We have so far looked at noise models where the mean expression level alone determined the noise – but it could not evolve independently. As seen in section 3.3.2, already simple regulatory networks are capable of encoding mean and noise in a (to some degree) independent fashion, i.e. there are different noise levels available for the same mean (and vice versa). To therefore show how the structure of the joint genotype-phenotype map matters for which values of mean and noise evolving populations will reach, we here look at the effects of correlations in the DoS between mean and noise.

As before, we summarize the effect of noise on fitness by the effective fitness. In this section, however, we treat noise and mean expression level as (a priori) independent traits, leading to a fitness landscape on the mean-noise plane (see Figure 3.7A). No noise model needs to be assumed, other than the Gaussian form of the expression distribution and the DoS. To be able to vary the relation between mean and noise in genotype space with a single parameter we model the DoS as a bivariate Gaussian (see Figure 3.7B):

$$G(\mu, \sigma) \propto \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{S}^{-1}(\mathbf{x} - \mathbf{m})\right), \quad (3.49)$$

$$\text{with } \mathbf{m} = \begin{pmatrix} m_\mu \\ m_\sigma \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} s_\mu^2 & \rho s_\mu s_\sigma \\ \rho s_\mu s_\sigma & s_\sigma^2 \end{pmatrix}. \quad (3.50)$$

where ρ is the correlation coefficient between mean and noise.

The distribution of phenotypes in steady state is then calculated in the monomorphic approximation (see section 3.2.2), where the probability of finding a population with a given phenotype ϕ solely depends on its effective fitness $\hat{f}(\phi)$, the population size N , and the DoS $g(\phi)$ (cf. equation 3.4):

$$p(\phi) = \frac{1}{Z} g(\phi) \cdot \hat{f}(\phi)^N, \quad (3.51)$$

$$p(\mu, \sigma) = \frac{1}{Z} g(\mu, \sigma) \cdot \hat{f}(\mu, \sigma)^N. \quad (3.52)$$

Figure 3.7C shows how these factors together give rise to a distribution of phenotypes. In Figure 3.7D, we see that the correlation in the DoS between mean and noise, can have pronounced effects on the level of mean and noise that a population displays. For high anti-correlation the DoS offers genotypes with high mean and low noise, which represents a combination with high fitness in this context. As the correlation coefficient increases, this kind of combinations get very rare and therefore the population settles on lower mean and higher noise genotypes.

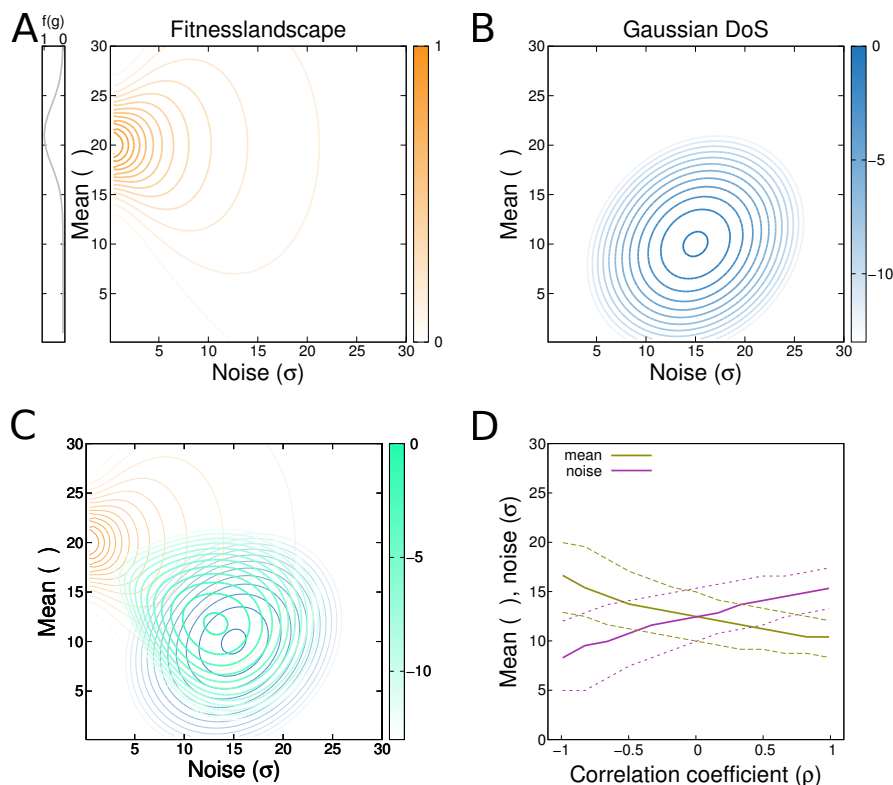


Figure 3.7: **(A)** When studying a freely evolving noise parameter in addition to the mean expression level one has to look at the fitness landscapes of combinations of these parameter values. Shown in gray on the left is the molecular fitness function which corresponds to the noiseless case. In the main plot, we can see that noise and mean expression combine in a non-trivial way to give the effective fitness for any combination of the two parameters. The plot shows normalized fitness, such that the highest fitness value is 1. **(B)** For similar reasons as in (A), we also need to look at the joint DoS of mean and noise parameters. In this example they are a bivariate Gaussian with significant correlation. The plot shows normalized log-density with a correlation coefficient of $\rho = 0.2$. **(C)** At a finite population size, the fitness landscape from (A) and (B) will combine into a distribution of observed phenotypes after evolution has reached a steady state (green). Shown is the log-probability normalized to its maximum. **(D)** By changing the correlation coefficient ρ between noise and mean in the DoS, while keeping the position of the maximal density the same, we can see how the genotypic correlation influences the level of noise observed in a population. Shown are the most probable mean and noise values (full lines) with the dotted lines indicating full width at half maximum. In all plots we use $m_\mu = 10, m_\sigma = 15, s_\mu^2 = s_\sigma = 5, \rho \in [-0.99, 0.99], N = 100$ and molecular fitness $f(g) = \mathcal{G}(g; 20, 3^2)$.

The results obtained so far by considering populations at steady state can be summarized as:

- Generically, evolving gene regulatory elements with noisy gene expression is not equivalent to evolving noiseless expression at a different effective population size. This is because the underlying molecular fitness function and noise combine in a nontrivial way to the effective fitness, which can both shift the region of selected phenotypes and change the stringency with which parts of sequence space are ‘carved out’ by selection.
- A central quantity to understand ensembles of populations at steady state is the density of states (for phenotypes in sequence space). For biophysical models of gene expression, these are computable both for noise and mean and also for their combination, yielding a two-dimensional density of states. The structure of this joint density is important in determining which noise and mean values will be observed at steady state.

From these observations we can state that gene expression noise and evolution can interact to a significant degree. The effects, however, are roughly of order unity. In the next sections, we will turn to the study of the dynamics of adaption and see that there are potentially larger effects, which could turn out to be biologically even more relevant.

3.6 Dynamics of evolution for noisy promoters

We now turn to the dynamics of adaptation and how they are influenced by the presence of gene expression noise. Previous studies have found that the formation of a transcription factor binding site can be prohibitively slow in some settings [Tuğrul *et al.*, 2015; Berg *et al.*, 2004]. These slow rates can be accelerated by various mechanisms, such as partially decayed old sites, or regulatory architectures that allow for a binding site to evolve in a longer stretch of DNA and still influence expression. This also helps to explain genomic evidence for the speed of adaptation in regulatory sequences [Dowell, 2010; Villar *et al.*, 2014]. In this section, we will see that noise can also accelerate adaptive dynamics.

To study how gene expression noise affects the dynamics of adaptation, we use a simple population genetics model including mutation, selection, and drift, similar to the ones used in [Tuğrul *et al.*, 2015] and [Berg *et al.*, 2004] and described in sections 3.6.1 and 3.6.2. To point out the effect of noise independently of the architecture that gave rise to it, we introduce an additional parameter of the fitness function in 3.6.3. With this parameter, which can be thought of as an averaging time for the selected output, we can tune the importance of noise for fitness, and thus evolution, within the same expression architecture. In addition, it is also an example for a more general dependence of fitness on the time trace of gene expression, as mentioned at the end of section 3.2.1. An expression model involving an activator and its binding site, which serves as an example for a small network where noise and mean can, to some degree, evolve independently, is introduced in section 3.6.4. Finally, section 3.6.5 shows that the presence of noise can speed up adaptation – potentially countering an intuition built from the results in the last sections.

3.6.1 Mutation rates in mismatch models

To see how regulatory sequences evolve under noise, we consider a small stretch of regulatory DNA sequence s of length $L = n_1 + n_2 + \dots + n_R$, accommodating R binding sites for regulatory proteins (RP), such as TFs and RNAPs. The binding of each RP is modeled with a mismatch model (cf. section 3.3.1), i.e. the binding site with k mismatches has the binding energy

$$E_r = E_x k_r, \quad r \in [1, R], \quad k_r \in [0, n_r]. \quad (3.53)$$

We only consider point mutations here; see [Tuğrul *et al.*, 2015] for a more general mutational regime. If such a change from one nucleotide to one of the three others occurs in a

binding site of length n , the probability of increasing the number of mismatches is

$$p_{k \rightarrow k+1} = \frac{n-k}{n} = 1 - \frac{k}{n}, \quad (3.54)$$

while the probability of decreasing it is

$$p_{k \rightarrow k-1} = \frac{k}{n} \frac{1}{3} = \frac{k}{3n}. \quad (3.55)$$

Thus, when considering all binding sites with potentially different lengths together, the probabilities for a single binding site are:

$$p_{k_r \rightarrow k_r+1} = \frac{n_r}{n_1 + \dots + n_R} \frac{n_r - k_r}{n_r} = \frac{n_r - k_r}{L} \quad (3.56)$$

$$p_{k_r \rightarrow k_r-1} = \frac{n_r}{n_1 + \dots + n_R} \frac{k_r}{n_r} \frac{1}{3} = \frac{k_r}{3L} \quad (3.57)$$

These probabilities then combine into the mutation rate matrix:

$$U_{k',k} = \begin{cases} L u p_{k \rightarrow k'} & \text{for } k' \neq k \\ -\sum_{k' \neq k} U_{k',k} & \text{for } k' = k \end{cases}, \quad (3.58)$$

with u being the mutation rate per base pair per generation.

3.6.2 Population genetics: dynamics of adaptation in monomorphic populations

As in the previous sections, we restrict ourselves to the monomorphic regime, where mutations are infrequent enough (and sweeps to fixation fast enough) to be able to always consider the population at a fixed state [Barton *et al.*, 2007; Desai and Fisher, 2007]. This also means we can view adaptation as a Markov jump process. We further simplify the setting to that of a single environment, meaning that we only have one fitness function to which the system is adapting. With these assumptions we can use the diffusion approximation for adaptive dynamics [Kimura, 1962]. After a mutant with a fitness advantage of Δf over the established genotype arises, its probability to grow to fixation p_{fix} in a population of effective size N is

$$p_{\text{fix}}(N, \Delta f) = \frac{1 - e^{-2\Delta f}}{1 - e^{-2N\Delta f}}. \quad (3.59)$$

Note that in a finite population, even mutants with a fitness disadvantage ($\Delta f < 0$) still have a positive fixation probability.

For a single binding site, the transition rates from one mismatch class to another, summarized in the matrix \mathbf{R} , are therefore

$$\mathbf{R}_{k',k} = \begin{cases} 2N U_{k',k} p_{\text{fix}}(N, \Delta f_{k',k}) & \text{for } k' \neq k \\ -\sum_{k' \neq k} \mathbf{R}_{k',k} & \text{for } k' = k \end{cases}, \quad (3.60)$$

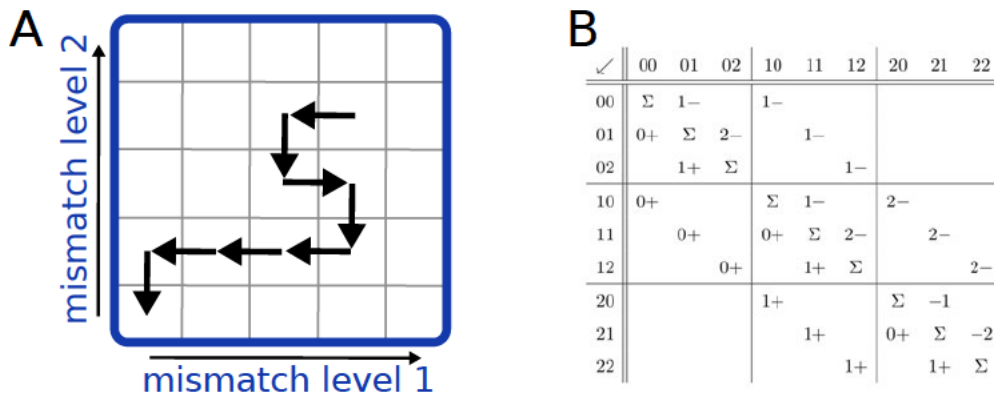


Figure 3.8: **Modeling the simultaneous evolution of several binding sites.** **(A)** A sample trace of genotypes in the two dimensional genotype space ($L = 4$) that a population could take to reach the lowest mismatch class for both binding sites (lower left corner). The transition rates in equation 3.60 determine the likelihood of any such trace. **(B)** A convenient scheme to organize more than one binding site into a mutation rate matrix U for two binding sites with length $L = 2$ (mismatch classes ranging from 0 to 2). Established sequences are indicated above the matrix and target sequences on the left side. Entries followed by a plus-sign (minus-sign) indicate an increase (decrease) in the mismatches starting from the number of mismatches indicated before the sign, empty entries are 0 as these genotypes are not mutational neighbours, and Σ is the normalizing entry in every column, ensuring preservation of probability.

and accordingly for the $(L + R) \times (L + R)$ -matrix accommodating the mismatch classes of all R binding sites (see Figure 3.8).

Considering a large ensemble of monomorphic populations at time t (in generations), we signify the state of this ensemble with $\psi(t)$, a vector in which every component represents the probability of observing a genotype with the corresponding mismatch class across the binding sites. The temporal change of $\psi(t)$ is then

$$\frac{d}{dt}\psi(t) = \mathbf{R} \cdot \psi \quad , \quad (3.61)$$

in the continuous time limit. This equation has the solution:

$$\psi(t) = e^{\mathbf{R}t} \cdot \psi(0) \quad , \quad (3.62)$$

which can be computed numerically once the fitness differences are known.

3.6.3 Fitness model with time averaging

In the previous sections, we studied the influence of changing biophysical parameters to emphasize the importance of noise structure on evolutionary quantities. Since we here consider freely evolving noise, determined by an evolving regulatory network, we use an additional parameter of the fitness function to vary the noise level (see Figure 3.9A). Mathematically, we will represent this by replacing the distribution $P(g; \mu, \sigma)$ in equation 3.1 with $P(g; \mu, \alpha\sigma)$, where $\alpha \in [0, 1]$. This then gives the following definition for effective fitness:

$$\hat{f}(\mu, \sigma; f(g), \alpha) = \int f(g)P(g; \mu, \alpha\sigma) dg, \quad \alpha \in [0, 1] \quad . \quad (3.63)$$

This new parameter α can be interpreted as a measure for how long the system integrates the gene expression level under selection, before evaluating its fitness. It should therefore not be viewed as a parameter of the evolving regulatory system but rather as an additional parameter of the fitness function or a property of the downstream network. We can interpret the integration time variable α in terms of the number of independent measurements n_τ needed to reach this reduction in standard deviation (reflecting the auto-correlation time τ of the expression level). We then get $\alpha\sigma = \sqrt{n_\tau\sigma^2}/n_\tau$, or $n_\tau = 1/\alpha^2$.

The underlying idea that not only the instantaneous level of gene expression, but its average in some window of time is relevant to selection may be especially relevant in developmental biology [Raj *et al.*, 2010]. In principle, also more complicated features of the dynamics in a window could be relevant [Hansen and O'Shea, 2015], but we will restrict ourselves to time averaging here.

3.6.4 Expression model for a small regulatory network

To study the influence of noise on the dynamics of adaptation, we look at a sequence-to-expression model that allows for several noise levels with the same mean (cf. section 3.3.2). We choose a minimal network consisting of an activating TF produced from a constitutive promoter and a downstream promoter that is sensitive to this activator and produces the output protein under selection (see Figure 3.9B). One of the mismatch-classes will determine the concentration of the activator c , whereas the other will determine the activators binding energy to the downstream promoter that generates the protein under selection g . We can then calculate

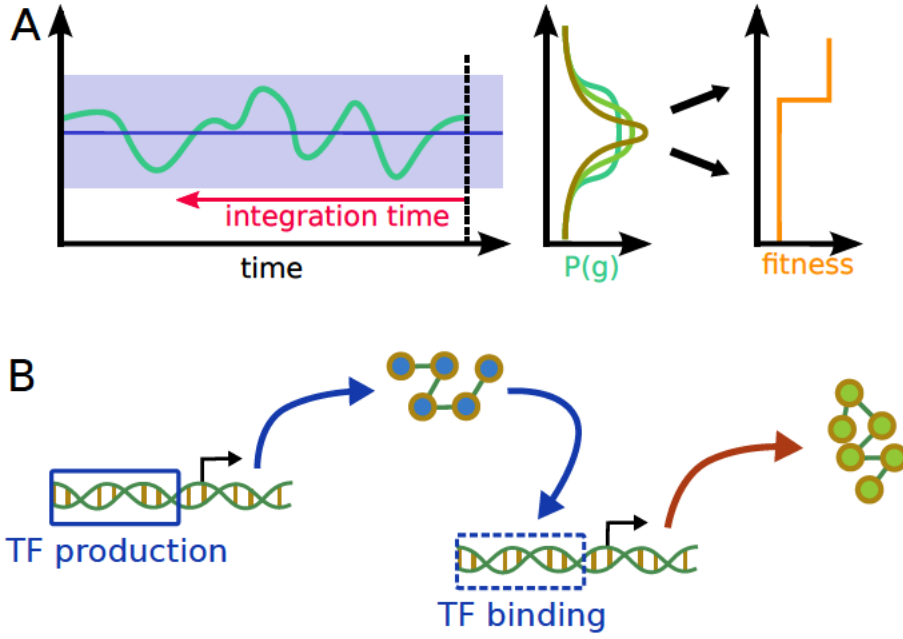


Figure 3.9: **(A)** The integration time (red), over which the expression levels are averaged before their fitness is evaluated can be modeled by altering the expression distribution that is used to evaluate effective fitness. Longer integration times correspond to narrower expression distributions. **(B)** A simple model for activator evolution. An activating TF (blue protein) is produced from a constitutive promoter, determining its concentration. This concentration together with its binding site sets the mean and noise of the protein under selection (green protein).

the first two moments of the expression distribution of the down-stream gene as follows:

$$\bar{c} = c_0 \cdot \frac{1}{1 + \exp[E_x k_1]} \quad , \quad (3.64)$$

$$\sigma_c^2 = \nu_c \bar{c} \quad , \quad (3.65)$$

$$\bar{g} = g_0 \cdot \frac{1}{1 + \exp[E_x k_2 - \log[\bar{c}]]} \quad , \quad (3.66)$$

$$\sigma_g^2 = \nu_g \bar{g} + (\partial_c \bar{g})^2 \sigma_c^2 \quad , \quad (3.67)$$

where mean expression levels are denoted with a bar ($\bar{\cdot}$) and σ are standard deviations. The parameters are $k_1, k_2 \in [0, 8]$, $E_x = 1$, $c_0 = 100$, $\nu_c = 4$, $g_0 = 100$, and $\nu_g = 4$. The form of the noise terms are motivated by a Poisson-like noise from bursty gene expression for the TF (with prefactor ν_c , see sections 3.2.2 and 1.3), and a combination of Poisson-like noise (with ν_g) and transmitted noise from the upstream TF for the downstream gene under selection (cf. section 1.4). This also means that, consistent with the results of Chapter 1, the noise level of both genes will depend on its mean expression levels. For the mean expression levels, we use a thermodynamic model of gene expression (cf. equation 3.43). The mismatch level of the upstream promoter (k_1) determines the expression level of the TF, and the mismatch level of

the downstream promoter (k_2) governs the binding strength of individual TF molecules to the binding site in this promoter. This allows for several combinations of TF levels and binding site strengths that result in the same mean output level of the downstream gene – yet the noise levels of it will be different.

We note that in this small network both, the mean and the noise of the downstream gene depend on both, the level of the upstream TF and the strength of the binding site. This means that even though selection only directly acts on the downstream gene, both promoters are expected to change during adaptation. Furthermore, the network has the ability to implement the same mean expression level in various ways and of those, some will display different levels of noise. Thus, when selecting on gene expression, not only the final outcome (in steady state) for mean and noise will depend on the architecture – but also the path a population takes to reach this state.

3.6.5 Trade-off in evolutionary dynamics

To see how noise changes the dynamics of adaptation, we will now look at the minimal activator network described in the previous section and use the population dynamics model from sections 3.6.1 and 3.6.2. The ensemble of populations starts out with the neutral distribution of genotypes, i.e. the stationary distribution that is reached if all fitness differences in equation 3.60 are set to zero. This distribution is dominated by sequences for which the expression levels of both proteins is low. From there, according to equation 3.62, we simulate the dynamics of adaptation forward in time until a steady state is reached. Figure 3.10A shows the dynamics of mean and noise for truncating selection on the output of the activator network. We see that the mean expression level of the selected output rises over time and eventually crosses the selection threshold. Noise, on the other hand has a less pronounced increase. The trajectory of the Fano factor shows that the populations only settle on the smallest possible value after adaptation has reached a steady state.

To emphasize the effects of evolving noise in the biophysical model, we compare it to a model where noise is not implemented in a biophysical way. To keep this alternative as comparable as possible, we use the same model for mean expression and a constant, i.e. genotype-independent, value for the variance:

$$\bar{c} = c_0 \cdot \frac{1}{1 + \exp[E_x k_1]} \quad , \quad (3.68)$$

$$\bar{g} = g_0 \cdot \frac{1}{1 + \exp[E_x k_2 - \log[\bar{c}]]} \quad , \quad (3.69)$$

$$\sigma_g^2 = V_e^2 \quad , \quad (3.70)$$

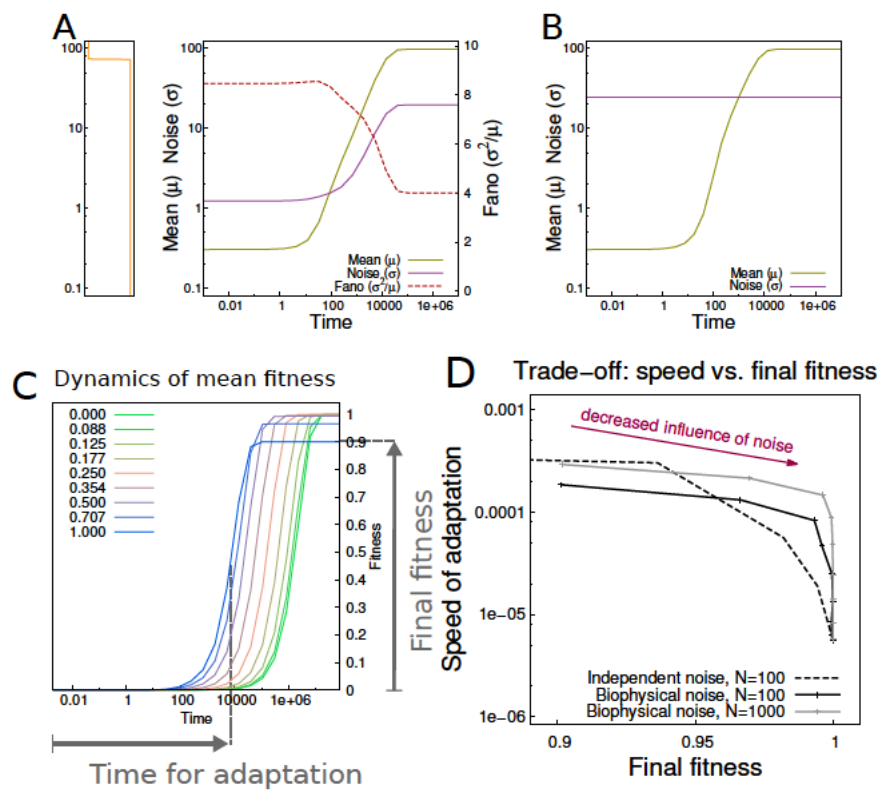


Figure 3.10: **Effect of integration time and noise structure on evolution.** (A) Dynamics of adaptation for mean and noise for an activator network with a biophysical noise model. Shown are average values over an ensemble of populations. The molecular fitness function is a step function with the step at $g = 73$, population size $N = 100$, mutation rate $u = 0.01$, and $\alpha = 1$. Time is measured in generations. (B) Same as (A) for a noise model with fixed noise. (C) Dynamics of adaptation for average fitness of the ensemble of populations for different integration times. The choices for α , indicated in the upper left corner, correspond to doubling the number of independent measurements n_τ . (D) Codependence of speed of adaptation and final fitness. Final fitness is the fitness the ensemble reaches at steady state and speed of adaptation is the inverse time to reach half the final fitness, see also (C). The influence of noise is changed continuously by varying (inverse) integration time α and curves are plotted as implicit functions of α . We compare two noise models: a biophysical model (solid lines), and a model with constant noise for comparison (dashed line). The results are identical for $\alpha = 0$, as expected. The benefit of higher noise (smaller α) is faster adaptation. The downside, however, is a ‘noise load’ (lower final fitness). This trade-off is more pronounced for the fixed noise model, as can be seen from comparing the curves for intermediate values of α . To compare the noise effects to those of population size, we also show the trade-off for the biophysical model for a larger population size (gray solid line, $N = 1000$), which generally speeds up adaptation. Note that for this plot, the value of the V_e is not important, provided that it is large enough to allow for tracing out the whole range of noise by tuning α .

with $V_e = 25$ and all other parameters as stated in section 3.6.4. In addition to the mathematical convenience, this is also equivalent to evolutionary models that incorporate non-genetic, or environmental variance but do not resolve molecular details [Bull, 1987; Zhang and Hill, 2005]. Figure 3.10B shows the dynamics of mean and noise for this model with fixed, non-evolving noise. We can already see that choosing a different noise model has changed the dynamics for the mean level.

To further explore this impact of noise on the evolutionary dynamics we make use of the model-independent way to tune the influence of noise on fitness developed in section 3.6.3. In Figure 3.10C, we show the dynamics of fitness for different, fixed (non-evolving) degrees of time averaging α . From this, we can see that prolonging integration time (by lowering α) greatly changes the dynamics of adaptation. For high values of α , fitness rises much earlier than for the noiseless case. On the other hand, quicker adaptation not only reaches its steady state earlier but also the final fitness is lower than for the cases where the impact of noise on fitness is smaller. Thus, for populations far away from a selection threshold, the benefit of higher noise is faster adaptation. This can be understood in the light of the wider effective fitness curves, that can exist at the flanking regions of fitness maxima (see figure 3.3A,B). As is expected from the analysis of the steady state in the preceding sections, higher noise does lower the fitness peak, i.e. the maximal obtainable fitness will deviate from the noiseless case if there are no solutions in genotype space that allow for a low enough noise level. In the case of longer time averaging, however, the populations are less constrained by the availability of low noise solutions for means above the selection threshold – thus reducing the deviation from maximal fitness due to noise (the noise load).

This trade-off is summarized in figure 3.10D by plotting the speed of adaptation against the final fitness for different values of α . We can now ask how severe this trade-off between the speed of adaptation and noise load is for the different noise models introduced above. For intermediate values of α , the speed-fitness trade-off is much less pronounced for the biophysical model than for the independent noise model. This can be understood from the flexibility that the biophysical model offers – high noise solutions, that are beneficial when the mean of the expression is much lower than the selection threshold can evolve to lower levels of noise as the mean proceeds to higher levels. The terms high and low noise solutions have to be interpreted in the sense that these noise levels are still constrained by the Poisson limit built into the biophysical model (see also the trajectory of the Fano factor in 3.10A). This type of flexibility is absent in the case of genotype-independent noise, where the noise level can not differ in early and late stages of adaptation.

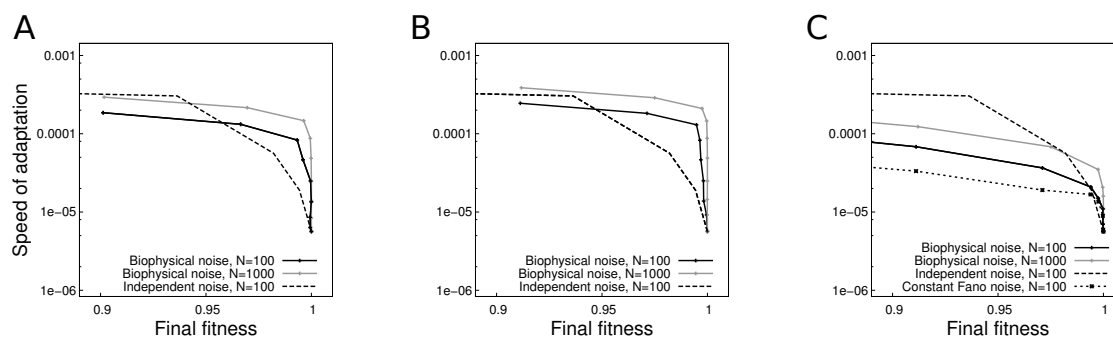


Figure 3.11: **Varying the relative importance of noise strengths.** To see how the noise strengths ν_c and ν_g influence the trade-off in evolutionary dynamics, (B) and (C) show two different extremes. For comparison, (A) shows the case of equal strength ($\nu_c = \nu_g = 4$, same as Figure 3.10D) and the model for independent noise ($V_e = 25$) is replotted in all three panels (dashed lines). As can be seen in (B), larger upstream noise ($\nu_c = 16$, $\nu_g = 4$) increases the degree to which the trade-off is alleviated. Increasing the downstream noise term ν_g in (C) ($\nu_c = 4$, $\nu_g = 16$) brings the result closer to a model with Poisson-like noise only, i.e. a constant Fano factor (additional dotted line, $\nu_c = \sigma_c = 0$, $\nu_g = 16$).

Finally, we can ask how the relative strengths of the different noise sources in the model, ν_c and ν_g , influence the evolutionary dynamics. Figure 3.11 shows that stronger upstream noise ($\nu_c > \nu_g$) will further alleviate the speed-fitness trade-off for intermediate values of α . This contrasts with the situation when $\nu_g > \nu_c$, where the situation starts to resemble that of a noise model with constant Fano factor for the gene under selection $\nu_c = 0$.

3.6.6 Discussion

Generally, one might have expected that the smoothing effect of noise on the fitness landscape (cf. section 3.2.3) makes evolution proceed slower, since the selective advantage of a certain difference in mean expression values is lessened by the effects of noise. Here we have shown that this is not necessarily the case since the flatter effective fitness function also stretches out further into regions of low-fitness phenotypes. Thus, the effect of slower progression due to shallower progression can be overcome and increased noise can even speed up adaptation compared to the noiseless case. This effect of noise is potentially much larger than the ones discussed in the steady state setting (cf. sections 3.4 and 3.5), as speed of adaptation varies over several orders of magnitude over the range of α .

Furthermore, as indicated in figure 3.10D, for the same noise level, a larger population will

also lead to faster adaptation. This means, however, that in terms of the speed of adaptation, the effects of noise and population size are actually opposite to what we would expect from a direct analogy to sampling noise (cf. section 3.2.2): high noise in a gene regulatory network under directional selection can speed up adaptation, whereas in small populations the time to find fitter genotypes will increase.

Several theoretical studies have shown how increased noise can be beneficial in varying environments [Bull, 1987; Zhang and Hill, 2005; Wolf *et al.*, 2015], especially when regulation is not able to track the changes in the environment. These cases are conceptually similar to the beginning of adaptation in our examples, when high fitness solutions are so far away from the bulk of the populations that only high noise can help sense and eventually reach them. This early benefit of reaching into a high-fitness region by noise also is key in the accelerated adaptive dynamics. It is also categorically different from other mechanisms that have been proposed to over-come the long waiting times for the arrival of transcription factor binding sites. Among those are decayed old binding sites, short binding motifs, and long regulatory regions [Tuğrul *et al.*, 2015; Berg *et al.*, 2004]. These mechanisms are mainly based on facilitating the finding of a binding site, rather than providing adaptive advantages to weakly binding sites, as can be achieved by gene expression noise.

Finally, this speed-up is also a qualitatively new example of beneficial effects of noise. It is sensing remote fitness peaks and provides ways to start moving towards them much earlier than in the noiseless case. This contrasts with situations where the beneficial effects of noise come from compensating for a lack of sufficient sensing, as for example suggested in [Wolf *et al.*, 2015]. In this paper, the authors show that fitness in a changing environment can be increased when an a downstream gene under selection gets linked to a noisy upstream TF. While this is a statement about the fitness of different genotypes, it also suggests an important mechanism for the evolutionary onset of gene regulation that is not unlike the results presented in this section, which only deal with a single environment.

For future research it will be interesting to investigate more general models of gene expression than the one studied here. The model of an evolving activator concentration and binding site in sections 3.6.4 and 3.6.5 is far too restrictive to model the full flexibility available to many biological systems. A first generalization can be to describe the production process of proteins in more detail by explicitly including mRNA in the model. This will allow the observation of interesting noise parameters over time, as the number of proteins per mRNA molecule can be linked to the parameters ν_c and ν_g of our model [Tkačik *et al.*, 2008a; Rieckh and Tkačik, 2014]. Furthermore, an exploration of how integration time could evolve

seems interesting. Here we treated it as a fixed property of the fitness function, but an interpretation as a property of the downstream network that reads the gene under selection could be an interesting extension. A central question here would be, what control mechanisms allow for time averaging and how easily they can adapt to different integration times. An interesting starting point could be provided by the mechanism of (spatial) averaging between neighboring cells [Sokolowski and Tkačik, 2015]. Finally, an extension of the framework for evolutionary dynamics developed here to also include changing environments could prove to be fruitful. On the one hand, this will enable us to elucidate the differences between the activating scheme used here and models of constitutive expression. On the other hand, such an extension will enable us to further elucidate the links to the stationary results from [Wolf *et al.*, 2015]. Specifically, it would be interesting to study for which types of regulatory interactions and environmental statistics noisy TFs can arise and stabilize in a population.

On the experimental side, a direct test of the predictions for the speed-up of adaptation due to noise could be attempted. Such an effort would profit from an experimental setup including a microfluidic device in combination with a microscopy platform that can accommodate a population large enough to make experimental evolution feasible.

Concerning the genetics of mean and noise in gene expression, there have recently been some efforts to characterize genetically closely related variants of promoters [Metzger *et al.*, 2015; Jones *et al.*, 2014]. However, the sheer size of genotype space might quickly outgrow experimentally feasible limits, emphasizing the need for predictive modeling based on measurements from only a sub-space of possible sequences. See [Jones *et al.*, 2014] for recent progress in this direction. Combining this kind of detailed molecular knowledge with modern bar-coding techniques to track very large populations (see for example [Levy *et al.*, 2015]) has the potential to produce even richer data-sets to elucidate the role of noise in evolution.

4 A genetic platform to study single cell stress response

In this chapter I report on work done in collaboration with Karin Mitosch and Tobias Bollenbach. All measurements in the microfluidic chamber have been performed by Karin Mitosch.

4.1 Introduction

Gene expression levels can be seen as fundamental phenotypic traits that selection can act on, as was discussed in the previous chapter. Depending on the evolutionary scenario, selection could either act on the instantaneous gene expression levels or some temporally averaged version thereof, and a full knowledge of these phenotypes thus requires us to measure complete single cell traces of gene expression as a function of time. The kind of data that can be collected this way is helpful for both, understanding the mean regulation of a gene, but also how deviation in single cells from this 'typical' behavior influences the non-genetic component of fitness.

Here, I will first explain a technique that can be used to obtain relevant reporter strains, and then an application wherein one of these strains was used to describe single cell expression dynamic and survival under stress.

4.2 A cloning strategy for the construction of chromosomal reporter constructs

To understand the dynamics and function of a gene, a common approach is to fuse the promoters of the gene of interest (POI) to a reporter gene, such as a fluorescent protein. For *E. coli* K12 (MG1655) a collection of intergenic regions, many of which contain important regulatory

regions and promoters, has been constructed [Zaslaver *et al.*, 2006]. The cloning strategy described in this section is designed to make use of this library of reporter plasmids (and similar ones) to facilitate the construction of single or multiple chromosomal reporter strains for various promoters from this library.

When studying single cell effects of gene expression in microbes it can often be beneficial to have chromosomally integrated reporters for gene expression, rather than reporter plasmids [Lin-Chao and Bremer, 1986; Paulsson and Ehrenberg, 2001; Pedraza and van Oudenaarden, 2005; Wong Ng *et al.*, 2010; Tal and Paulsson, 2012] – especially when studying them in stressed conditions. Constructing strains with chromosomal reporters in the bacterium *E. coli* can be challenging because integrations get less efficient as the DNA molecule that needs to be integrated gets longer [Kuhlman and Cox, 2010]. Another disadvantage of classical recombineering [Datsenko and Wanner, 2000] is the use of long primers (70 nucleotides and longer), which are expensive and harder to design and handle than shorter ones, which are usually used in PCRs. The technique described here circumvents these two issues by dividing the procedure of getting the desired resistance-marker-promoter-reporter construct into the chromosome into two steps. The first one deals with the integration of a long construct using long primers. In the second step, this long construct will serve as a ‘landing platform’ for a much shorter fragment carrying the POI. The first, inefficient step only has to be performed once, and after the successful integration of the platform, the resulting strain can then be used multiple times to produce working chromosomal reporter constructs.

The construct from the first step serves as a platform for accepting shorter segments (obtained by PCR with shorter primers) in the second step. For this to work, the strategy needs three key parts that differ from the conventional recombineering protocol: (1) a selectable marker for the platform in the first step, (2) a selectable marker that only works once the desired final construct has been completed in the second step but is non-functional in the first step, and (3) homology regions for the second integration that can be accommodated even with short primers.

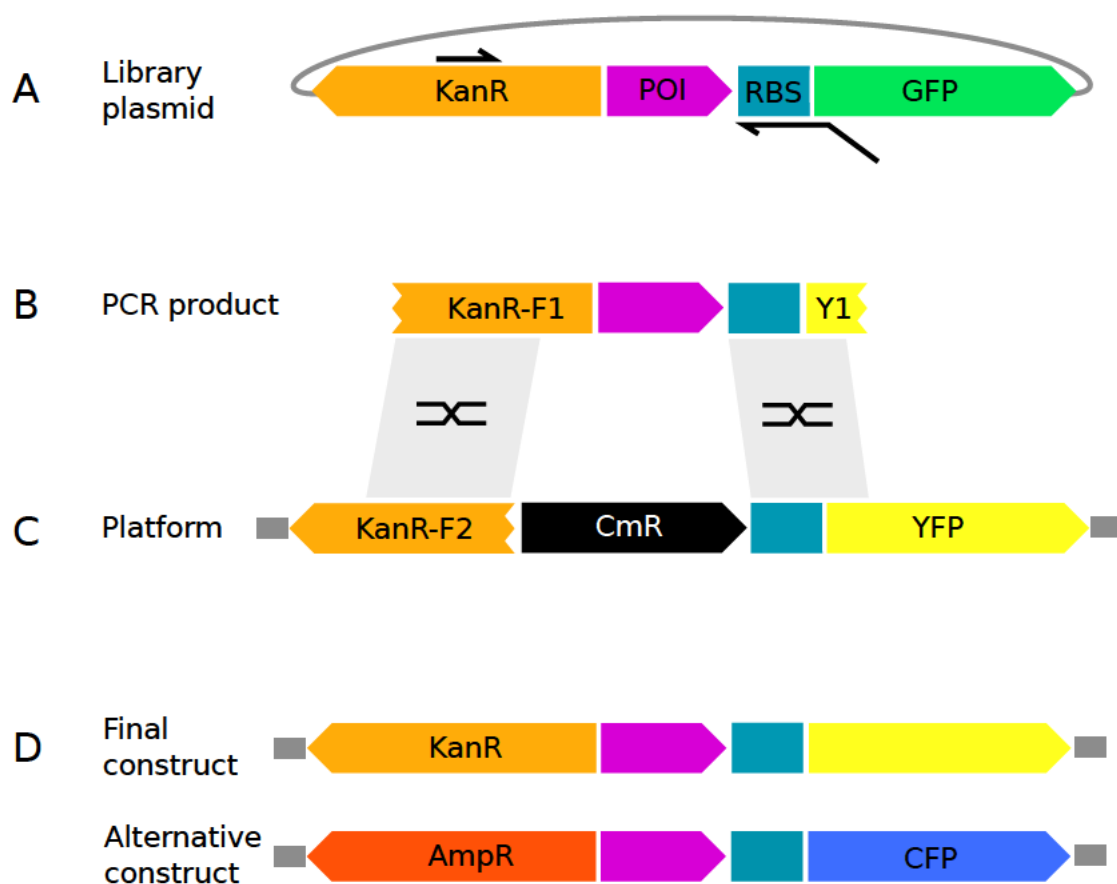


Figure 4.1: **(A) Plasmid promoter library.** The relevant parts for the further use of the reporter plasmids in the library from [Zaslaver *et al.*, 2006] to produce chromosomal reporter fusions are the kanamycin resistance cassette (KanR), the intergenic region (POI), and the fluorescence gene including its 5'-UTR (RBS-GFP). **(B) PCR product for homologous recombination.** The homology regions are in the coding region of the kanamycin resistance and the 5'-UTR plus the first few codons of the YFP (Y1). The YFP-part comes partly from the primer, as indicated in (A). **(C) Integration platform.** The platform including the non-functional kanamycin resistance cassette (KanR-F2) and the reporter gene (YFP), which can differ from the original one, has been integrated into the chromosome (thick gray lines) using the chloramphenicol resistance cassette (CmR). The platform can now be used to accept the linear DNA fragment from (B) that has been amplified from a chosen library plasmid in (A). **(D) Desired reporter construct.** Upon successful integration, the kanamycin resistance cassette is completed by joining KanR-F1 and KanR-F2 and can serve as a selectable marker; at the same time, this deletes the chloramphenicol resistance cassette. Since the 5'-UTR is used for homology, the process can be used to swap reporter genes (such as YFP or CFP instead of GFP). For the use of two reporters on the same chromosome, changing the resistance marker (e.g. to ampicillin resistance, AmpR) in the library plasmid is necessary.

4.2.1 Construction of a plasmid containing the platform

To construct the platform, the promoterless plasmid from [Zaslaver *et al.*, 2006], with a pZA origin [Lutz and Bujard, 1997], was used. First, the reporter gene was changed from GFPmut2 to either a YFP or a CFP variant (sequences obtained from Tobias Bergmiller, as described in [Cox *et al.*, 2010]), which was achieved by using primers GR-cYFP-1 and GR-cYFP-2 for YFP, and GR-cCFP-1 and GR-cCFP-2 for CFP (see table 4.1) and the restriction enzymes HindIII and NdeI (New England Biolabs). This resulted in plasmid pFIL-koy and pFIL-koc. These two variants of fluorescent proteins were chosen since they are monomeric and fast maturing (maturation times are below 10 minutes); furthermore they can be used in channels with reduced background fluorescence in many conditions, and if used simultaneously in one set-up have minimal bleed-through.

The second modification is to provide a selectable marker for the platform that will be knocked out upon successful integration of the final reporting construct. For this, the chloramphenicol resistance cassette from the Lutz&Bujard-library [Lutz and Bujard, 1997] was put between XhoI and BamHI restriction sites with primers GR-CmR-1 and GR-CmR-2, and cloned into the plasmid pFIL-koy and pFIL-koc, resulting in the plasmids pFIL-kcy and pFIL-kcc.

In a third step, the intact kanamycin resistance cassette was replaced by a defunct fragment, starting after the start codon of its protein coding region. For this, whole-plasmid PCR was used with primers GR-CmR-1 and GR-KnF-2, which also contains an XhoI restriction site. This yields the plasmid containing the desired platforms pFIL-fcy and pFIL-fcc.

After completion of these plasmids, the platforms were integrated into two different chromosomal locations (galK and intS) using lambda-red-recombineering as described in [Datsenko and Wanner, 2000], with the recombineering plasmid pKD46 and primers GR-intS-1 and GR-intS-2, or GR-galK-1 and GR-galK-2. Finally, all integrated platforms were checked for mutations by sequencing the PCR product obtained by using primers GR-intS-up, GR-intS-dn, GR-galK-up, GR-galK-dn on the chromosomal DNA.

4.2.2 Chromosomal integration of promoters of interest

The step that will be performed several times after establishing the platform from the last section on the chromosome is the actual integration for the promoters of interest. For this, the POI and the necessary homology regions, the promoter driving the kanamycin resistance gene and its start codon are simply amplified via PCR and the primers GR-MKan-1 and GR-YFP-RBS or GR-CFP-RBS. This provides a comparatively short linear fragment, that still has large ho-

GR-cYFP-1	AGAAA GGATCC GAGAAGAACTTTTCACTGGAG
GR-cYFP-2	ATGAC CTCGAG CTGAATGAACTGCAGGAC
GR-cCFP-1	GTCCG GGATCC TCTAGATTTAAG
GR-cCFP-2	CTCGAG G GGATCC TCTAGATT
GR-CmR-1	AGATA CTCGAG GTGAAGACGAAAGGG
GR-CmR-2	AGAAT CTCGAG TAGACGTCGATATCTGGCG
GR-KnF-2	AGAAT CTCGAG GATATCTGGCGAAAATGAGAC
GR-intS-1	CCGTAGATTTACAGTTCGTCAATGGTTCGCTTCAGATCGTTGACAGCCGCA- -GAGTCAGTGAGCGAGGAAGC
GR-intS-2	ATAGTTGTTAAGGTCGCTCACTCCACCTTCTCATCAAGCCAGTCCGCCCA- -TGAAGTCAGCCCCATACGAT
GR-galK-1	GTTTGCGCGCAGTCAGCGATATCCATTTTCGCGAATCCGGAGTGTAAGAA- -GAGTCAGTGAGCGAGGAAGC
GR-galK-2	ACCATCGGGTGCCAGTGCGGGAGTTTCGTTTCAGCACTGTCCTGCTCCTTG- -TGAAGTCAGCCCCATACGAT
GR-intS-up	GTA CTTACCCCGCACTCCAT
GR-intS-dn	TGTT CAGCACACCAATAGAGG
GR-galK-up	CCTACTCTATGGGCTGGCAC
GR-galK-dn	GGAAAAGTAAAGTCGCACCCC
GR-MKan-1	GCGATACCGTAAAGCACGAG
GR-YFP-RBS	GAACAGTTCTTCACCT-TTGCTCATATGTATATCTCCTTC
GR-CFP-RBS	GACCAGGATAGGAACACACCAGTAAACAGCTCCTCGCCC- -TTGCTCATATGTATATCTCCTTCTTAAATCTAGAG

Table 4.1: Primers for the construction and application of integration platforms. Restriction sites are surrounded by spaces. Integration primers contain two parts, separated by a hyphen, that correspond to chromosomal homology regions and PCR annealing regions.

mology regions and completes a selectable marker upon successful integration. This can now be used in a standard recombineering protocol to produce the desired chromosomal reporter strains. For the production of strains with different reporter genes (in our case CFP and YFP) the same template plasmid can be used – only one of the primers should be changed to provide a longer homology reaching into the coding region.

After the identification of successful integrands, the sequence of the integrated promoter can be checked by using the same primers as when checking the integration of the platform (e.g. GR-intS-1 and GR-intS-2). Additionally, checking for growth rate and unchanged sensitivity to stresses is advisable.

The described method can be used to study the details of regulation on complex promoters (cf. Chapter 1) and fitness effects stemming from different dynamics of expression in promoter variants (cf. Chapter 3). For this it is necessary to have a method to generate a diverse set of similar promoters and put them in the same context [Kinney *et al.*, 2010]. Furthermore, in the quest for a more detailed understanding of how natural pathways perform their function, an efficient way to produce reporter constructs is beneficial. In the next section a simple application will be presented.

4.3 Application: Promoter activity and survival under acid stress in single cells

In this section, the cloning strategy described in the last section is used to study complex promoters (cf. Chapter 1) and demonstrate that gene expression is a phenotype that selection can act upon (cf. Chapter 3).

4.3.1 Bacterial stress response and cross-protection

Bacteria can find themselves in harsh environments, where they not only encounter single stress factors but rather combinations of multiple stresses. They have evolved a number of different stress response systems to increase their chance of survival under stress. Some of these systems offer protection not only against one stress, but also against other, subsequently encountered challenges. This phenomenon is called cross-protection [Al-Nabulsi *et al.*, 2015; Jenkins *et al.*, 1988; Leyer and Johnson, 1993; McMahon *et al.*, 2007]. Under antibiotic-stress, the metabolism and gene expression of microorganisms can change markedly [Brazas and Hancock, 2005; Kwon *et al.*, 2010; Sangurdekar *et al.*, 2011], which can also be the reason for cross-protective effects.

It has previously been observed that stress response genes tend to have an increased level of cell-to-cell variability [Newman *et al.*, 2006; Silander *et al.*, 2012]. A cross-protecting effect for a subsequent antibiotic stress has been reported for nutrient pre-stress [Arnoldini *et al.*, 2014] and for noisily expressed antibiotic resistance genes [El Meouche *et al.*, 2016]. In these cases, high gene expression is beneficial for survival under the subsequent stress condition; however, also a case where the opposite, namely high expression of a gene is harmful for survival has been reported [Ni *et al.*, 2012]. Here, we show that antibiotics can also cross-protect from other,

non-antibiotic, stressors, and the survival time under the second stress is influenced by gene expression noise.

4.3.2 Single cell survival can be predicted by activity of the acid stress promoter P_{gadB}

As a pre-stress that induces a stress response, we chose trimethoprim (TMP), a folic-acid synthesis inhibitor. Since TMP cross-protects against acid stress and up-regulates the expression from the *gadB* promoter (P_{gadB}) [Mitosch, Rieckh, Bollenbach; unpublished], we wanted to know if information on the expression from this promoter with single cell resolution can predict survival of individual cells. To this end we constructed a reporter for the *gadB* promoter using a YFP-based platform at the *intS*-locus ($P_{gadB} - YFP$) into the chromosome, as described in the previous section. P_{gadB} controls the expression of two proteins, a glutamate decarboxylase (GadB) and a glutamate:4-aminobutyrate antiporter (GadC). Both are known to have a key role in survival at low pH [Richard and Foster, 2004]. The GadB protein catalyzes the binding of protons to glutamate. The product Gamma-aminobutyric acid (GABA) is then exchanged for glutamate by the antiporter GadC. This lowers the intracellular proton concentration [Hersh *et al.*, 1996; Tsai *et al.*, 2013].

The activity of the P_{gadB} -reporter construct under TMP and acid stress was then recorded in a microfluidic device (see Figure 4.2A). This showed a strong up-regulation within three hours after TMP addition, accompanied by a large variation in expression level: some cells showed no detectable change, while some others rose about 30-fold (see Figure 4.2A).

We then asked how well the response induced by TMP predicts the time until cell lysis (termed survival time) after the stress was switched to hydrochloric acid (HCl). We observed a strong correlation between *gadB* expression in single cells just before the HCl was added and the survival time under acid stress (Figure 4.2B). A two-fold increase in *gadB* expression (and presumably in GadB) prolonged survival on average by slightly less than 2 hours.

We then wanted to see if a simple averaging scheme (i.e. a prolonged integration time for the P_{gadB} -signal), as proposed in the previous chapter, can improve the prediction of survival time. For this, we compare the RMS-error for two predictors in Figure 4.2C. The first predictor applies a uniform filter for a specified time window before the addition of the second stress (the acid stress). As a second predictor we use multiple linear regression over YFP values from the same time windows. Figure 4.2C shows that the best predictor for survival with the uniform filter is already the instantaneous value of P_{gadB} -expression at the beginning of the acid stress,

whereas linear regression is able to still exploit some information from the dynamics before the second stress (roughly 30% reduction in RMS-error for the full time scale). How much of this improvement is due to overfitting will have to be determined once more data is available.

In this work we have thus demonstrated a previously unknown cross-protective effect from an antibiotic towards another stress.¹ Moreover, we have demonstrated that knowledge of the activity of a single gene can predict survival in stressful conditions. For future research, we plan to use the cloning strategy described here to efficiently construct multiple reporters for genes in different response pathways (acid response, folate synthesis) and record the response to a number of pre-stressors. This data will then be used to understand in more detail which, potentially non-linear, features of the dynamics of gene expression contribute to the survival of single cells.

¹This result will be presented in more detail elsewhere [Mitosch, Rieckh, Bollenbach; unpublished].

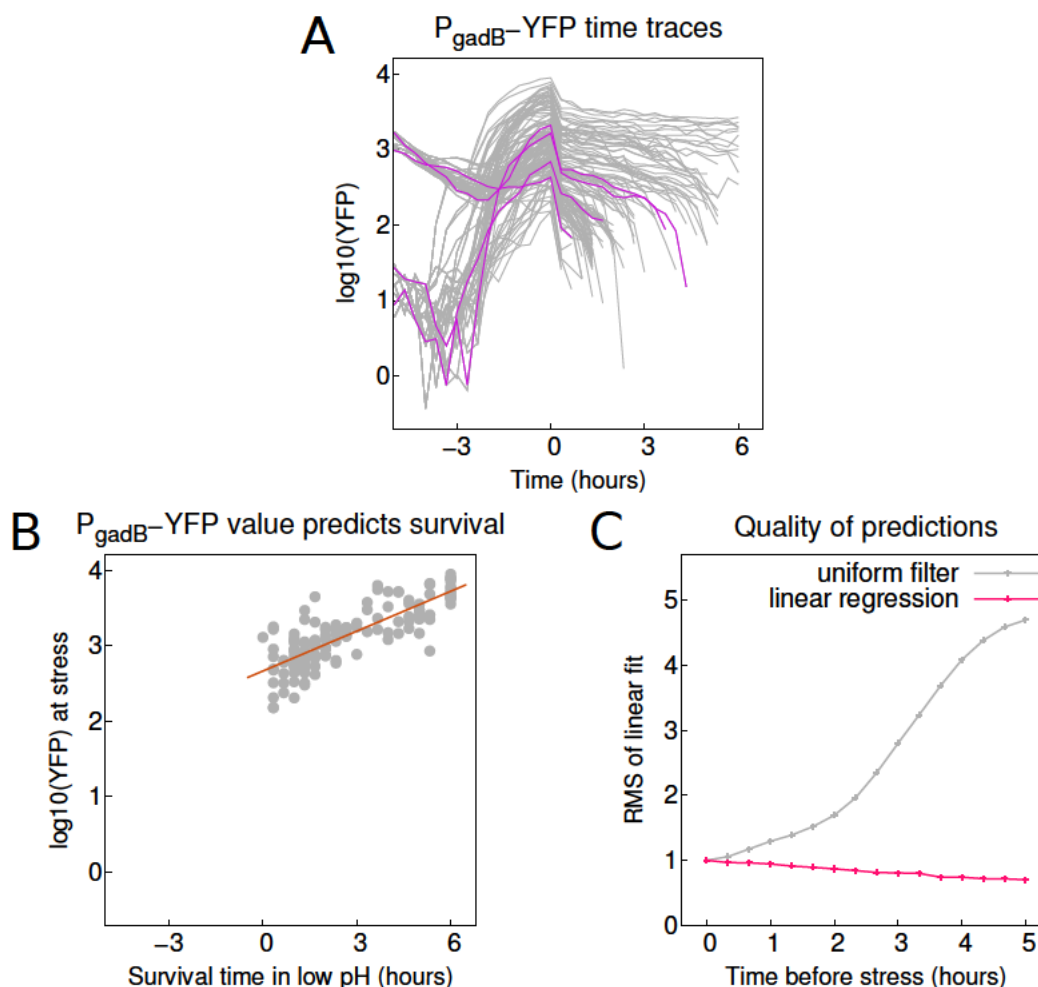


Figure 4.2: **(A) Expression from P_{gadB} over time.** Expression of $P_{gadB} - YFP$ in 122 single cells after addition of pre-stress at time -3h ($0.5 \mu\text{g/mL}$ TMP) and in the low pH stress at time 0h (switch to medium with HCl added to pH 3). The end of the YFP signal also indicates cell lysis and therefore survival time. Four typical traces are highlighted (magenta). **(B) Expression of $P_{gadB} - YFP$ predicts survival time.** Expression values of $P_{gadB} - YFP$ at the moment of pH drop plotted against single cell survival time in HCl stress, including a least squares linear fit to the data. Pearson correlation coefficient is 0.73 for 122 data points. **(C) Unweighted averaging does not improve predictive power.** The RMS-error for two model fits that take the past history of promoter activity into account are shown. One model, shown in gray, applies a uniform filter for a specified time window before the addition of the second stress (horizontal axis). The second model (pink line) is a multiple linear regression over YFP values from the same time windows. All YFP values are background-corrected and reported in arbitrary units.

4.3.3 Experimental set-up

Strains and culture conditions

We used E.coli K-12 strain MG1655 as the wildtype strain. All experiments were performed in M9 minimal medium (1x M9 salts, 1mM MgSO₄, 0.1mM CaCl₂, supplemented with 4g/L glucose and 0.1% ampicase). Antibiotic concentrations used for selection and bacterial glycerol stocks were: kanamycin, 25 $\mu\text{g}/\text{mL}$; ampicillin, 50 $\mu\text{g}/\text{mL}$; spectinomycin, 100 $\mu\text{g}/\text{mL}$. For the acid stress, the pH of the M9 minimal medium was adjusted to pH 3 with hydrogen chloride (HCl). All chemicals were ordered from Sigma Aldrich.

Microfluidics and time-lapse microscopy

Bacteria were inoculated from frozen stocks at a dilution of 1:1000 to 1:5000 and grown until an OD₆₀₀ of 0.05 to 0.1. Then they were diluted 1:100 and loaded into the microfluidics chamber (CellASIC ONIX, Merck Millipore). This normally led to spatially well separated single cells for the beginning of the experiment. All experiments were performed in a heated chamber at 30°C. Data acquisition was started after 1-2 hours. Images were taken every 10 to 20 minutes in 100x magnification with a EMCCD camera (Hamamatsu) on Nikon Eclipse Ti-E (Nikon) with Lumencor light engine (Lumencor), using NIS-Elements software (Nikon). Excitation and emission wavelength (in nm) for YFP were 513/17 and LP 520, BP 542/27, respectively.

Microfluidics data analysis

Time traces were analyzed using an adapted version of the MATLAB program *SchnitzCells* [Young *et al.*, 2012]. YFP expression levels in Figure 4.2 were determined by dividing the total fluorescence signal from a cell by its cell area. Autofluorescence was subtracted as the mean fluorescence expression of a microcolony with only the segmentation fluorophore present. Survival time was determined as the last time point at which fluorescence of the segmentation color was still above detection threshold. Maturation times of GFP, YFP and CFP were below 10 minutes in our conditions.

Bibliography

- [Acar *et al.*, 2008] M. Acar, J.T. Mettetal, and A. van Oudenaarden, “Stochastic switching as a survival strategy in fluctuating environments,” *Nat Genet*, 40(4):471–5, 2008.
- [Al-Nabulsi *et al.*, 2015] A.A. Al-Nabulsi, T.M. Osaili, R.R. Shaker, A.N. Olaimat, Z.W. Jaradat, N.A. Zain Elabedeen, and R.A. Holley, “Effects of osmotic pressure, acid, or cold stresses on antibiotic susceptibility of *Listeria monocytogenes*,” *Food Microbiology*, 46:154–60, 2015.
- [Alberts *et al.*, 2002] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, Garland Science, New York, 2002.
- [Aquino and Endres, 2010] G. Aquino and R. G. Endres, “Increased accuracy of ligand sensing by receptor internalization,” *Phys Rev E*, 81(2):021909, 2010.
- [Arnoldini *et al.*, 2014] M. Arnoldini, I.A. Vizcarra, R. Pena-Miller, N. Stocker, M. Diard, V. Vogel, R.E. Beardmore, W.-D. Hardt, and M. Ackermann, “Bistable Expression of Virulence Genes in *Salmonella* Leads to the Formation of an Antibiotic-Tolerant Subpopulation,” *PLoS Biology*, 12:e1001928, 2014.
- [Babu *et al.*, 2004] M.M. Babu, N.M. Luscombe, L. Aravind, M. Gerstein, and S.A. Teichmann, “Structure and evolution of transcriptional regulatory networks,” *Curr Opin Struct Biol*, 14(3):283–91, 2004.
- [Bajic and Poyatos, 2012] D. Bajic and J.F. Poyatos, “Balancing noise and plasticity in eukaryotic gene expression,” *BMC Genomics*, 13:343, 2012.
- [Balaban, 2011] N.Q. Balaban, “Persistence: mechanisms for triggering and enhancing phenotypic variability,” *Curr Opin Genet Dev*, 21(6):768–75, 2011.
- [Bandiera *et al.*, 2016] L. Bandiera, S. Furini, and E. Giordano, “Phenotypic Variability in Synthetic Biology Applications: Dealing with Noise in Microbial Gene Expression,” *Front Microbiol*, 7:479, 2016.

- [Barton *et al.*, 2007] N.H. Barton, D.E.G. Briggs, J.A. Eisen, D.B. Goldstein, and N.H. Patel, *Evolution*, Cold Spring Harbor Laboratory Press, 2007.
- [Barton and Coe, 2009] N.H. Barton and J.B. Coe, “On the application of statistical physics to evolutionary biology,” *J Theor Biol*, 259(2):317–24, 2009.
- [Beaumont *et al.*, 2009] H.J. Beaumont, J. Gallie, C. Kost, G.C. Ferguson, and P.B. Rainey, “Experimental evolution of bet hedging,” *Nature*, 462(7269):90–3, 2009.
- [Bel *et al.*, 2009] G. Bel, B. Munsky, and I. Nemenman, “The simplicity of completion time distributions for common complex biochemical processes,” *Phys Biol*, 7:016003, 2009.
- [Berg and Purcell, 1977] H.C. Berg and E.M. Purcell, “Physics of chemoreception,” *Biophys J*, 20(2):193–219, 1977.
- [Berg *et al.*, 2004] J. Berg, S. Willmann, and M. Lässig, “Adaptive evolution of transcription factor binding sites,” *BMC Evol Biol*, 4:42, 2004.
- [Berg and von Hippel, 1985] O.G. Berg and P.H. von Hippel, “Diffusion-controlled macromolecular interactions,” *Annu Rev Biophys Biophys Chem*, 14:131–58, 1985.
- [Berg and von Hippel, 1987] O.G. Berg and P.H. von Hippel, “Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters,” *J Mol Biol*, 193(4):723–50, 1987.
- [Bialek and Setayeshgar, 2005] W. Bialek and S. Setayeshgar, “Physical limits to biochemical signaling,” *Proc Natl Acad Sci USA*, 102(29):10040–5, 2005.
- [Bialek and Setayeshgar, 2008] W. Bialek and S. Setayeshgar, “Cooperativity, sensitivity, and noise in biochemical signaling,” *Phys Rev Lett*, 100:258101, 2008.
- [Bialek, 2013] William Bialek, *Biophysics: Searching for Principles*, Princeton University Press, 2013.
- [Bintu *et al.*, 2005] L. Bintu, N.E. Buchler, H.G. Garcia, T. Hwa, J. Kondev, and R. Phillips, “Transcriptional regulation by the numbers: models,” *Curr Opin Genet Dev*, 15:116–24, 2005.
- [Bird, 1995] A.P. Bird, “Gene number, noise reduction and biological complexity,” *Trends Genet*, 11(3):94–100, 1995.
- [Blahut, 1972] R.E. Blahut, “Computation of channel capacity and rate-distortion functions,” *IEEE Trans Info Th*, 18:460–73, 1972.

- [Blake *et al.*, 2006] W.J. Blake, G. Balazsi, M.A. Kohanski, F.J. Isaacs, K.F. Murphy, Y. Kuang, C.R. Cantor, D.R. Walt, and J.J. Collins, “Phenotypic Consequences of Promoter-Mediated Transcriptional Noise,” *Mol Cell*, 24(6):853–65, 2006.
- [Blake *et al.*, 2003] W.J. Blake, M. Kaern, C.R. Cantor, and J.J. Collins, “Noise in eukaryotic gene expression,” *Nature*, 422:633–7, 2003.
- [Bowsher and Swain, 2012] C.G. Bowsher and P.S. Swain, “Identifying sources of variation and the flow of information in biochemical networks,” *Proc Natl Acad Sci USA*, 109(20):7615–6, 2012.
- [Brazas and Hancock, 2005] M.D. Brazas and R.E.W. Hancock, “Using microarray gene signatures to elucidate mechanisms of antibiotic action and resistance,” *Drug Discovery Today*, 10:1245–52, 2005.
- [Brewster *et al.*, 2012] R.C. Brewster, D.L. Jones, and R. Phillips, “Tuning Promoter Strength through RNA Polymerase Binding Site Design in *Escherichia coli*,” *PLoS Comput Biol*, 8(12):e1002811, 2012.
- [Bull, 1987] J.J. Bull, “Evolution of phenotypic variance,” *Evolution*, 41(2):303–15, 1987.
- [Buratowski, 2009] S. Buratowski, “Progression through the RNA polymerase II CTD cycle,” *Mol Cell*, 36(4):541–6, 2009.
- [Burga *et al.*, 2011] A. Burga, M.O. Casanueva, and B. Lehner, “Predicting mutation outcome from early stochastic variation in genetic interaction partners,” *Nature*, 480(7376):250–3, 2011.
- [Cai *et al.*, 2006] L. Cai, N. Friedman, and X.S. Xie, “Stochastic protein expression in individual cells at the single molecule level,” *Nature*, 440(7082):358–62, 2006.
- [Carey *et al.*, 2013] L.B. Carey, D. van Dijk, P.M.A. Slood, J.A. Kaandorp, and E. Segal, “Promoter sequence determines the relationship between expression level and noise,” *PLoS Biol*, 11(4):e1001528, 2013.
- [Cepeda-Humerez *et al.*, 2015] S.A. Cepeda-Humerez, G. Rieckh, and G. Tkačik, “Stochastic proofreading mechanism alleviates crosstalk in transcriptional regulation,” *Phys Rev Lett*, 115:248101, 2015.
- [Charlebois, 2015] D.A. Charlebois, “Effect and evolution of gene expression noise on the fitness landscape,” *Phys Rev E*, 92(2):022713, 2015.

- [Charlebois *et al.*, 2011] D.A. Charlebois, N. Abdennur, and M. Kaern, “Gene expression noise facilitates adaptation and drug resistance independently of mutation,” *Phys Rev Lett*, 107(21):218101, 2011.
- [Cheong *et al.*, 2011] R. Cheong, A. Rhee, C.J. Wang, I. Nemenman, and A. Levchenko, “Information transduction capacity of noisy biochemical signaling networks,” *Science*, 334(6054):354–8, 2011.
- [Cho *et al.*, 1999] H. Cho, T.K. Kim, H. Mancebo, W.S. Lane, O. Flores, and D. Reinberg, “A protein phosphatase functions to recycle RNA polymerase II,” *Genes Dev*, 13(12):1540–52, 1999.
- [Coulon *et al.*, 2010] A. Coulon, O. Gandrillon, and G. Beslon, “On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter,” *BMC Sys Biol*, 4:2, 2010.
- [Cover and Thomas, 2006] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Interscience, 2006.
- [Cox *et al.*, 2010] R.S. Cox, M.J. Dunlop, and M.B. Elowitz, “A synthetic three-color scaffold for monitoring genetic regulation and noise,” *Journal of Biological Engineering*, 4:10, 2010.
- [Crick, 1970] F. Crick, “Central dogma of molecular biology,” *Nature*, 227(5258):561–3, 1970.
- [Datsenko and Wanner, 2000] K.A. Datsenko and B.L. Wanner, “One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products,” *PNAS*, 97:6640–5, 2000.
- [Davidson and Surette, 2008] C.J. Davidson and M.G. Surette, “Individuality in Bacteria,” *Annu Rev Genet*, 42:253–68, 2008.
- [de Ronde *et al.*, 2009] W.H. de Ronde, B.C. Daniels, A. Mugler, N.A. Sinitsyn, and I. Nemenman, “Mesoscopic statistical properties of multistep enzyme-mediated reactions,” *IET Syst Biol*, 3(5):429, 2009.
- [de Ronde *et al.*, 2010] W.H. de Ronde, F. Tostevin, and P.R. ten Wolde, “Effect of feedback on the fidelity of information transmission of time-varying signals,” *Phys Rev E*, 82:031914, 2010.
- [de Ronde *et al.*, 2011] W.H. de Ronde, F. Tostevin, and P.R. ten Wolde, “Multiplexing biochemical signals,” *Phys Rev Lett*, 107:048101, 2011.

- [de Ronde *et al.*, 2012] W.H. de Ronde, F. Tostevin, and P.R. ten Wolde, “Feed-forward loops and diamond motifs lead to tunable transmission of information in the frequency domain,” *Phys Rev E*, 86:021913, 2012.
- [Desai and Fisher, 2007] M.M. Desai and D.S. Fisher, “Beneficial mutation selection balance and the effect of linkage on positive selection,” *Genetics*, 176(3):1759–98, 2007.
- [Dessaud *et al.*, 2008] E. Dessaud, A. P. McMahon, and J. Briscoe, “Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network,” *Development*, 135(15):2489–503, 2008.
- [Dobrzynski and Bruggeman, 2009] M. Dobrzynski and F.J. Bruggeman, “Elongation dynamics shape bursty transcription and translation,” *Proc Natl Acad Sci USA*, 106(8):2583–8, 2009.
- [Donaldson-Matasci *et al.*, 2010] M.C. Donaldson-Matasci, C.T. Bergstrom, and M. Lachmann, “The fitness value of information,” *Oikos*, 219(2):219–30, 2010.
- [Dowell, 2010] R.D. Dowell, “Transcription factor binding variation in the evolution of gene regulation,” *Trends Genet*, 26(11):468–75, 2010.
- [Dubuis *et al.*, 2013] J.O. Dubuis, G. Tkačik, E.F. Wieschaus, T. Gregor, and W. Bialek, “Positional information, in bits,” *Proc Natl Acad Sci USA*, 110:16301–8, 2013.
- [Earnest *et al.*, 2013] T. M. Earnest, E. Roberts, M. Assaf, K. Dahmen, and Z. Luthey-Schulten, “DNA looping increases the range of bistability in a stochastic model of the lac genetic switch,” *Phys Biol*, 10(2):026002, 2013.
- [Egloff *et al.*, 2012] S. Egloff, M. Dienstbier, and S. Murphy, “Updating the RNA polymerase CTD code: adding gene-specific layers,” *Trends Genet*, 28(7):333–41, 2012.
- [Egloff and Murphy, 2008] S. Egloff and S. Murphy, “Cracking the RNA polymerase II CTD code,” *Trends Genet*, 24:280–8, 2008.
- [El Meouche *et al.*, 2016] I. El Meouche, Y. Siu, and M.J. Dunlop, “Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells,” *Scientific Reports*, 6:1–9, 2016.
- [Eldar *et al.*, 2009] A. Eldar, V.K. Chary, P. Xenopoulos, M.E. Fontes, O.C. Loson, J. Dworkin, Piggot P.J., and M.B. Elowitz, “Partial penetrance facilitates developmental evolution in bacteria,” *Nature*, 460(7254):510–4, 2009.

- [Eldar and Elowitz, 2010] A. Eldar and M.B. Elowitz, “Functional roles for noise in genetic circuits,” *Nature*, 467(7312):167–73, 2010.
- [Elowitz *et al.*, 2002] M.B. Elowitz, A.J. Levine, E.D. Siggia, and P.S. Swain, “Stochastic gene expression in a single cell,” *Science*, 297(5584):1183–6, 2002.
- [Endres and Wingreen, 2008] R.G. Endres and N.S. Wingreen, “Accuracy of direct gradient sensing by single cells,” *Proc Natl Acad Sci USA*, 105(41):15749–54, 2008.
- [Fakhouri *et al.*, 2010] W.D. Fakhouri, A. Ay, R. Sayal, J. Dresch, E. Dayringer, and D.N. Arnosti, “Deciphering a transcriptional regulatory code: modeling short-range repression in the *Drosophila* embryo,” *Mol Syst Biol*, 6:341, 2010.
- [Fay and Wittkopp, 2007] J.C. Fay and P.J. Wittkopp, “Evaluating the role of natural selection in the evolution of gene regulation,” *Heredity*, 100(X):191–9, 2007.
- [Friedlander *et al.*, 2015] T. Friedlander, R. Prizak, C.G. Guet, N.H. Barton, and Tkačik. G., “Intrinsic limits to gene regulation by global crosstalk,” *arXiv.org*, page 1506.06925, 2015.
- [Garcia and Phillips, 2011] H.G. Garcia and R. Phillips, “Quantitative dissection of the simple repression input-output function,” *Proc Natl Acad Sci USA*, 108(29):12173–8, 2011.
- [Garcia *et al.*, 2012] H.G. Garcia, A. Sanchez, J.Q. Boedicker, M. Osborne, J. Gelles, J. Kondev, and R. Phillips, “Operator sequence alters gene expression independently of transcription factor occupancy in bacteria,” *Cell Rep*, 2(1):150–61, 2012.
- [Geertz *et al.*, 2012] M. Geertz, D. Shore, and S.J. Maerkl, “Massively parallel measurements of molecular interaction kinetics on a microfluidic platform,” *Proc Natl Acad Sci USA*, 109(41):16540–5, 2012.
- [Gerland and Hwa, 2002] U. Gerland and T. Hwa, “On the selection and evolution of regulatory DNA motifs,” *J Mol Evol*, 55(4):386–400, 2002.
- [Gerland *et al.*, 2002] U. Gerland, J.D. Moroz, and T. Hwa, “Physical constraints and functional characteristics of transcription factor–DNA interaction,” *Proc Natl Acad Sci USA*, 99:12015–20, 2002.
- [Gillespie, 2000] D.T. Gillespie, “The chemical Langevin equation,” *J Chem Phys*, 113(1):297, 2000.

- [Giorgetti, 2010] L. Giorgetti, “Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs,” *Mol Cell*, 37:418–28, 2010.
- [Golding *et al.*, 2005] I. Golding, J. Paulsson, S.M. Zawilski, and E.C. Cox, “Real-time kinetics of gene activity in individual bacteria,” *Cell*, 123(6):1025–36, 2005.
- [Gregor *et al.*, 2007] T. Gregor, D.W. Tank, E.F. Wieschaus, and W. Bialek, “Probing the Limits to Positional Information,” *Cell*, 130(1):153–64, 2007.
- [Grönlund *et al.*, 2013] A. Grönlund, P. Lötstedt, and J. Elf, “Transcription factor binding kinetics constrain noise suppression via negative feedback,” *Nat Commun*, 4:1864, 2013.
- [Gutierrez *et al.*, 2009] P.S. Gutierrez, D. Monteoliva, and L. Diambra, “Role of cooperative binding on noise expression,” *Phys Rev E*, 80:011914, 2009.
- [Gutierrez *et al.*, 2012] P.S. Gutierrez, D. Monteoliva, and L. Diambra, “Cooperative binding of transcription factors promotes bimodal gene expression response,” *PLoS ONE*, 7(9):e44812, 2012.
- [Hansen and O’Shea, 2015] A.S. Hansen and E.K. O’Shea, “Limits on information transduction through amplitude and frequency regulation of transcription factor activity,” *Elife*, 4:e06559, 2015.
- [He *et al.*, 2010] X. He, M.A.H. Samee, C. Blatti, and S. Sinha, “Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression,” *PLOS Comput Biol*, 6:e1000935, 2010.
- [Hersh *et al.*, 1996] B.M. Hersh, F.T. Farooq, D.N. Barstad, D.L. Blankenhorn, and J.L. Slonczewski, “A glutamate-dependent acid resistance gene in *Escherichia coli*,” *J Bacteriol*, 178:3978–81, 1996.
- [Hoekstra and Coyne, 2007] H.E. Hoekstra and J.A. Coyne, “The locus of evolution: evo devo and the genetics of adaptation,” *Evolution*, 61(5):995–1016, 2007.
- [Hopfield, 1974] J.J. Hopfield, “Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity,” *Proc Natl Acad Sci USA*, 71:4135–9, 1974.
- [Hopfield *et al.*, 1976] J.J. Hopfield, T. Yamane, V. Yue, and S.M. Coutts, “Direct experimental evidence for kinetic proofreading in amino acylation of tRNA-Ile,” *Proc Natl Acad Sci USA*, 73(4):1164–8, 1976.

- [Hormoz, 2013] S. Hormoz, “Cross talk and interference enhance information capacity of a signaling pathway,” *Biophys J*, 104(5):1170–80, 2013.
- [Houchmandzadeh *et al.*, 2002] B. Houchmandzadeh, E. Wieschaus, and S. Leibler, “Establishment of developmental precision and proportions in the early *Drosophila* embryo,” *Nature*, 415:798–802, 2002.
- [Hu *et al.*, 2011] B. Hu, D.A. Kessler, W.-J. Rappel, and H. Levine, “Effects of Input Noise on a Simple Biochemical Switch,” *Phys Rev Lett*, 107:148101, 2011.
- [Hu *et al.*, 2012] B. Hu, D.A. Kessler, W.-J. Rappel, and H. Levine, “How input fluctuations reshape the dynamics of a biological switching system,” *Phys Rev E*, 86:061910, 2012.
- [Ito *et al.*, 2009] Y. Ito, H. Toyota, K. Kaneko, and T. Yomo, “How selection affects phenotypic fluctuation,” *Mol Syst Biol*, 5:264, 2009.
- [Iwasa, 1988] Y. Iwasa, “Free fitness that always increases in evolution.,” *J Theor Biol*, 135(3):265–81, 1988.
- [Iyer-Biswas *et al.*, 2009] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, “Stochasticity of gene products from transcriptional pulsing,” *Phys Rev E*, 79(3):031911, 2009.
- [Jacob and Monod, 1961] F. Jacob and J. Monod, “Genetic regulatory mechanisms in the synthesis of proteins,” *J Mol Biol*, 3(3):318–56, 1961.
- [Janssens *et al.*, 2006] H. Janssens, S. Hous, J. Jaeger, A.R. Kim, E. Myasnikova, D. Sharp, and J. Reinitz, “Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster* even skipped gene,” *Nat Genet*, 38:1159–65, 2006.
- [Jenkins *et al.*, 1988] D.E. Jenkins, J.E. Schultz, and A. Matin, “Starvation-induced cross protection against heat or H₂O₂ challenge in *Escherichia coli*,” *J Bacteriol*, 170:3910–4, 1988.
- [Johnson *et al.*, 2005] J.M. Johnson, S. Edwards, D. Shoemaker, and E.E. Schadt, “Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments,” *Trends Genet*, 21(2):93–102, 2005.
- [Jones *et al.*, 2014] D.L. Jones, R.C. Brewster, and R. Phillips, “Promoter architecture dictates cell-to-cell variability in gene expression,” *Science*, 346(6216):1533–6, 2014.
- [Jost *et al.*, 2013] D. Jost, A. Nowojewski, and E. Levine, “Regulating the many to benefit the few: Role of weak small RNA targets,” *Biophys J*, 104(8):1773–82, 2013.

- [Jothi *et al.*, 2009] R. Jothi, S. Balaji, A. Wuster, J.A. Grochow, J. Gsponer, T.M. Przytycka, L. Aravind, and M.M. Babu, “Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture,” *Mol Syst Biol*, 5:294, 2009.
- [Kandhavelu *et al.*, 2011] M. Kandhavelu, H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-Price, O. Yli-Harja, and A.S. Ribeiro, “In vivo kinetics of transcription initiation of the *lar* promoter in *Escherichia coli*. Evidence for a sequential mechanism with two rate-limiting steps,” *BMC Syst Biol*, 5:149, 2011.
- [Karmakar, 2010] R. Karmakar, “Conversion of graded to binary response in an activator-repressor system,” *Phys Rev E*, 81:021905, 2010.
- [Kepler and Elston, 2001] T.B. Kepler and T.C. Elston, “Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations,” *Biophys J*, 81(6):3116–36, 2001.
- [Kimura, 1962] M. Kimura, “On the probability of fixation of mutant genes in a population,” *Genetics*, 47:713–9, 1962.
- [Kinney *et al.*, 2010] J.B. Kinney, A. Murugan, C.G. Callan, and E.C. Cox, “Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence,” *Proc Natl Acad Sci USA*, 107(20):9158–63, 2010.
- [Koh and Dunlop, 2012] R.S. Koh and M.J. Dunlop, “Modeling suggests that gene circuit architecture controls phenotypic variability in a bacterial persistence network,” *BMC Syst Biol*, 6:47, 2012.
- [Kuhlman *et al.*, 2007a] T. Kuhlman, Z. Zhang, M.H. Saier, and T. Hwa, “Combinatorial transcriptional control of the lactose operon of *Escherichia coli*,” *Proc Natl Acad Sci USA*, 104(14):6043–8, 2007.
- [Kuhlman *et al.*, 2007b] T. Kuhlman, Z. Zhang, M.H. Jr Saier, and T. Hwa, “Combinatorial transcriptional control of the lactose operon of *Escherichia coli*,” *Proc Natl Acad Sci USA*, 104:6043–8, 2007.
- [Kuhlman and Cox, 2010] T.E. Kuhlman and E.C. Cox, “Site Specific Chromosomal Integration of Large Synthetic Constructs,” *Nucleic Acids Research*, 38(6):e92, 2010.
- [Kussell and Leibler, 2005] E. Kussell and S. Leibler, “Phenotypic diversity, population growth, and information in fluctuating environments,” *Science*, 309(5743):2075–8, 2005.

- [Kwon *et al.*, 2010] Y.K. Kwon, M.B. Higgins, and J.D. Rabinowitz, "Antifolate-Induced Depletion of Intracellular Glycine and Purines Inhibits Thymineless Death in *E. coli*," *ACS Chem Biol*, 5:787–95, 2010.
- [Larson, 2011] D.R. Larson, "What do expression dynamics tell us about the mechanism of transcription?," *Curr Opin Genet Dev*, 21(5):591–9, 2011.
- [Lehner, 2008] B. Lehner, "Selection to minimise noise in living systems and its implications for the evolution of gene expression," *Mol Syst Biol*, 4:170, 2008.
- [Lehner, 2010] B. Lehner, "Conflict between noise and plasticity in yeast," *PLoS Genet*, 6(11):e1001185, 2010.
- [Lenormand *et al.*, 2009] T. Lenormand, D. Roze, and F. Rousset, "Stochasticity in evolution," *X*, 24:157–65, 2009.
- [Levine *et al.*, 2007] J. Levine, H.Y. Kueh, and L. Mirny, "Intrinsic fluctuations, robustness, and tunability in signaling cycles," *Biophys J*, 92(12):4473–81, 2007.
- [Levine and Davidson, 2005] M. Levine and E.H. Davidson, "Gene regulatory networks for development," *Proc Natl Acad Sci USA*, 102(14):4936–42, 2005.
- [Levy *et al.*, 2015] S.F. Levy, J.R. Blundell, S Venkataram, D.A. Petrov, D.S. Fisher, and G. Sherlock, "Quantitative evolutionary dynamics using high-resolution lineage tracking," *Nature*, 519(7542):181–6, 2015.
- [Leyer and Johnson, 1993] G.J. Leyer and E.A. Johnson, "Acid adaptation induces cross-protection against environmental stresses in *Salmonella typhimurium*," *Appl Environ Microbiol*, 59:1842–7, 1993.
- [Li *et al.*, 2009] G.-W. Li, O.G. Berg, and J. Elf, "Effects of macromolecular crowding and DNA looping on gene regulation kinetics," *Nat Phys*, 5:294–7, 2009.
- [Li *et al.*, 2008] X.Y. Li, S. MacArthur, R. Bourgon, D. Nix, D.A. Pollard, V.N. Iyer, A. Hechmer, L. Simirenko, M. Stapleton, C.L. Luengo-Hendriks, H.C. Chu, N. Ogawa, W. Inwood, V. Sementchenko, A. Beaton, R. Weiszmann, S.E. Celniker, D.W. Knowles, T. Gingeras, T.P. Speed, M.B. Eisen, and M.D. Biggin, "Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm," *PLOS Biol*, 6:e27, 2008.
- [Lin-Chao and Bremer, 1986] S. Lin-Chao and H. Bremer, "Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*," *Mol Gen Genet*, 203:143–9, 1986.

- [Little *et al.*, 2013] S.C. Little, M. Tikhonov, and T. Gregor, “Precise developmental gene expression arises from globally stochastic transcriptional activity,” *Cell*, 154:789–800, 2013.
- [Liu *et al.*, 2011] X. Liu, D.A. Bushnell, D.A. Silva, X. Huang, and R.D. Kornberg, “Initiation complex structure and promoter proofreading,” *Science*, 333(6042):633–7, 2011.
- [Lutz and Bujard, 1997] R. Lutz and H. Bujard, “Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements,” *Nucleic Acids Res*, 25(6):1203–10, 1997.
- [MacKay, 2003] David J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- [Maerkl and Quake, 2007] S.J. Maerkl and S.R. Quake, “A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors,” *Science*, 315:233–7, 2007.
- [Mancini *et al.*, 2013] F. Mancini, C. H. Wiggins, M. Marsili, and A. M. Walczak, “Time-dependent information transmission in a model regulatory circuit,” *Phys Rev E*, 88:022708, 2013.
- [McGinnis and Krumlauf, 1992] W. McGinnis and R. Krumlauf, “Homeobox genes and axial patterning,” *Cell*, 68:283–302, 1992.
- [McMahon *et al.*, 2007] M.A.S. McMahon, J. Xu, J.E. Moore, I.S. Blair, and D.A. McDowell, “Environmental Stress and Antibiotic Resistance in Food-Related Pathogens,” *Appl Environ Microbiol*, 73:211–7, 2007.
- [Metzger *et al.*, 2015] B.P. Metzger, D.C. Yuan, J.D. Gruber, F. Duvéau, and P.J. Wittkopp, “Selection on noise constrains variation in a eukaryotic promoter,” *Nature*, 521:344–7, 2015.
- [Milo *et al.*, 2010] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer, “BioNumbers – the database of key numbers in molecular and cell biology,” *Nucl Acids Res*, 38:D750–3, 2010.
- [Mirny, 2010] L. Mirny, “Nucleosome-mediated cooperativity between transcription factors,” *Proc Natl Acad Sci USA*, 107(52):22534–9, 2010.
- [model for the statistical fluctuations of protein numbers in a microbial population, 1978]
A model for the statistical fluctuations of protein numbers in a microbial population, “Berg, O.G.,” *J Theor Biol*, 71(4):587–603, 1978.

- [Mugler *et al.*, 2013] A. Mugler, F. Tostevin, and P.R. ten Wolde, “Spatial partitioning improves the reliability of biochemical signaling,” *Proc Natl Acad Sci USA*, 110(15):5927–32, 2013.
- [Müller and Basler, 2000] B. Müller and K. Basler, “The repressor and activator forms of *Cu-bitus interruptus* control Hedgehog target genes through common generic Gli-binding sites,” *Development*, 127(14):2999–3007, 2000.
- [Müller *et al.*, 2013] J. Müller, B.A. Hense, T.M. Fuchs, M. Utz, and C. Pötzsche, “Bet-hedging in stochastically switching environments.,” *J Theor Biol*, 336:144–57, 2013.
- [Neuert *et al.*, 2013] G. Neuert, B. Munsky, R.Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, “Systematic identification of signal-activated stochastic gene regulation,” *Science*, 339(6119):584–7, 2013.
- [Newman *et al.*, 2006] J.R.S. Newman, S. Ghaemmaghami, J. Ihmels, D.K. Breslow, M. Noble, J.L. DeRisi, and J.S. Weissman, “Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise,” *Nature*, 441:840–6, 2006.
- [Ni *et al.*, 2012] M. Ni, A.L. Decrulle, F. Fontaine, A. Demarez, F. Taddei, and A.B. Lindner, “Pre-disposition and epigenetics govern variation in bacterial survival upon stress,” *PLoS Genetics*, 8:e1003148, 2012.
- [Nock *et al.*, 2012] A. Nock, J.M. Ascano, M.J. Barrero, and S. Malik, “Mediator-regulated transcription through the +1 nucleosome,” *Mol Cell*, 48(6):837–48, 2012.
- [Ozbudak *et al.*, 2002] E.M. Ozbudak, M. Thattai, I. Kurtser, A.D. Grossman, and A. van Oudenaarden, “Regulation of noise in the expression of a single gene,” *Nat Genet*, 31(1):69–73, 2002.
- [Paixao and Bauer, 2015] T. Paixao and U. Bauer, “Using Reeb graphs to investigate fitness landscapes,” Technical report, IST Austria, 2015.
- [Parker *et al.*, 2011] D.S. Parker, M.A. White, A.I. Ramos, B.A. Cohen, and S. Barolo, “The cis-regulatory logic of Hedgehog gradient responses: Key roles for Gli binding affinity, competition, and cooperativity,” *Sci Signal*, 4(176):ra38, 2011.
- [Paulsson, 2004] J. Paulsson, “Summing up the noise in gene networks,” *Nature*, 427:415–8, 2004.
- [Paulsson and Ehrenberg, 2001] J. Paulsson and M. Ehrenberg, “Noise in a minimal regulatory network: plasmid copy number control,” *Q Rev Biophys*, 34:1–59, 2001.

- [Peccoud and Ycart, 1995] J. Peccoud and B. Ycart, “Markovian modelling of gene products synthesis,” *Theor Pop Biol*, 48:222–34, 1995.
- [Pedraza and Paulsson, 2008] J.M. Pedraza and J. Paulsson, “Effects of molecular memory and bursting on fluctuations in gene expression,” *Science*, 319(5861):339–43, 2008.
- [Pedraza and van Oudenaarden, 2005] J.M. Pedraza and A. van Oudenaarden, “Noise propagation in gene networks,” *Science*, 307(5717):1965–9, 2005.
- [Prelich, 2002] G. Prelich, “RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function,” *Eukaryot Cell*, 1(2):153–62, 2002.
- [Ptashne and Gann, 2002] M. Ptashne and A. Gann, *Genes and Signals*, Cold Spring Harbor Press, New York, 2002.
- [Raj *et al.*, 2006] A. Raj, C.S. Peskin, D. Tranchina, D.Y. Vargas, and S. Tyagi, “Stochastic mRNA Synthesis in Mammalian Cells,” *PLoS Biol*, 4(10):e309, 2006.
- [Raj *et al.*, 2010] A. Raj, S.A. Rifkin, E. Andersen, and A. van Oudenaarden, “Variability in gene expression underlies incomplete penetrance,” *Nature*, 463(7283):913–8, 2010.
- [Raj and van Oudenaarden, 2008] A. Raj and A. van Oudenaarden, “Nature, nurture, or chance: Stochastic gene expression and its consequences,” *Cell*, 135(2):216–26, 2008.
- [Raser and O’Shea, 2004] J.M. Raser and E.K. O’Shea, “Control of Stochasticity in Eukaryotic Gene Expression,” *Science*, 304(5678):1811–4, 2004.
- [Raser and O’Shea, 2005] J.M. Raser and E.K. O’Shea, “Noise in gene expression: origins, consequences, and control,” *Science*, 309(5743):2010–3, 2005.
- [Richard and Foster, 2004] H. Richard and J.W. Foster, “Escherichia coli Glutamate- and Arginine-Dependent Acid Resistance Systems Increase Internal pH and Reverse Transmembrane Potential,” *J Bacteriol*, 186:6032–41, 2004.
- [Rieckh and Tkačik, 2014] G. Rieckh and G. Tkačik, “Noise and information transmission in promoters with multiple internal States,” *Biophys J*, 106(5):1194–204, 2014.
- [Rieke *et al.*, 1997] F. Rieke, D. Warland, R. Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*, The MIT press, 1997.
- [Rivoire and Leibler, 2011] O. Rivoire and S. Leibler, “The value of information for populations in varying environments,” *J Stat Phys*, 142(6):1124–66, 2011.

- [Rockel *et al.*, 2013] S. Rockel, M. Geertz, K. Hens, B. Deplancke, and S.J. Maerkl, “iSLIM: a comprehensive approach to mapping and characterizing gene regulatory networks,” *Nucl Acids Res*, 41:e52, 2013.
- [Romero *et al.*, 2012] I.G. Romero, I. Ruvinsky, and Y. Gilad, “Comparative studies of gene expression and the evolution of gene regulation,” *Nat Rev Genet*, 13(7):505–16, 2012.
- [Ronen *et al.*, 2002] M. Ronen, R. Rosenberg, B.I. Shraiman, and U. Alon, “Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics,” *Proc Natl Acad Sci USA*, 99(16):10555–60, 2002.
- [Rosenfeld *et al.*, 2005] N. Rosenfeld, J.W. Young, U. Alon, P.S. Swain, and M.B. Elowitz, “Gene regulation at the single-cell level,” *Science*, 307(5717):1962–5, 2005.
- [Rotem *et al.*, 2010] E. Rotem, A. Loinger, I. Ronin, I. Levin-Reisman, C. Gabay, N. Shoresh, O. Biham, and N.Q. Balaban, “Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence,” *Proc Natl Acad Sci USA*, 107(28):12541–6, 2010.
- [Ruess *et al.*, 2015] J. Ruess, F. Parise, A. Miliadis-Argeitis, M. Khammash, and J. Lygeros, “Iterative experiment design guides the characterization of a light-inducible gene expression circuit,” *Proc Natl Acad Sci USA*, 112(26):8148–53, 2015.
- [Sakata *et al.*, 2009] A. Sakata, K. Hukushima, and K. Kaneko, “Funnel landscape and mutational robustness as a result of evolution under thermal noise,” *Phys Rev Lett*, 102(14):148101, 2009.
- [Sanchez *et al.*, 2013] A. Sanchez, S. Choubey, and J. Kondev, “Regulation of noise in gene expression,” *Annu Rev Biophys*, 42:469–91, 2013.
- [Sanchez *et al.*, 2011a] A. Sanchez, H.G. Garcia, D. Jones, R. Phillips, and J. Kondev, “Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression,” *PLoS Comput Biol*, 7(3):e1001100, 2011.
- [Sanchez *et al.*, 2011b] A. Sanchez, H.G. Garcia, D. Jones, R. Phillips, and J. Kondev, “Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression,” *PLoS Comput Biol*, 7(3):e1001100, 2011.
- [Sanchez and Kondev, 2008] A. Sanchez and J. Kondev, “Transcriptional control of noise in gene expression,” *Proc Natl Acad Sci USA*, 105(13):5081–6, 2008.

- [Sanchez *et al.*, 2011c] A. Sanchez, M.L. Osborne, L.J. Friedman, J. Kondev, and J. Gelles, “Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules,” *EMBO J*, 30(19):3940–6, 2011.
- [Sandelin *et al.*, 2004] A. Sandelin, W. Alkema, P. Engström, W.W. Wasserman, and B. Lenhard, “JASPAR: an open-access database for eukaryotic transcription factor binding profiles,” *Nucl Acids Res*, 32:D91–4, 2004.
- [Sangurdekar *et al.*, 2011] D.P. Sangurdekar, Z. Zhang, and A.B. Khodursky, “The association of DNA damage response and nucleotide level modulation with the antibacterial mechanism of the anti-folate drug Trimethoprim,” *BMC Genomics*, 12:583, 2011.
- [Sato *et al.*, 2003] K. Sato, Y. Ito, T. Yomo, and K. Kaneko, “On the relation between fluctuation and response in biological systems,” *Proc Natl Acad Sci USA*, 100(24):14086–90, 2003.
- [Savir and Tlusty, 2013] Y. Savir and T. Tlusty, “The ribosome as an optimal decoder: a lesson in molecular recognition,” *Cell*, 153:471–9, 2013.
- [Segal and Widom, 2009] E. Segal and J. Widom, “What controls nucleosome positions,” *Trends Genet*, 25(8):335–42, 2009.
- [Sella and Hirsh, 2005] G. Sella and A.E. Hirsh, “The application of statistical physics to evolutionary biology,” *Proc Natl Acad Sci USA*, 102(27):9541–6, 2005.
- [Setty *et al.*, 2003] Y. Setty, A. E. Mayo, , M. G. Surette, and U. Alon, “Detailed map of a cis-regulatory input function,” *Proc Natl Acad Sci USA*, 100(13):7702–7, 2003.
- [Shahrezaei and Swain, 2008] V. Shahrezaei and P.S. Swain, “Analytical distributions for stochastic gene expression,” *Proc Natl Acad Sci USA*, 105(45):17256–61, 2008.
- [Shannon, 1948] C.E. Shannon, “A mathematical theory of communication,” *Bell Syst Tech J*, 27:379–423&623–56, 1948.
- [Shannon and Weaver, 1949] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, U Illinois Press, 1949.
- [Shea and Ackers, 1985a] M.A. Shea and G.K. Ackers, “The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation,” *J Mol Biol*, 181:211–30, 1985.

- [Shea and Ackers, 1985b] M.A. Shea and G.K. Ackers, "The O_R control system of bacteriophage lambda. A physical-chemical model for gene regulation," *J Mol Biol*, 181(2):211–30, 1985.
- [Sherman and Cohen, 2014] M.S. Sherman and B.A. Cohen, "A computational framework for analyzing stochasticity in gene expression," *PLoS Comput Biol*, 10(5):e1003596, 2014.
- [Silander *et al.*, 2012] O.K. Silander, N. Nikolic, A. Zaslaver, A. Bren, I. Kikoin, U. Alon, and M. Ackermann, "A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in *Escherichia coli*," *PLoS Genetics*, 8:e1002443, 2012.
- [Singh *et al.*, 2012] A. Singh, B.S. Razoogy, R.D. Dar, and L.S. Weinberger, "Dynamics of protein noise can distinguish between alternate sources of gene-expression variability," *Mol Syst Biol*, 8:607, 2012.
- [So *et al.*, 2011] L.-H. So, A. Ghosh, C. Zong, L.A. Sepulveda, R. Segev, and I. Golding, "General properties of transcriptional time series in *Escherichia coli*," *Nat Gen*, 43:554–60, 2011.
- [Sokolowski and Tkačik, 2015] T.R. Sokolowski and G. Tkačik, "Optimizing information flow in small genetic networks. IV. Spatial coupling.," *Phys Rev E*, 91(6):062710, 2015.
- [Stiller and Hall, 2002] J.W. Stiller and B.D. Hall, "Evolution of the RNA polymerase II C-terminal domain," *Proc Natl Acad Sci USA*, 99(9):6091–6, 2002.
- [Stormo and Fields, 1998] G.D. Stormo and D.S. Fields, "Specificity, free energy and information content in protein-DNA interactions," *Trends Biochem Sci*, 23(3):109–13, 1998.
- [Stormo and Zhao, 2010] G.D. Stormo and Y. Zhao, "Determining the specificity of protein-DNA interactions," *Nat Rev Genet*, 11(11):751–60, 2010.
- [Suter *et al.*, 2011] D.M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, "Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics," *Science*, 332(6028):472–4, 2011.
- [Swain *et al.*, 2002] P.S. Swain, M.B. Elowitz, and E.D. Siggia, "Intrinsic and extrinsic contributions to stochasticity in gene expression," *Proc Natl Acad Sci USA*, 99(20):12795–800, 2002.
- [Taft *et al.*, 2007] R.J. Taft, M. Pheasant, and J.S. Mattick, "The relationship between non-protein-coding DNA and eukaryotic complexity," *Bioessays*, 29(3):288–99, 2007.

- [Tal and Paulsson, 2012] S. Tal and J. Paulsson, “Evaluating quantitative methods for measuring plasmid copy numbers in single cells,” *Plasmid*, 67:167–73, 2012.
- [Taylor *et al.*, 2007] S.F. Taylor, N. Tishby, and W. Bialek, “Information and fitness,” *arXiv*, 0712.4382, 2007.
- [Tkačik and Bialek, 2016] G. Tkačik and W. Bialek, “Information processing in biological systems,” *Annu Rev Cond Matt Phys*, 7(12):89–117, 2016.
- [Tkačik *et al.*, 2008a] G. Tkačik, C.G. Callan, and W. Bialek, “Information capacity of genetic regulatory elements,” *Phys Rev E*, 78:011910, 2008.
- [Tkačik *et al.*, 2008b] G. Tkačik, C.G. Callan, and W. Bialek, “Information flow and optimization in transcriptional regulation,” *Proc Natl Acad Sci USA*, 105(34):12265–70, 2008.
- [Tkačik *et al.*, 2008c] G. Tkačik, T. Gregor, and W. Bialek, “The Role of Input Noise in Transcriptional Regulation,” *PLoS ONE*, 3(7):e2774, 2008.
- [Tkačik and Walczak, 2011] G. Tkačik and A.M. Walczak, “Information transmission in genetic regulatory networks: a review,” *J Phys: Condens Matter*, 23:153102, 2011.
- [Tkačik *et al.*, 2009] G. Tkačik, A.M. Walczak, and W. Bialek, “Optimizing information flow in small genetic networks,” *Phys Rev E*, 80:031920, 2009.
- [Tkačik *et al.*, 2012] G. Tkačik, A.M. Walczak, and W. Bialek, “Optimizing information flow in small genetic networks. III. A self-interacting gene,” *Phys Rev E*, 85:041903, 2012.
- [Todeschini *et al.*, 2014] A.L. Todeschini, A. Georges, and R.A. Veitia, “Transcription factors: specific DNA binding and specific gene regulation,” *Trends Genet*, 30:211–9, 2014.
- [Tostevin *et al.*, 2012] F. Tostevin, W.H. de Ronde, and P.R. ten Wolde, “Reliability of frequency and amplitude decoding in gene regulation,” *Phys Rev Lett*, 108:108104, 2012.
- [Tostevin and ten Wolde, 2009] F. Tostevin and P.R. ten Wolde, “Mutual Information between Input and Output Trajectories of Biochemical Networks,” *Phys Rev Lett*, 102:218101, 2009.
- [Tostevin and ten Wolde, 2010] F. Tostevin and P.R. ten Wolde, “Mutual information in time-varying biochemical systems,” *Phys Rev E*, 81:061917, 2010.
- [Tsai *et al.*, 2013] M.-F. Tsai, P. McCarthy, and C. Miller, “Substrate selectivity in glutamate-dependent acid resistance in enteric bacteria,” *PNAS*, 110:5898–902, 2013.
- [Tsimring, 2014] L.S. Tsimring, “Noise in biology,” *Rep Prog Phys*, 77(2):026601, 2014.

- [Tuğrul *et al.*, 2015] M. Tuğrul, T. Paixão, N.H. Barton, and G. Tkačik, “Dynamics of transcription binding site evolution,” *PLoS Genetics*, 11(11):e1005639, 2015.
- [van Kampen, 1981] N.G. van Kampen, *Stochastic processes in physics and chemistry*, North Holland, Amsterdam, 1981.
- [van Nimwegen, 2003] E. van Nimwegen, “Scaling laws in the functional content of genomes,” *Trends Genet*, 19(9):479–84, 2003.
- [Venkatesh and Workman, 2015] S. Venkatesh and J.L. Workman, “Histone exchange, chromatin structure and the regulation of transcription,” *Nat Rev Mol Cell Biol*, 16(3):178–89, 2015.
- [Villar *et al.*, 2014] D. Villar, P. Flicek, and D.T. Odom, “Evolution of transcription factor binding in metazoans – mechanisms and functional implications,” *Nat Rev Genet*, 15(4):221–33, 2014.
- [Vinuelas *et al.*, 2013] J. Vinuelas, G. Kaneko, A. Coulon, E. Vallin, V. Morin, C. Mejia-Pous, J.-J. Kupiec, G. Beslon, and O. Gandrillon, “Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts,” *BMC Biol*, 11:15, 2013.
- [von Hippel and Berg, 1986] P.H. von Hippel and O.G. Berg, “On the specificity of DNA-protein interactions,” *Proc Natl Acad Sci USA*, 83(6):1608–12, 1986.
- [von Hippel *et al.*, 1974] P.H. von Hippel, A. Revzin, C.A. Gross, and A.C. Wang, “Non-specific DNA binding of genome regulating proteins as a biological control mechanism: 1. The lac operon: equilibrium aspects,” *Proc Natl Acad Sci USA*, 71(12):4808–12, 1974.
- [Walczak *et al.*, 2005] A.M. Walczak, M. Sasai, and P.G. Wolynes, “Self-consistent proteomic field theory of stochastic gene switches,” *Biophys J*, 88(2):828–50, 2005.
- [Walczak *et al.*, 2010] A.M. Walczak, G. Tkačik, and W. Bialek, “Optimizing information flow in small genetic networks. II. Feed-forward interactions,” *Phys Rev E*, 81:041905, 2010.
- [Wang and Zhang, 2011] Z. Wang and J. Zhang, “Impact of gene expression noise on organismal fitness and the efficacy of natural selection,” *Proc Natl Acad Sci USA*, 108(16):E67–76, 2011.
- [Weber *et al.*, 2014] C.M. Weber, S. Ramachandran, and S. Henikoff, “Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase,” *Mol Cell*, 53(5):819–30, 2014.

- [White *et al.*, 2012] M.A. White, D.S. Parker, S. Barolo, and B.A. Cohen, “A model of spatially restricted transcription in opposing gradients of activators and repressors,” *Mol Syst Biol*, 8:614, 2012.
- [Wolf *et al.*, 2015] L. Wolf, O.K. Silander, and E. van Nimwegen, “Expression noise facilitates the evolution of gene regulation,” *eLife*, 4:e05856, 2015.
- [Wong Ng *et al.*, 2010] J. Wong Ng, D. Chatenay, J. Robert, and M.G. Poirier, “Plasmid copy number noise in monoclonal populations of bacteria,” *Phys Rev E*, 81:011909, 2010.
- [Wunderlich and Mirny, 2009] Z. Wunderlich and L.A. Mirny, “Different gene regulation strategies revealed by analysis of binding motifs,” *Trends Genet*, 25(10):434–40, 2009.
- [Yang and Ko, 2012] H.-T. Yang and M.S.H. Ko, “Stochastic modeling for the expression of a gene regulated by competing transcription factors,” *PLoS ONE*, 7(3):e32376, 2012.
- [Yosef and Regev, 2011] N. Yosef and A. Regev, “Impulse Control: Temporal Dynamics in Gene Transcription,” *Cell*, 144(6):886–96, 2011.
- [Young *et al.*, 2012] J.W. Young, J.C.W. Locke, A. Altinok, N. Rosenfeld, T. Bacarian, P.S. Swain, E. Mjolsness, and M.B. Elowitz, “Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy,” *Nature Protocols*, 7:80–8, 2012.
- [Zaslaver *et al.*, 2006] A. Zaslaver, A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M.G. Surette, and U. Alon, “A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*,” *Nature Methods*, 3:623–8, 2006.
- [Zenklusen *et al.*, 2008] D. Zenklusen, D.R. Larson, and R.H. Singer, “Single-RNA counting reveals alternative modes of gene expression in yeast,” *Nat Struct Mol Biol*, 15(12):1263–71, 2008.
- [Zhang *et al.*, 2012] J. Zhang, L. Chen, and T. Zhou, “Analytical distribution and tunability of noise in a model of promoter progress,” *Biophys J*, 102(6):1247–57, 2012.
- [Zhang and Hill, 2005] X.S. Zhang and W.G. Hill, “Evolution of the environmental component of the phenotypic variance: stabilizing selection in changing environments and the cost of homogeneity,” *Evolution*, 59(6):1237–44, 2005.
- [Zheng *et al.*, 2011] W. Zheng, T.A. Gianoulis, K.J. Karczewski, H. Zhao, and M. Snyder, “Regulatory Variation Within and Between Species,” *Annu Rev Genomics Hum Genet*, 12(1):327–46, 2011.

- [Zhou and Zhang, 2012] T. Zhou and J. Zhang, “Analytical Results for a Multistate Gene Model,” *SIAM J Appl Math*, 72(3):789–818, 2012.
- [Zhuravel *et al.*, 2010] D. Zhuravel, D. Fraser, S. St-Pierre, L. Tepliakova, W.L. Pang, J. Hasty, and M. Kaern, “Phenotypic impact of regulatory noise in cellular stress-response pathways,” *Syst Synth Biol*, 4(2):105–16, 2010.
- [Ziv *et al.*, 2007] E. Ziv, I. Nemenman, and C.H. Wiggins, “Optimal Signal Processing in Small Stochastic Biochemical Networks,” *PLoS ONE*, 2(10):e1077, 2007.