

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

A quantitative analysis of genetic transcriptional and post-transcriptional regulation in Escherichia coli

Permalink

<https://escholarship.org/uc/item/07z035t7>

Author

Kuhlman, Thomas Edward

Publication Date

2007

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

A Quantitative Analysis of Genetic
Transcriptional and Post-Transcriptional
Regulation in *Escherichia coli*

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Physics

by

Thomas Edward Kuhlman

Committee in charge:

Professor Terence Hwa, Chair
Professor Jeff Hasty
Professor Herbert Levine
Professor William Loomis
Professor Doug E. Smith

2007

The dissertation of Thomas Edward Kuhlman is approved,
and it is acceptable in quality and form for publication on
microfilm:

Chair

University of California, San Diego

2007

DEDICATION

For Stella

TABLE OF CONTENTS

Signature Page.....	iii
Dedication.....	iv
Table of Contents.....	v
List of Figures.....	ix
List of Tables.....	xi
Acknowledgement.....	xii
Vita.....	xiv
Abstract.....	xv
Chapter 1. Combinatorial Transcriptional Control of the Lactose Operon of <i>E. coli</i>	1
1.1. Introduction.....	1
1.2. Repression by LacR.....	3
1.3. Activation by CRP.....	8
1.4. Combinatorial Control.....	13
1.5. Activation by cAMP-CRP is non-cooperative.....	16
1.6. PDE provides insulation to variations in cAMP.....	17
1.7. DNA looping increases the sensitivity of the IPTG response.....	18
1.8. Cooperativity between CRP and LacR enhances sensitivity to IPTG.....	21
1.9. DNA looping is assisted by CRP.....	27
1.10. Hyper-sensitivity of IPTG response resulted from multiple factors.....	28
1.11. The importance of being quantitative.....	29
1.12. Acknowledgement.....	31
1.13. References.....	32

Chapter 2. Combinatorial Transcriptional Control of the Lactose Operon of *E. coli*:

Experimental and Theoretical Methods.....	38
2.1. Plasmids and Strains.....	38
2.2. B-galactosidase assay.....	39
2.3. Promoter activity.....	40
2.4. Activation of the regulatory proteins.....	44
2.5. Thermodynamic modeling.....	49
2.6. Activation by cAMP-CRP.....	51
2.7. Simple repression by LacR.....	53
2.8. LacR-mediated DNA looping.....	54
2.9. Combinatorial control by LacR and cAMP-CRP.....	57
2.10. Comparison to alternative models.....	60
2.11. Data analysis.....	65
2.12. Acknowledgement.....	67
2.13. References.....	68
Chapter 3. Autoregulation as a Strategy in Synthetic Reporter Constructs.....	72
3.1. Introduction.....	72
3.2. Sensitivity and dynamic range of a system with constitutive expression of repressor.....	74
3.3. Sensitivity and dynamic range of a system with autoregulatory control....	78
3.4. Experimental implementation of autoregulatory control.....	81
3.5. Autoregulation decreases the effective sensitivity of P_L promoters.....	85
3.6. Fine tuning for enhanced performance.....	86

3.7. Autoregulation in nature.....	89
3.8. Materials and Methods: Plasmids.....	90
3.9. Medium, growth, and measurements.....	91
3.10. Data analysis.....	92
3.11. References.....	94
Chapter 4. Quantitative Characteristics of Gene Regulation by Small RNA.....	98
4.1. Introduction.....	98
4.2. Coupled degradation leads to the threshold-linear response.....	100
4.3. Quantitative experimental characterization of RyhB.....	106
4.4. Reporter target of RyhB exhibits threshold-linear expression.....	112
4.5. Irreversible binding suffices for threshold-linear response.....	114
4.6. Hierarchical cross-talk between sRNA targets.....	119
4.7. Properties of sRNA repression vs. protein repression.....	125
4.8. sRNA regulation is subject to dynamic control.....	127
4.9. sRNA may exhibit tight repression of fluctuations.....	129
4.10. Small RNA regulation may be highly sensitive.....	134
4.11. Materials and Methods: Strains & Plasmids.....	135
4.12. Medium, growth, measurements.....	139
4.13. Data analysis.....	140
4.14. Acknowledgement.....	143
4.15. References.....	144
Chapter 5. Noise and Silence in Small RNA Regulation.....	150
5.1. Introduction.....	150

5.2. Model for sRNA regulation.....	152
5.3. Mass-action model.....	153
5.4. Noise in target gene expression.....	156
5.5. Population diversity at the threshold – Model.....	157
5.6. Population diversity at the threshold – Experiment.....	164
5.7. Noise characteristics of sRNA vs. protein regulation.....	173
5.8. Materials and Methods: Strains & plasmids.....	177
5.9. Medium, growth, measurements.....	177
5.10. Acknowledgement.....	180
5.11. References.....	181

LIST OF FIGURES

Figure 1.1: Dependence of Plac activity on the inducers.....	6
Figure 1.2: <i>Plac</i> activity of <i>E. coli</i> MG1655 cells and $\Delta cyaA$ mutants grown with various sugar sources.....	11
Figure 1.3: Indirect effect of cAMP.....	12
Figure 1.4: cAMP dependence of IPTG response.....	15
Figure 1.5: Combinatorial control of Plac activity for strain TK310.....	15
Figure 1.6: IPTG response in the absence of DNA looping.....	22
Figure 1.7: Schematics of the <i>lac</i> promoter.....	22
Figure 1.8: Combinatorial control of Plac activity for strain TK310.....	25
Figure 1.9: Combinatorial control of Plac activity for strain TK250.....	26
Figure 2.1: Determination of the limit of β -Gal activity detection.....	43
Figure 2.2: Different states of LacR-IPTG binding.....	48
Figure 2.3: Comparison to alternative models.....	64
Figure 3.1: Architecture of constitutive and autoregulatory expression systems.....	77
Figure 3.2: Effect of autoregulation on reporter expression.....	80
Figure 3.3: Expression profiles of experimental constitutive and autoregulated expression systems.....	83
Figure 3.4: Population distribution of $P_{Ltet-O1}$ autoregulatory expression system as determined by flow cytometry.....	84
Figure 3.5: Sequences of $P_{Ltet-O1}$ and $P_{Llac-O1}$ compared to consensus sequences.....	88
Figure 4.1: Threshold-linear response of a target gene.....	104
Figure 4.2: Example of raw data used to compile Fig. 4.3a.....	109

Figure 4.3: Threshold-linear response of a reporter target of RyhB.....	110
Figure 4.4: Activity dependent repression of chromosomal target.....	113
Figure 4.5: Threshold linear response of a reporter target of RNA-OUT.....	117
Figure 4.6: Model for indirect interaction between targets of a small RNA	122
Figure 4.7: Crosstalk between different targets of a common sRNA.....	123
Figure 4.8: Comparison between sRNA and protein mediated repression.....	132
Figure 4.9: Mean and Fano factor of the protein level in the silenced state.....	133
Figure 5.1: Noise indices for target mRNA and protein.....	155
Figure 5.2: Diversity in gene expression near the transition.....	161
Figure 5.3: Random walk picture.....	162
Figure 5.4: Phase diagram.....	163
Figure 5.5: Bulk response of a target gene to small RNA regulation.....	166
Figure 5.6: Histograms of single cell fluorescence in response to small RNA regulation, obtained by flow cytometry.....	169
Figure 5.7: Single cell expression level of $P_{L-lac}-gfp$	170
Figure 5.8: Flow cytometry data for RyhB/crsodB-gfp and RyhB/crsodB*-gfp.....	171
Figure 5.9: Single cell expression levels of $P_{L-lac}-crsodB-gfp$ at different levels of ryhB expression.....	172
Figure 5.10: Noise properties in the silenced state.....	176

LIST OF TABLES

Table 1.1: List of strains and plasmids used in this study.....	5
Table 1.2: Hill parameters for the IPTG response.....	7
Table 1.3: Hill parameters for the cAMP response.....	7
Table 1.4: Parameters for the model with DNA looping.....	7
Table 4.1: Model parameters.....	105
Table 4.2: Best-fit parameters of the data in Fig. 4.3 <i>a</i>	111
Table 4.3: Best-fit parameters of the data in Fig. 4.5 <i>a</i>	118
Table 4.4: Bacterial strains and plasmids used in this study.....	138
Table 5.1: Strains & Plasmids.....	179

ACKNOWLEDGEMENT

I would like to acknowledge the help and support of the entire Hwa group, especially Dr. Terry Hwa for his patience and guidance throughout the past four years of research. I thank Dr. Erel Levine for many hours of theoretical tutelage and amusing conversations. I additionally owe a great deal of gratitude to Dr. Zhongge Zhang, without whose exceedingly patient guidance none of the experimental work described here would have ever been possible.

In addition, I am indebted to Milton H. Saier, Jr., Bob Schleif, Bill Loomis, Nicholas Buchler, Antoine Danchin, Hernan Garcia, Kathleen Matthews, Stefan Oehler, Rob Phillips, Agnes Ullman, Susan Gottesman, R.W. Simons, and N. Kleckner for their helpful and insightful comments on this work, as well as Natalie Ostroff and Jeff Hasty for advice and hospitality.

Chapter 1 contains material in Kuhlman T, Zhang Z, Saier MH, Jr, Hwa T. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc Natl Acad Sci* 2007; 104(14): 6043-6048, used with permission. The work presented in Chapters 1 and 4 were supported by the DOE (Grant No. DE-FG02-03-ER63691), the NSF (Grant No. MCB-0417721), and by the NIH (Grant No. GM070915). We also acknowledge support by the NSF PFC-sponsored Center for Theoretical Biophysics (Grants No. PHY-0216576 and PHY-0225630). The dissertation author was the primary investigator and author of this work.

Chapter 2 contains material from the supplementary information contained in Kuhlman T, Zhang Z, Saier MH, Jr, Hwa T. Combinatorial transcriptional control of the

lactose operon of *Escherichia coli*. *Proc Natl Acad Sci* 2007; 104(14): 6043-6048, used with permission. The dissertation author was the primary investigator and author of this work.

Chapter 4 contains material in Levine E, Kuhlman T, Zhang Z, Hwa T. Quantitative features of gene regulation by small RNA. *PLoS Biology*, submitted, used with permission. The work in this chapter was supported by the NSF PFC-sponsored Center for Theoretical Biophysics (Grants No. PHY-0216576 and PHY-0225630), by the DOE through grant DE-FG02-03ER63691, and by the NIH through grant GM070915. The dissertation author was the primary experimental investigator of this work.

Chapter 5 contains material in Levine E, Kuhlman T, Zhang Z, Huang M, Huang Y, Hwa T. Noise and Silence in small RNA regulation. In preparation, used with permission. The work presented in this chapter was supported by the NSF PFC-sponsored Center for Theoretical Biophysics (Grants No. PHY-0216576 and PHY-0225630). We also acknowledge additional support by NSF Grant MCB. The dissertation author was the primary experimental investigator of this work.

VITA

2000	Bachelor of Arts, Ball State University
2000-2003	Teaching Assistant, Department of Physics, University of California, San Diego
2002	Master of Science, University of California, San Diego
2003-2007	Research Assistant, University of California, San Diego
2007	Doctor of Philosophy, University of California, San Diego

Publications

Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Philips R (2005). Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev* 15(2): 125-135.

Kuhlman T, Zhang Z, Saier MH, Jr., Hwa T (2007). Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc Natl Acad Sci* 104(14): 6043-6048.

Levine E, Kuhlman T, Zhang Z, Hwa T (2007). Quantitative characteristics of gene regulation by small RNA. *PLoS Biology*, Submitted.

Levine E, Kuhlman T, Zhang Z, Huang M, Huang Y, Hwa T (2007). Noise and silence in small RNA regulation. In Preparation.

ABSTRACT OF THE DISSERTATION

A Quantitative Analysis of Genetic Transcriptional
and Post-Transcriptional Regulation in *Escherichia coli*

by

Thomas Edward Kuhlman

Doctor of Philosophy in Physics

University of California, San Diego, 2007

Professor Terence Hwa, Chair

It is the goal of systems biology to understand the behavior of the whole in terms of knowledge of the parts. This is hard to achieve in many cases due to the difficulty of characterizing the many constituents and their complex web of interactions involved in biological systems. In this dissertation we report attempts to confront this difficulty by applying quantitative methods to experimentally characterize and theoretically model the web of interactions involved in transcriptional and post-transcriptional regulation in *E. coli*.

In Chapters 1 and 2, *Combinatorial Transcriptional Control of the Lactose Operon of Escherichia coli*, we study the combinatorial control of the *lac* promoter by

the regulators LacR and CRP. We are able to account for the combinatorial control of the *lac* promoter quantitatively. Specifically, our analysis indicates that the sensitivity of the inducer response results from LacR-mediated DNA looping, which is significantly enhanced by CRP.

In Chapter 3, *Autoregulation as a Strategy in Synthetic Reporter Constructs*, we predict that autoregulation of a transcription factor used to control expression of a reporter gene can lead to significantly decreased sensitivity of the response of the reporter to inducer. We show that this is indeed the case by constructing and assaying the response of two autoregulated reporter systems using the transcription factors LacR and TetR.

In Chapter 4, *Quantitative Characteristics of Gene Regulation by Small RNA*, we study quantitatively two classes of bacterial small RNAs (sRNA) in *Escherichia coli*. We demonstrate that sRNA provide a novel mode of gene regulation with characteristics distinct from those of protein-mediated gene regulation. These include a threshold-linear response with a tuneable threshold and a built-in capability for hierarchical cross talk.

In Chapter 5, *Noise and Silence in Small RNA Regulation*, we study in detail noise properties of a gene targeted by an sRNA. We predict that sRNA-mediated gene silencing is efficient in attenuating noise. In contrast, the onset of gene expression is dominated by fluctuations, leading to noise induced phenotypic diversity. Using a plasmid-borne synthetic sRNA-target pair in *E. coli*, we present evidence for possible realization of the expected bi-modality in growing cells.

CHAPTER 1

Combinatorial Transcriptional Control of the Lactose Operon of *Escherichia coli*

1.1 Introduction

The *lac* promoter of *E. coli* is one of the most extensively studied systems of molecular biology¹⁻⁶. Knowledge and insight gained from these studies have shaped much of how we now think about gene regulation. It is well known that *E. coli* cells repress the expression of the *lac* operon when glucose is abundant in the growth medium. Only when the glucose level is low *and* the lactose level is high is the operon fully expressed. Thus the regulation of this operon represents an example of “combinatorial control” widely seen in prokaryotes and eukaryotes^{7,8}. In this case, the combinatorial control is implemented molecularly by two transcription factors, the Lac repressor LacR which represses transcription and the cyclic-AMP receptor protein CRP which activates transcription. Activation by CRP requires the inducer cAMP, which is used by *E. coli* cells as a signal of glucose shortage⁹⁻¹⁵. Repression by LacR is activated in a nearly all-or-none manner upon varying the amount of lactose or one of the several synthetic inducers in growth medium with poor carbon sources^{16,17}. The fold-change in repression is very large, (> 1000-fold), and has been shown to involve LacR-mediated DNA looping^{5,6,18-21}.

Here we investigate quantitatively the competing effects of activation and repression on the *lac* promoter *in vivo*. We focus on two perplexing issues: (i) According to biochemical studies²²⁻²⁵, LacR-inducer interaction is only weakly cooperative. By

what mechanism(s) does the observed induction response become so abrupt ^{26,27}? (ii) Despite the well-known role CRP plays in activating transcription ²⁸, structural studies ^{29,30} suggest that CRP enhances *repression* by facilitating the LacR-mediated DNA looping. Moreover, *in vitro* biochemical studies indicate that CRP stabilizes LacR-DNA binding ³¹⁻³³. What functional role(s) does CRP actually play in the control of this operon? We approached these issues by first identifying mutants of *E. coli* MG1655 that allowed us to directly control the activities of the activators and repressors by varying the levels of two inducers in the growth medium. We then characterized the promoter activity systematically for numerous combinations of the inducers. The gene expression data obtained clearly reveal the effect of CRP in enhancing the steepness of the inducer response. We developed a thermodynamic model of gene regulation ^{8,34-36}, incorporating the known molecular mechanisms of LacR-induced DNA looping and its coupling to CRP through DNA bending ^{32,33}. The success of the model is manifested in its ability to describe the complex co-dependence of gene expression on the two inducer levels quantitatively by invoking a *single* parameter, the cooperativity between CRP and LacR-mediated DNA looping, with the fitted value of the cooperativity agreeing well with that determined from *in vitro* biochemical measurements ^{32,33}. Our study presents a proof of concept that the complicated web of interactions coupling repressors, activators, promoters, and DNA loops *in vivo* can be quantitatively dissected, provided that the right modeling, together with a precise sequence of experiments on a systematically picked set of mutants are carried out.

1.2 Repression by LacR

We first characterized quantitatively the activity of the *lac* promoter (*Plac*) subject to varying degrees of repression by the Lac repressor, LacR, for *E. coli* cells in the exponential growth phase. In our experiments, the activity of LacR was modulated by the synthetic gratuitous inducer isopropyl- β -D-thiogalactopyranoside (IPTG)³⁷. We performed the β -galactosidase assays for wild-type *E. coli* MG1655 cells in M9 minimal medium with 0.5% glucose as the carbon source and up to 1 mM of IPTG. “*Plac* activity”, defined here as the product of the β -galactosidase activity and the cell-doubling rate (see Ch. 2.1, *β -galactosidase assay*), is plotted against the corresponding IPTG concentrations as the black crosses in Fig. 1.1*a*.

The data points are fitted to the Hill function

$$\alpha_{IPTG} = b_{IPTG} \cdot \frac{1 + f_{IPTG} \cdot ([IPTG] / C_{IPTG})^{m_{IPTG}}}{1 + ([IPTG] / C_{IPTG})^{m_{IPTG}}}, \quad [1.2.1]$$

shown as the black line in Fig. 1.1*a*. The Hill function is used here (and elsewhere in the Results section) merely to extract the *qualitative* features of the promoter activity; the appropriate quantitative description of the data presented will be provided below in the Analysis section. The features of the IPTG-dependent promoter activity (the “IPTG response”) are conveyed by the best-fit Hill parameters listed in Table 1.2 (row 1): The overall fold change, quantified by $f_{IPTG} \sim 1200$, is in good agreement with previous studies^{5,6,18-20,38,39}. However, the “sensitivity” of the response, quantified by the slope of the transition region in a log-log plot and given approximately by the Hill coefficient

($m_{\text{IPTG}} \approx 4.5$), is much larger than the expected behavior based on the known biochemistry; see below.

One factor contributing to the hyper-sensitivity observed is a positive-feedback effect due to the expression of *lacY* which encodes the Lac permease. This effect is analogous to the well known effect on the induction of the *lac* promoter due to LacY expression¹⁶. Although the uptake of IPTG does not require LacY, Jensen et al⁴⁰ showed that the expression of LacY by the *lac* promoter nevertheless led to enhanced sensitivity. We therefore deleted *lacY* from *E. coli* MG1655 to form strain TK150 (see Table 1.1), and repeated the β -galactosidase assays for this strain. *Plac* activities obtained are plotted as the cyan crosses in Fig. 1.1a. Fitting again to the Hill form yields the cyan line with the best-fit parameters shown in Table 1.2 (row 2). The IPTG response of the $\Delta lacY$ mutant exhibits a broader transition, with $m_{\text{IPTG}} \approx 2.6$.

Another possible cause of this hyper-sensitivity is a suggested cooperative interaction between the Lac repressor and the activator CRP which also binds in the promoter region³¹⁻³³. To investigate this possibility, we deleted the *crp* gene from *E. coli* TK150 to form strain TK230 (Table 1.1), and repeated the β -galactosidase assays for this strain. *Plac* activities obtained are plotted as the green circles in Fig. 1.1a. Fitting again to the Hill form yields the green line with the best-fit parameters shown in Table 1.2 (row 3). The IPTG response of the $\Delta crp \Delta lacY$ double mutant is broader still, with $m_{\text{IPTG}} \approx 2$.

Table 1.1: List of strains and plasmids used in this study.

Strain	Genotype	Derived From	Comments
TK110	$\Delta cpdA$	MG1655	deletion from 3174100 to 3174793
TK120	Δcrp	MG1655	deletion from 3483856 to 3484369
TK130	$\Delta cyaA$	MG1655	deletion from 3988871 to 3991173
TK140	$\Delta lacI$	MG1655	deletion from 365748 to 366642
TK150	$\Delta lacY$	MG1655	deletion from 361259 to 362274
TK200	$\Delta lacI \Delta lacY$	TK140	
TK210	$\Delta cpdA \Delta lacY$	TK110, TK150	
TK220	$\Delta crp \Delta lacI$	TK110, TK140	
TK230	$\Delta crp \Delta lacY$	TK120, TK150	
TK240	$\Delta cyaA \Delta lacI$	TK130, TK140	
TK250	$\Delta cyaA \Delta lacY$	TK130, TK150	
TK310	$\Delta cyaA \Delta cpdA \Delta lacY$	TK110, TK130, TK150	
TK320	$\Delta cyaA \Delta lacI \Delta lacY$	TK130, TK200	
TK330	$\Delta crp \Delta lacI \Delta lacY$	TK120, TK200	
BW25113	$lacI^A rmB_{T14}$ $\Delta lacZ_{WJ16} hsdR514$ $araBAD_{AH33}$ $\Delta rhaBAD_{LD78}$	K12 BD792	Ref. 73
TK400	$\Delta cyaA$	BW25113	deletion from 3988871 to 3991173
TK500	$\Delta cyaA \Phi(P_{crp}-lacZ)$	TK400	P_{crp} amplified from 3483424 to 3483721
TK600	$\Delta cyaA \Phi(P_{lac\Delta crp}-lacZ)$	TK400	CRP operator at -61.5 altered from TGTGAGTTAGCTCACT to CAGACGTTAGCTCACT

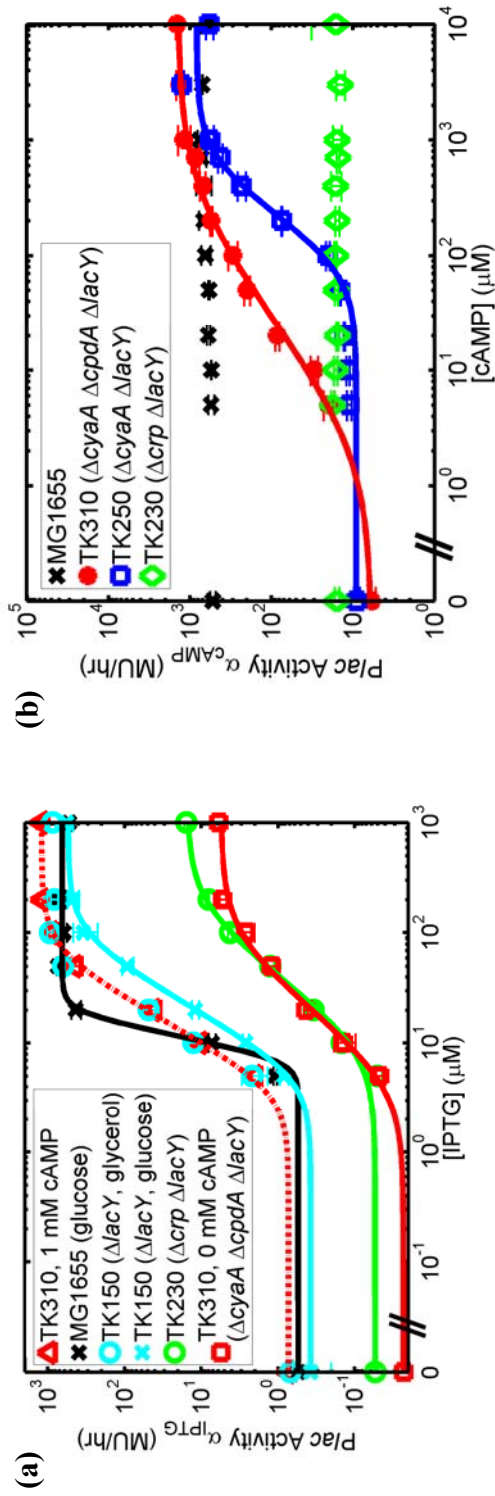


Figure 1.1: Dependence of Plac activity on the inducers. (a) The IPTG response for *E. coli* MG1655 and various Δ lacY mutants, grown in minimal M9 medium with varying amounts of IPTG and 0.5% glucose except for the cyan circles (0.5% glycerol). No cAMP was added to the medium except for the red triangles where 1mM cAMP was added. The lines are best fits to the Hill function (Eq. [1.2.1]). (b) The cAMP response for *E. coli* MG1655 and various Δ cyaA mutants, grown in minimal M9 medium + 0.5% glucose, 1mM IPTG and varying amount of cAMP. The lines are best fits to the Hill function (Eq. [1.2.1]).

Table 1.2: Hill parameters for the IPTG response. Parameters derived from fit of the IPTG dependence of the Plac activity to the Hill form (Eq. [1.2.1]), for various strains of *E. coli* derived from MG1655, grown in medium with 0.5% glucose and varying amounts of IPTG. For the last row, 1mM cAMP was also added to the growth medium.

Strain	b_{IPTG} (MU/hr)	f_{IPTG}	C_{IPTG} (μM)	m_{IPTG}
MG1655	0.5	1200 ± 80	20 ± 4	4.5 ± 0.7
TK150 (ΔlacY)	0.4	1445 ± 185	100 ± 18	2.6 ± 0.2
TK230 ($\Delta\text{crp } \Delta\text{lacY}$)	0.05	255 ± 19	150 ± 15	2.0 ± 0.1
TK310 ($\Delta\text{cyaA } \Delta\text{cpdA } \Delta\text{lacY}$) no cAMP	0.02	238 ± 26	90 ± 5	2.0 ± 0.2
TK310, 1 mM cAMP	0.7	1600 ± 180	70 ± 8	2.8 ± 0.1

Table 1.3: Hill parameters for the cAMP response. The first two rows give parameters derived from fit of cAMP response of *cyaA* mutants to the Hill function (Eq. [1.3.1]). The last row gives best estimates of these parameters according to the thermodynamic model and the known biochemical parameters; see Chapter 2.6, *Activation by cAMP-CRP* for details.

Strain	b_{cAMP} (MU/hr)	f_{cAMP}	C_{cAMP} (μM)	m_{cAMP}
TK250 ($\Delta\text{cyaA } \Delta\text{lacY}$)	9.1	91 ± 5	645 ± 43	2.1 ± 0.1
TK310($\Delta\text{cyaA } \Delta\text{cpdA } \Delta\text{lacY}$)	5.9	240 ± 13	320 ± 32	1.0 ± 0.1
Prediction based on thermodynamics	---	> 20	10 – 1000	1

Table 1.4: Parameters for model with DNA looping. Parameters derived from the fit of IPTG response of given strains to the thermodynamic model with DNA looping (Eq. [1.7.1]).

Strain	K_{IPTG} (μM)	L_0
TK230 ($\Delta\text{crp } \Delta\text{lacY}$)	15.7 ± 1.5	4.1
TK310 ($\Delta\text{cyaA } \Delta\text{cpdA } \Delta\text{lacY}$)	12.3 ± 1.4	3.7
TK250 ($\Delta\text{cyaA } \Delta\text{lacY}$)	17.8 ± 1.7	5.0

1.3 Activation by CRP

We next characterized the dependence of the *Plac* activity to different degrees of activation by cAMP-CRP (the “cAMP response”). To avoid possible complications due to interaction between LacR and CRP, all experiments were performed under saturating IPTG concentration (1 mM) to disable LacR-operator interaction. This was complemented by direct deletion of LacR in some cases (see below). We also deleted *lacY* in all subsequent experiments in order to avoid possible feedback. In all cases discussed below where we directly compared the *Plac* activity of *lacY*[−] and *lacY*⁺ cells, differences of no more than 2-fold were obtained (data not shown).

One way to manipulate the cellular level of cAMP-CRP is to subject cells to different levels of cAMP in the medium, and rely on the diffusion of cAMP into cells. This approach requires shutting off the endogenous synthesis of cAMP by the enzyme adenylate cyclase (AC), encoded by *cyaA*⁴¹⁻⁴⁴. Setty et al²⁷ attempted this approach by growing cells in medium with 0.2% glucose and varying levels of cAMP, expecting that AC activity would be repressed via catabolite repression^{10,14,15,45-49}. However, they observed only a few-fold change in *Plac* activity despite large variations in the extracellular cAMP levels (0 to 20 mM)²⁷. The observed change was surprisingly small given that > 50-fold difference in *Plac* activity was obtained between the wild-type and Δcrp strains³. In fact, a nearly 10-fold difference in *Plac* activity can be seen by simply growing wild-type cells on various sugars; see Fig. 1.2 (blue bars).

We reasoned that the small change in *Plac* activity obtained by Setty et al might have resulted from the incomplete repression of AC activity by glucose uptake, and

hence repeated the experiment with the deletion of *cyaA*. *E. coli* TK250 strain ($\Delta cyaA \Delta lacY$, see Table 1.1) was grown in M9 minimal medium with 0.5% glucose, 1mM IPTG, and up to 10mM cAMP. β -galactosidase activity was assayed as described above. The resulting *Plac* activity displays a smooth sigmoidal dependence as shown in Fig. 1.1b (blue squares). [Almost identical results (not shown) were obtained for *LacR* null mutant (strain TK320), indicating that LacR is indeed not a factor with saturating IPTG (1 mM) in the growth medium.] A ~ 100 -fold difference is seen between the low and high cAMP concentrations, comparable to the difference in *Plac* activity reported between *crp* null mutant and wild-type strains of *E. coli* ³. In contrast, *Plac* activity of the wild-type (black symbols in Fig. 1.1b) as well as the $\Delta lacY$ mutant (not shown) grown in glucose displayed only ~ 3 -fold change over the same range of cAMP levels, similar to the aforementioned finding by Setty et al ²⁷. As a negative control, we show in Fig. 1.1b (green circles) the promoter activity obtained for the *crp*⁻ strain (TK230); its lack of cAMP dependence indicates that the observed cAMP dependence for the *crp*⁺ strain was mediated primarily by cAMP-CRP. Additional negative controls on possible indirect effects of cAMP variations on *Plac* activity are shown in Fig. 1.3b. We found the variation of CRP expression and the non-specific effects of CRP on *Plac* activity to be small (~ 2 -fold), compared to the 100-fold difference observed for the $\Delta cyaA$ mutants over the same range of cAMP concentrations.

The cAMP response exhibited by strain TK250 was analyzed by fitting to the Hill function,

$$\alpha_{cAMP} = b_0 \cdot \frac{1 + f_{cAMP} \cdot ([cAMP] / C_{cAMP})^{m_{cAMP}}}{1 + ([cAMP] / C_{cAMP})^{m_{cAMP}}}, \quad [1.3.1]$$

and plotted as the blue line in Fig. 1.1*b*. The best-fit parameters are shown in Table 1.3 (row 1). The sensitivity of the cAMP response, characterized by the Hill coefficient $m_{\text{cAMP}} \approx 2$, is in disagreement with the non-cooperative nature of cAMP-CRP interaction⁵⁰⁻⁵³. This suggests a nonlinear relationship between the extracellular and intracellular cAMP concentrations in ΔcyaA strain, and prompted us to look for additional factors regulating intracellular cAMP levels.

One such factor is cAMP degradation catalyzed by the enzyme cAMP-phosphodiesterase (PDE)⁵⁴⁻⁵⁸, encoded by *cpdA*⁵⁹. We deleted the *cpdA* gene to obtain strain TK310 ($\Delta\text{cyaA} \Delta\text{lacY} \Delta\text{cpdA}$) (see Table 1.1) and repeated the β -galactosidase assay and analysis. The cAMP response obtained (red circles in Fig. 1.1*b*) is more gradual than that of *cpdA*⁺ cells (blue squares). Fitting the data to the Hill form [1.3.1] yielded the red line with the Hill parameters listed in Table 1.3 (row 2). Specifically, the Hill coefficient $m_{\text{cAMP}} \approx 1$ is consistent with the naïve expectation based on the non-cooperative nature of cAMP-CRP interaction⁵⁰⁻⁵³.

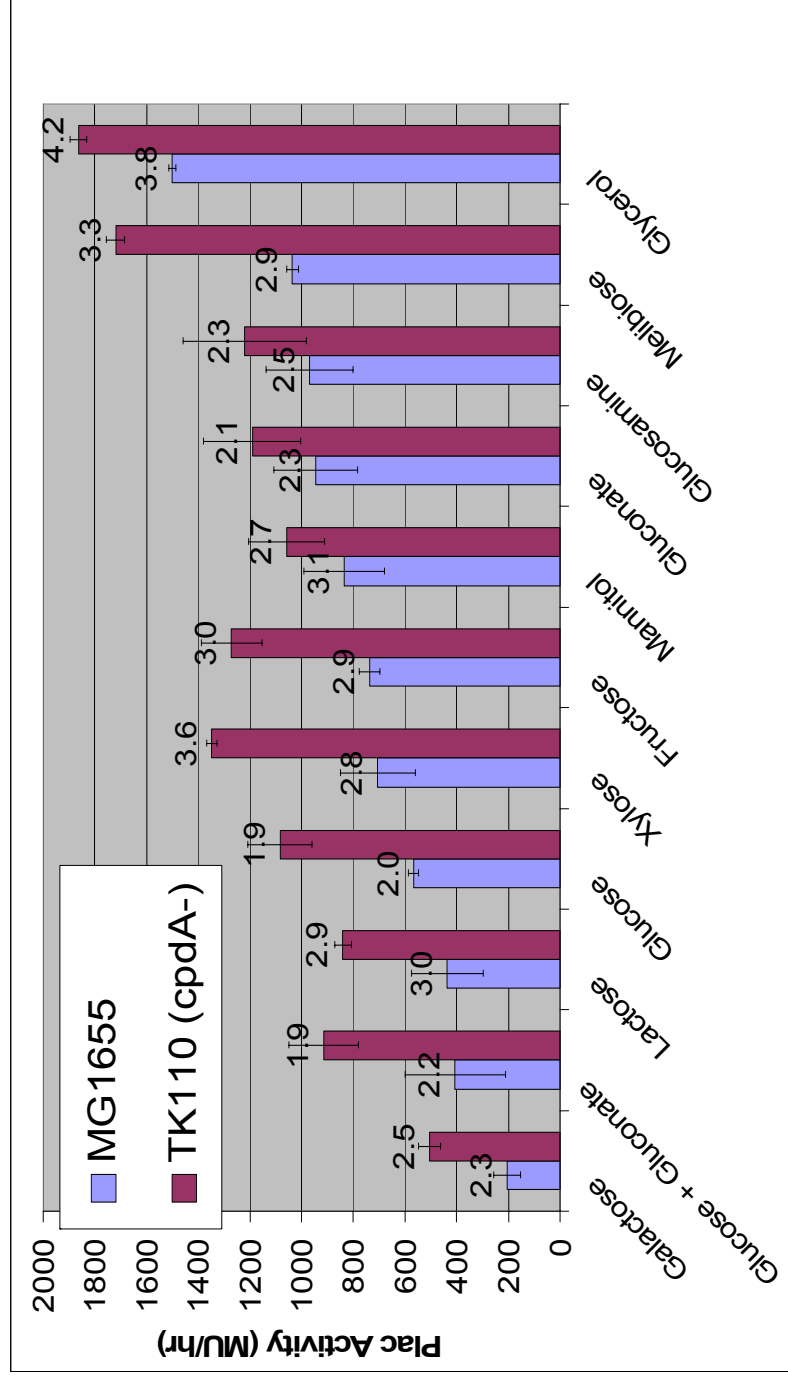


Figure 1.2: Plac activity of *E. coli* MG1655 cells and Δ cydA mutants grown with various sugar sources. The cells are grown in minimal M9 medium with 1mM IPTG and with 0.5% of the indicated sugars as the sole carbon source. The number above each bar is the observed doubling times (in hours) for that strain and medium.

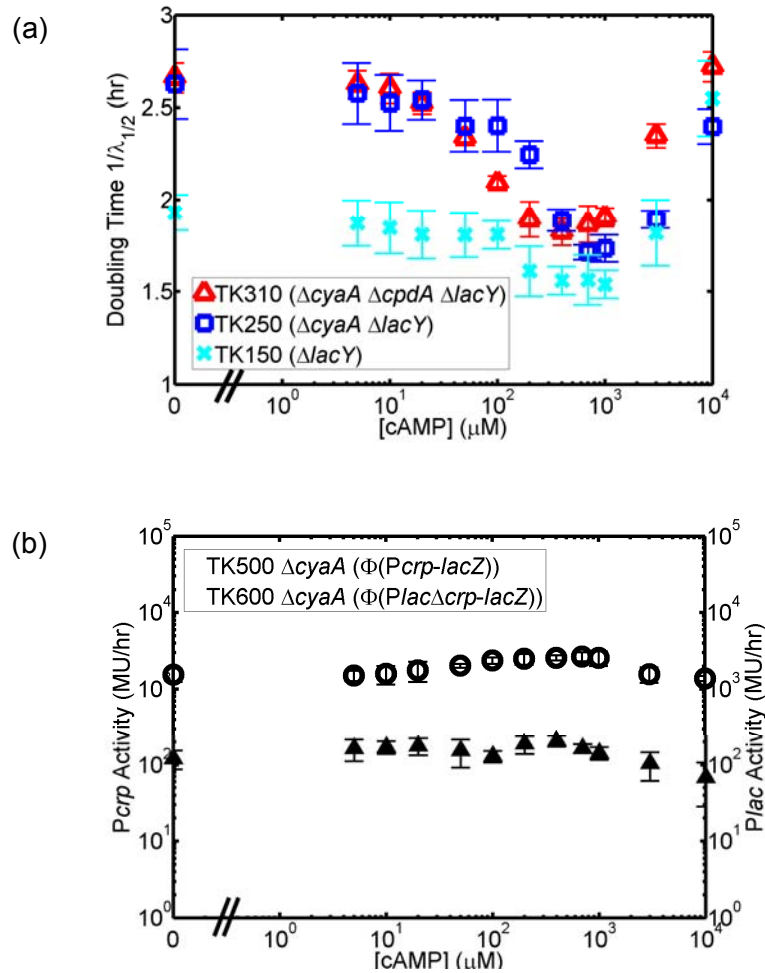


Figure 1.3: Indirect effect of cAMP. (a) The doubling time $1/\lambda_{1/2}$ of various mutant strains of *E. coli* MG1655 grown in M9 minimal medium with 0.5% glucose and varying amounts of cAMP. For the *cyaA* strains (blue and red symbols), the growth rate varied by as much as 40%, with the most rapid growth in medium with 1mM cAMP. (b) The activity of the *crp* promoter was determined by the β -galactosidase assay for strain TK500 (BW25113 cells bearing *cyaA* deletion and chromosomal insertion of the *crp* promoter transcriptionally fused to *lacZ*). The activity was found to vary by no more than 2-fold in medium containing 0-10 mM cAMP, with maximal activity at ≈ 1 mM cAMP. The indirect effect of cAMP-CRP on the *lac* promoter was determined by the β -galactosidase assay for strain TK600 (BW25113 cells bearing *cyaA* deletion and chromosomal insertion of a mutated *lac* promoter:*lacZ* construct, with the promoter containing a mutated CRP operator site; see Table 1.1). The activity of this promoter was nearly unchanged in media containing 0-1mM cAMP, and decreasing only by 2-fold in medium with 1-10 mM cAMP. (1mM IPTG was added to the medium to alleviate repression by LacR.) The results indicate that the cAMP dependence observed for wild-type *lac* promoter (Fig. 1.1b) is primarily due to cAMP-CRP exerted through the CRP operator site.

1.4 Combinatorial Control

We next investigated the co-dependence of *Plac* activity on the two regulators, LacR and cAMP-CRP. We showed above that the $\Delta crp \Delta lacY$ strain (TK230) could be used to characterize the bare IPTG response, while the $\Delta lacY \Delta cyaA \Delta cpdA$ strain (TK310) could be used to characterize the bare cAMP response. In order to characterize the co-dependence of the promoter on IPTG and cAMP, we first verified that strain TK310 exhibited nearly the same IPTG dependence as TK230 in growth medium with no cAMP added; compare† the red squares and green circles in Fig. 1.1a. Fitting the IPTG-response of TK310 (red squares) to the Hill form [1.2.1] yields the solid red line; the corresponding parameters are provided in Table 1.3 (row 4).

We repeated the β -galactosidase assay and analysis for TK310 cells grown in media with various *combinations* of IPTG and cAMP concentrations. As evidence of interaction between IPTG-mediated and cAMP-mediated regulations, we show in Fig. 1.1a (red triangles) the IPTG response for TK310 strain in growth medium containing 1 mM cAMP. This response is nearly indistinguishable from the physiological *Plac* activity exhibited by $cyaA^+ cpdA^+$ cells in glycerol medium (cyan circles). The dashed red line is the best fit to the Hill function [1.2.1], with parameters listed in Table 1.3 (row 5). Note that the overall fold-change (f_{IPTG}) is increased from $< 250x$ when no cAMP was added in the medium, to $> 1500x$ with 1mM cAMP in the medium. The latter fold-change is comparable to those obtained for wild-type *E. coli* cells grown in the absence or saturating concentration of IPTG (black crosses in Fig. 1.1a, and Refs. 5,6,18-20,38,39). Additionally, the sensitivity of the IPTG response increased from

$m_{\text{IPTG}} \approx 2.0$ (Table 1.2 row 4) for TK310 cells grown in the absence of cAMP to $m_{\text{IPTG}} \approx 2.8$ (Table 1.2 row 5) for the same cells grown in 1mM cAMP. Fitting the IPTG responses of these cells obtained at a variety of cAMP concentrations, we found a trend of increasing Hill coefficient (from 2 to 3) and fold-change (from 250x to 1800x) for cAMP levels from 1 μM to 1 mM; see Figs. 1.4. The complete co-dependence of *Plac* activity on IPTG and cAMP is shown as the 3D plot in Fig. 1.5a.

[†] However, *cpdA* mutants show a 2-fold overall reduction in gene expression for unknown reasons.

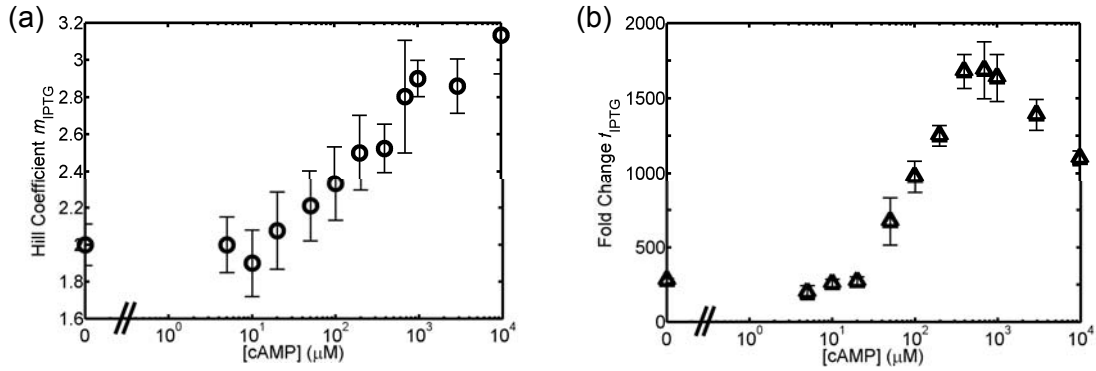


Figure 1.4: cAMP dependence of IPTG response. Values of the Hill parameters m_{IPTG} (a) and f_{IPTG} (b) obtained from fitting the IPTG response of TK310 cells grown in media with the indicated cAMP concentrations

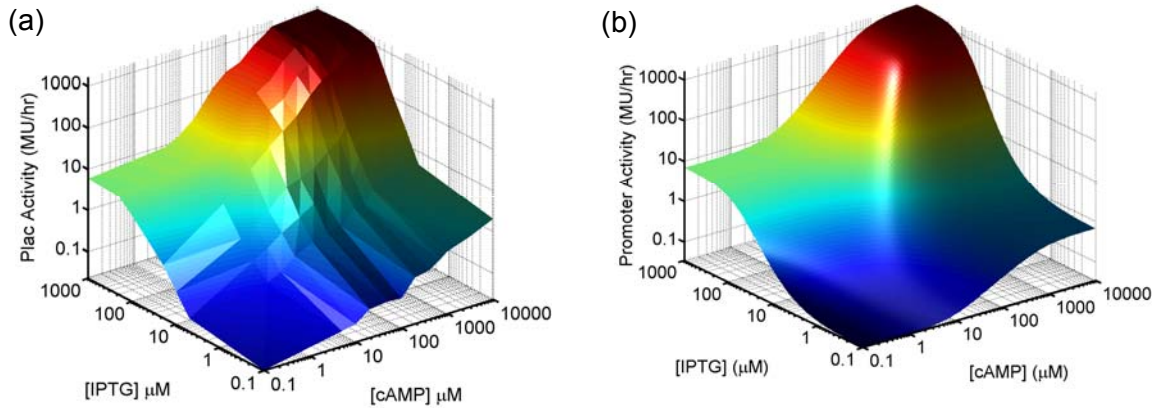


Figure 1.5: Combinatorial control of *Plac* activity for strain TK310. (a) Three-dimensional plot of *Plac* activity for strain TK310 ($\Delta lacY \Delta cyaA \Delta cpdA$) grown in M9 medium with varying amounts of cAMP and IPTG. (b) Predicted form of *Plac* activity according to the extended thermodynamic model. The function is generated using only five parameters: K_{IPTG} and L_0 (Table 1.4, row 2) for the bare IPTG response, C_{cAMP} and f_{cAMP} (Table 1.3, row 2) for the bare cAMP response, and $\Omega = 10.3$ for the co-dependence.

Analysis

We have seen that the IPTG and cAMP responses of various mutant strains of *E. coli* MG1655 fitted well to Hill functions, with the Hill parameters summarized in Tables 1.2 and 1.3. However, the Hill function itself has been invoked so far without justification; it was merely a familiar form used to quantify key features of the response, e.g., the overall fold-change and sensitivity. Below we will analyze and interpret the results obtained in light of the rich knowledge on the molecular biology of the *lac* promoter and the biochemistry of the associated components using a thermodynamic model of transcriptional regulation; see Refs. 8,36 and Chapter 2.5, *Thermodynamic modeling*.

1.5 Activation by cAMP-CRP is non-cooperative

The cAMP response found for strain TK310 ($\Delta cyaA \Delta cpdA \Delta lacY$) exhibited a broad transition (red circles in Fig. 1.1*b*), well fitted by the Hill form (Fig. 1*b*, red line). This is in agreement with the thermodynamic model which predicts the Hill form for response to simple activation by cAMP-CRP; see Bintu et al ³⁵ and Chapter 2.6, *Activation by cAMP-CRP*. The Hill coefficient $m_{cAMP} \approx 1$ obtained is in good agreement with the biochemistry finding that it takes one cAMP molecule to activate the CRP dimer ⁵¹⁻⁵³.

The thermodynamic model further relates the other Hill parameters f_{cAMP} and C_{cAMP} to the biochemical parameters that describe CRP-mediated transcriptional activation; see Chapter 2.6, *Activation by cAMP-CRP*. Best estimates of f_{cAMP} and C_{cAMP}

based on knowledge of the biochemical parameters are given in Table 1.3 (row 3). The available information is not sufficient for a quantitative comparison of the parameter C_{cAMP} , whose value depends on the CRP-operator binding affinity *in vivo* as well as the relation between the intra- and extra- cellular cAMP concentrations, both of which can be estimated only very crudely^{9,60}; see Chapter 2.6, *Activation by cAMP-CRP*. The parameter f_{cAMP} , which describes the maximal fold-change in the cAMP response, is given by the cooperativity of CRP and RNA polymerase (RNAP) interaction according to the thermodynamic model. Our result $f_{\text{cAMP}} = 240 \pm 13$ is significantly larger than the cooperativity factor of ~ 20 obtained from *in vitro* biochemical measurements^{61,62}. This discrepancy is analogous to one noted earlier by Beckwith et al.³. It may have resulted from the accumulation of a number of small factors. For example, in addition to recruiting RNAP, CRP was shown to stimulate the transition of promoter DNA from the close to open conformation, thereby enhancing the transcription rate by $\sim 50\%$ ⁶¹. Also, the autoregulation of CRP expression may account for ~ 2 -fold difference (see Fig. 1.3*b* and caption), and another $2\sim 3$ fold difference may be attributed to the deletion of *cpdA* (compare the vertical ranges of the blue and red lines in Fig. 1.1*b*).

1.6 PDE provides insulation to variations in cAMP

We are not aware of any significant phenotype reported for cells with a PDE deletion. Only small differences of 2 – 3 fold in *Plac* activity were seen between the wild-type and ΔcpdA strains (blue and red bars in Fig. 1.2), while no systematic trend can be seen in the growth rates of the two strains (the numbers on top of the bars in Fig. 1.2)

either. However, the effect of *cpdA* expression on the *cyaA* mutant is striking. Comparison of the blue and red lines in Fig. 1.1*b* suggests that PDE expression *insulates* the cell from extracellular cAMP variations of up to 100 μ M. This may be important for cells in environments where AC activity is significantly repressed.

1.7 DNA looping increases the sensitivity of the IPTG response

The IPTG response of the $\Delta crp \Delta lacY$ double mutant (TK230) exhibited a reduced sensitivity (Fig. 1.1*a* green circles) compared to the *lacY* mutant (TK150, cyan crosses in Fig. 1.1*a*). The difference is due to the activated CRP in the latter strain (which has *cyaA* intact and hence synthesizes cAMP endogenously). We will discuss the effect of CRP shortly; for now, we first discuss the IPTG response in the absence of CRP, i.e., that of strain TK230. This response is well fitted by the Hill form (green line), with the Hill coefficient $m_{IPTG} \approx 2.0$ (Table 1.2 row 3). While a cooperative IPTG response with Hill coefficient ~ 2 is widely quoted in the molecular biology literature²⁶, and moreover a cooperative IPTG-LacR interaction was suggested based on a structural study of LacR²⁹, *in vitro* biochemical studies of IPTG-LacR binding in fact found a non-cooperative IPTG-LacR interaction, which became only weakly cooperative (Hill coefficient = 1.4 $\sim$ 1.6) in the presence of operator DNA fragments²²⁻²⁵. We reasoned that the apparent cooperativity ($m_{IPTG} \sim 2$) observed *in vivo* might have resulted from LacR-mediated *DNA looping*[‡] known to occur in wild-type *E. coli* cells^{6,18-20}. The corresponding

[‡] We note that the increased in cooperativity by DNA looping was discussed recently by Vilar & Saiz⁶³ in the context of the lysis-lysogeny control of Phage λ . However, the mechanism of cooperativity there is the oligomerization of regulatory proteins made possible through DNA looping. In contrast here, the

thermodynamic model is developed in Chapter 2.8, *LacR-mediated DNA looping*, with the following form of the IPTG response

$$\alpha_{IPTG} = \alpha_0 / \left[1 + \frac{R}{(1 + [IPTG] / K_{IPTG})^2} + \frac{R \cdot L_0}{(1 + [IPTG] / K_{IPTG})^4} \right]. \quad [1.7.1]$$

Although Eq. [1.7.1] is not of the Hill form, it resembles a spectrum of different Hill functions, with the maximal fold-change

$$f_{IPTG} = 1 + R \cdot (L_0 + 1), \quad [1.7.2]$$

and with the sensitivity also increasing with the parameters R and L_0 ; see Chapter 2.8, *LacR-mediated DNA looping*. The parameter R depends on the LacR concentration and LacR-operator affinity; it gives the fold-repression between zero and saturating IPTG levels in the absence of DNA looping ($L_0 = 0$). The values of $R \approx 20 - 50$ can be inferred from two previous studies by Oehler et al ^{6,20}, who characterized the activities of *lac* promoters bearing various Lac operator mutations; see Chapter 2.8, *LacR-mediated looping*. In the analysis below, we will use $R \approx 50$; results of similar quality (not shown) can be obtained for $R \approx 20$. The parameter L_0 describes the enhancement in the “local concentration” of LacR molecules due to DNA looping ^{6,35}. There is currently no direct measurement of L_0 for the *lac* promoter in the absence of CRP binding.

Before applying Eq. [1.7.1] to our results, we first discuss its form in the absence of DNA looping ($L_0 = 0$). In a very recent report, Oehler et al ⁶⁴ used various mutants to characterize the IPTG responses of the *lac* promoter incapable of DNA looping. Their

tetramerization of LacR is independent of DNA looping. What DNA looping enhances here is the *sensitivity* to IPTG rather than LacR.

data are reproduced in Fig. 1.6a (red triangles and blue squares) and shown to fit well to Eq. [1.7.1] with $L_0 = 0$ (solid line). The data are equally well-fitted in Fig. 1.6b by the Hill form [1.2.1] with an effective Hill coefficient of 1.5, in agreement with results of previous *in vitro* studies²⁴. Together, the goodness of these fits validates the above quantitative description of LacR-operator interaction based on the thermodynamic model.

We next confront DNA looping by fitting the IPTG response of strain TK230 (the green circles in Fig. 1.1a) to Eq. [1.7.1], with two fitting parameters, K_{IPTG} and L_0 . (The saturation value α_0 is fixed by the promoter activity at 1mM IPTG.) The best-fit curve (not shown) is nearly indistinguishable from the green line in Fig. 1.1a even though Eq. [1.7.1] contains one fewer fitting parameter than Eq. [1.2.1]. In particular, the observed sensitivity of the IPTG dependence is well reproduced. The best-fit parameters obtained are shown in Table 1.4 (row 1). The value of the loop parameter simply reflects the maximal fold-change of the IPTG response as given by Eq. [1.7.2]. The value of the IPTG-LacR dissociation constant K_{IPTG} obtained is comparable to the results of *in vitro* biochemical studies; see Refs. 4, 65, 66 and Chapter 2.4, *Activation of the regulatory proteins*. Similarly, we found that the DNA looping model described the IPTG response of strain TK310 ($\Delta\text{cyaA } \Delta\text{cpdA } \Delta\text{lacY}$) very well in medium with no cAMP. The corresponding best-fit parameters are given in Table 1.4 row 2; the best-fit curve (not shown) is nearly indistinguishable from the lower red line in Fig. 1.1a.

1.8 Cooperativity between CRP and LacR enhances sensitivity to IPTG

Strain TK310, which exhibited the expected cAMP response in medium with 1mM IPTG, and the expected IPTG response in medium with no cAMP added, was used

to investigate the *combinatorial control* of *Plac* activity by LacR and cAMP-CRP. According to the thermodynamic model, if there is no interaction between cAMP-CRP mediated activation and LacR mediated repression, then the co-dependence of *Plac* activity on IPTG and cAMP would simply be the algebraic product of the IPTG response and the cAMP response, i.e., the product of the red lines in Fig. 1.1*a* and Fig. 1.1*b*. A manifestation of this product form is a simple vertical shift of the IPTG responses of the same strain at different cAMP concentrations in the log-log plots. However, this clearly cannot be the case for strain TK310, as both the sensitivity and the maximal fold-change of its IPTG response increase with increasing cAMP levels (Figs. 1.4 and Fig. 1.2); see also Fig. 2.3*a* for a direct comparison.

Various forms of interactions between CRP and LacR in the *lac* promoter region have been previously proposed or reported^{29,31-33,67,68}: Since CRP causes a 90°-130° bend in DNA^{20,69,70}, the binding of CRP to its site in the *lac* promoter could potentially bring the Lac operators O1 and O3 into closer proximity, promoting the formation of a DNA loop between them²⁹; see Figs. 1.7. The existence of this interaction is supported by a computational study based on the available LacR-operator and CRP-operator co-crystal structures³⁰. Moreover, Hudson & Fried³² and Vossen et al³³ found cooperativity in the formation of a ternary complex of CRP, LacR, and *lac* promoter DNA *in vitro*, with the simultaneous presence of LacR and cAMP-CRP increasing the affinity of both proteins for their respective binding sites by a cooperative factor (Ω) of 4-12 fold.

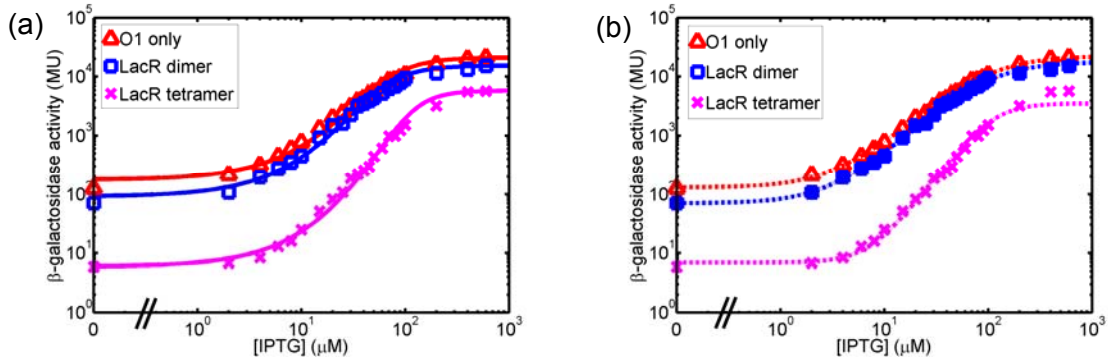


Figure 1.6: IPTG response in the absence of DNA looping. Data from Oehler *et al.*⁶⁴ with overexpressed LacR. Strains include wild-type promoter and repressor (LacR tetramer; magenta crosses), wild-type promoter with dimeric repressor (LacR dimer; blue squares), and wild-type repressor with promoter including O1 binding site only (O1 only; red triangles); **(a)** fit of the nonlooping mutants, red triangles and blue squares, to Eq. [1.7.1] with $L_0 = 0$, yielding $R \approx 200$ and $K_{\text{IPTG}} \approx 6 \mu\text{M}$. The wild-type fit to the looping model (with $R = 200$ and L_0 fixed by the fold-change); **(b)** Fit of the same data to the Hill form. O1 only and LacR dimer strains yield Hill coefficient $m_{\text{IPTG}} = 1.5$; wild-type strain yields $m_{\text{IPTG}} = 2.5$.

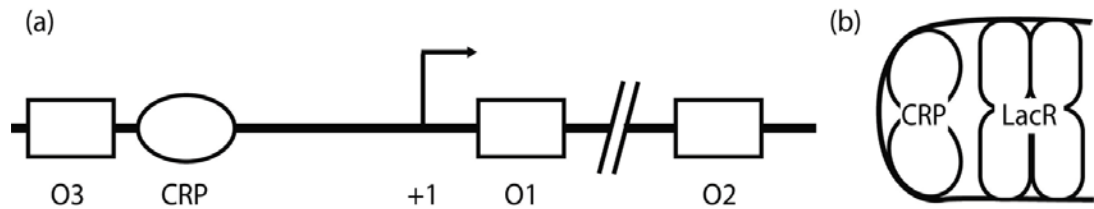


Figure 1.7: Schematics of the *lac* promoter. **(a)** The main Lac operator O1 is located at position +11 from the transcriptional start. The auxiliary Lac operator O2 is located at +411 and O3 at -81. The CRP operator, at position -62.5, is located in between O1 and O3. **(b)** Illustration of the hypothesized interaction between CRP and LacR: The binding of CRP to its operator site bends the DNA in a way to facilitate the looping between O1 and O3.

To see whether CRP-assisted DNA looping may quantitatively explain the observed co-dependency of *Plac* activity, we extended the thermodynamic model with DNA looping to include the CRP-dependent looping effect; see Chapter 2.9, *Combinatorial control by LacR and cAMP-CRP*. The extended model predicts the IPTG response to have the same form as Eq. [1.7.1], but with the prefactor α_0 replaced by the Hill function α_{cAMP} for strain TK310 (Eq. [1.3.1] with parameters given by Table 1.3 row 2), and with the loop parameter L_0 replaced by

$$L = L_0 \cdot \frac{1 + \Omega \cdot [\text{cAMP}] / C_{\text{cAMP}}}{1 + [\text{cAMP}] / C_{\text{cAMP}}} \quad [1.8.1]$$

where Ω is the cooperativity factor of the CRP-LacR-loop interaction.

To test the predictions of the extended thermodynamic model, we used Eq. [1.7.2] to infer the value of the loop parameter L from the maximal fold-change of the IPTG response at each cAMP value; the result is plotted as the circles in Fig. 1.8a. We then used Eq. [1.7.1] to compute the expected IPTG responses using these values of L . The results are shown as the dashed lines in Fig. 1.8b for several representative cAMP values, together with the experimental data for TK310 cells grown in media with the corresponding cAMP concentrations. Note that the dashed lines are *not* fits of the data to the model. The striking congruence between the dashed lines and the data points (for $[\text{cAMP}] \leq 1$ mM) shows that the cAMP-dependent sensitivity, which requires a series of different Hill coefficients within the Hill function description (Fig. 1.8a), can be naturally described by the DNA looping model in terms of the cAMP-dependent loop parameter L with no other adjustable parameters.

To further examine whether the cAMP-dependence of L conforms to the known energetics of CRP-assisted DNA looping, we fitted the inferred values of L (Fig. 1.8a circles) to Eq. [1.8.1], by allowing only the parameter Ω to vary. The result of the fit (Fig. 1.8a solid line) described the data well, except again at the highest cAMP level (10 mM) where cell growth significantly slowed down and other effects (e.g., over expression of nucleoid proteins) might significantly affect DNA bending and looping⁷¹. The value of the best-fit parameter obtained, $\Omega = 10.3 \pm 0.1$, is in agreement with the 4-12 fold cooperativity found *in vitro*^{31,33}; Fig. 1.5b shows the 3D plot of the expected *Plac* activity according to the extended thermodynamic model. Collectively, these results provide strong quantitative support for the hypothesis of CRP-assisted repression by DNA looping³¹⁻³³.

Finally, we repeated the analysis for the combinatorial control of TK250 cells which are *cpdA*⁺. Despite the unexpected cooperative nature of the cAMP response of this strain (blue symbols in Fig. 1.1b with $m_{cAMP} \approx 2$), the IPTG responses at various cAMP levels (Fig. 1.9b) are still well described by Eq. [1.7.1] (with the best-fit values of K_{IPTG} and L_0 given in Table 1.4 row 3). Moreover, except for the highest cAMP concentrations again, the cAMP-dependence of the loop parameter L agrees well with the corresponding Hill function, with similar value of the cooperativity factor, $\Omega \approx 9.7$; see Fig. 1.9a and caption. These results indicate that *cpdA* expression only affects the activation of CRP by extracellular cAMP, but not the interaction of cAMP-CRP with DNA looping itself.

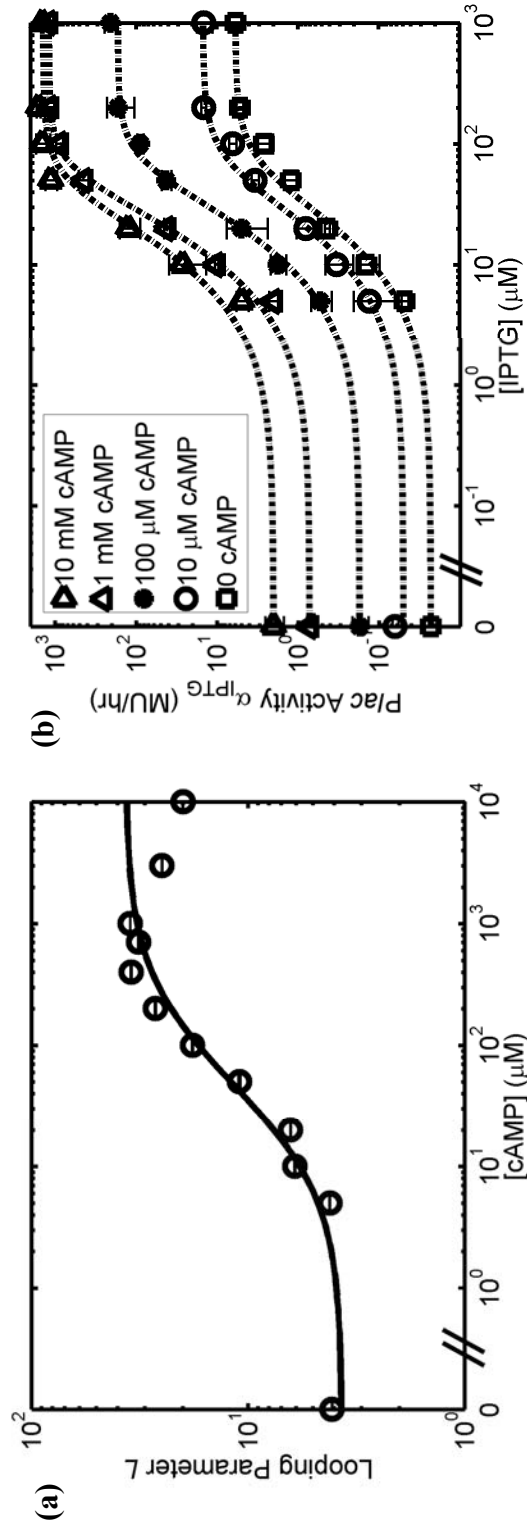


Figure 1.8: Combinatorial control of Plac activity for strain TK310. (a) The cAMP-dependent loop parameter L is inferred from Eq. [1.7.2] by fixing the maximal fold-change f_{IPTG} to be the ratio of α_0 and the observed promoter activity in medium with no IPTG. α_0 is generated by the bare cAMP response α_{cAMP} using the parameters in Table 1.3 row 2. The result obtained (circles) is fitted to Eq. [1.8.1] using a single fitting parameter $\Omega = 10.3 \pm 0.1$, with $C_{cAMP} = 320 \mu M$ (Table 1.3 row 2) and $L_0 = 3.7$ (Table 1.4 row 2). (b) The IPTG responses obtained at different cAMP levels (symbols) are plotted together with predictions of the CRP-assisted DNA-looping model (Eq. [1.7.1]) with no adjustable parameters. We used $R = 50$, $K_{IPTG} = 12.3 \mu M$ (Table 1.4 row 2), with values of L taken from (a) and α_{cAMP} for α_0 (Eq. [1.3.1]) with parameters values given in Table 1.3 row 2).

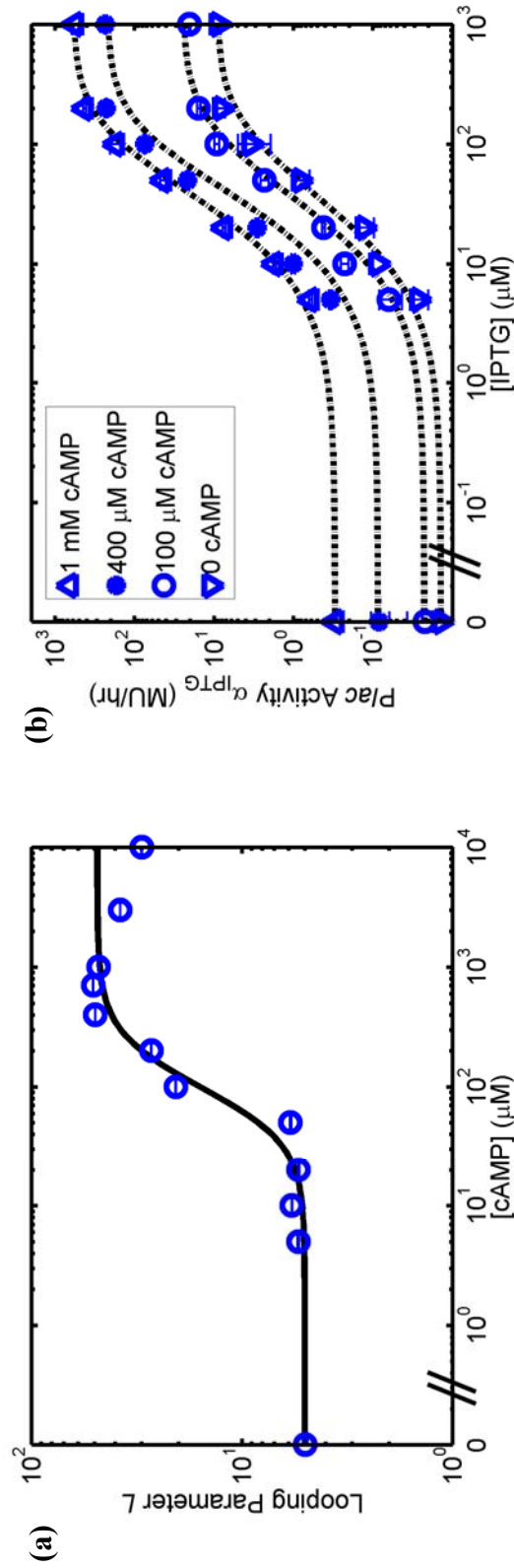


Figure 1.9: Combinatorial control of *Plac* activity for strain TK250. (a) The cAMP-dependent loop parameter L is inferred from Eq. [1.7.2] by fixing the maximal fold-change f_{IPTG} to be the ratio of α_0 and the observed promoter activity in medium with no IPTG. α_0 is generated by the bare cAMP response α_{cAMP} using the parameters in Table 1.3, row 1. The result obtained (circles) is fitted to Eq. [2.9.10] using a single fitting parameter, Ω , determined to be 9.7, with $C_{cAMP} = 645 \mu M$ (Table 1.3, row 1) and $L_0 = 5.0$ (Table 1.4, row 3). (b) The IPTG responses obtained at different cAMP levels (symbols) are plotted together with predictions of the DNA-looping model (Eq. [1.7.1]) with no adjustable parameters. We used $R = 50$, $K_{IPTG} = 17.8 \mu M$ (Table 1.4, row 3), with values of L taken from (a) and α_{cAMP} (Eq. [1.3.1] with parameters in Table 1.3, row 1) for α_0 (see Chapter 2, *Experimental and Theoretical Methods*).

Discussion

Systematic, quantitative characterization of gene expression may yield unique insights on the mechanisms of gene regulation. A major obstacle in this effort is to quantify the amount of *active* transcription factors in cells in a given environment. As attempted earlier by Setty et al ²⁷, we tackled this problem for the *lac* promoter by controlling the amount of inducers IPTG and cAMP extracellularly, but using *cyaA* mutant of *E. coli* which cannot synthesize cAMP endogenously. One danger of this approach is the array of mechanisms the cell may employ to control the intracellular effector levels. We encountered two such mechanisms in our study: Through the positive feedback of the Lac permease, *E. coli* cells sensitize themselves to the presence of inducers of the Lac repressor, while through the activity of cAMP-phosphodiesterase (PDE), *E. coli* cells appear to insulate themselves from low levels of cAMP in the environment. The LacY effect is well known for induction with lactose in the growth medium ^{16,17}. However, the PDE effect observed here is previously unknown. It may yield important clues towards a comprehensive understanding of cAMP transport and control central to carbon metabolism in *E. coli*.

1.9 DNA looping is assisted by CRP

The innate regulation of *Plac* activity is revealed in *E. coli* cells bearing *cyaA lacY cpdA* triple deletion. We found that with a saturating amount of IPTG in the medium, the cAMP-CRP-mediated activation behaved in accordance with the thermodynamic model, given the non-cooperative nature of the cAMP-CRP interaction.

Also, in the absence of cAMP, the LacR-mediated repression behaved in accordance with the thermodynamic model with DNA looping. Moreover, the co-dependence on IPTG and cAMP can be explained by the extended thermodynamic model with a single cooperativity parameter Ω whose value is of the order found by *in vitro* experiments. Both the quality of the fit and the simplicity of the model (i.e., one parameter for co-dependence) are superior compared to other simple but *ad-hoc* models, as we show explicitly in Chapter 2.10, *Comparison to alternative models*, (Figs. 2.3). These results strongly support the hypothesis of CRP-assisted DNA looping^{29,31}, and also shed light on the seemingly conflicting roles CRP plays in the regulation of the *lac* promoter: CRP is an activator in that it increases gene expression at each IPTG level. However it increases gene expression in a nonlinear way such that the maximal fold-change and the sensitivity of the IPTG response are both enhanced.

1.10 Hyper-sensitivity of IPTG response resulted from multiple factors

Even in the absence of *lacY* mediated feedback, *Plac* activity depended sensitively on the IPTG level with apparent Hill coefficient of $2.5 \sim 3$ at physiological intracellular cAMP concentrations. This is rather remarkable given that the basic interaction of IPTG with the Lac tetramer is not cooperative at all in the absence of operator DNA²⁴. While the involvement of DNA looping in the IPTG hyper-sensitivity is shown qualitatively by the experiment of Oehler et al⁶⁴, the results and analysis of this study reveal the complex way in which this hyper-sensitivity is attained: First, the non-cooperative IPTG-LacR binding becomes weakly cooperative (effective Hill coefficient $\sim$

1.5) in the presence of operator DNA [Ref. 22-25 and Fig. 1.6], due presumably to the allosteric coupling between inducer and operator binding (see Chapter 2.4, *Activation of the regulatory proteins*). The involvement of DNA looping then increases the cooperativity (effective Hill coefficient ~ 2), as both dimeric units of the Lac tetramer are required to bind specifically to operators in order to form DNA loops. Finally, in the physiological range of intracellular cAMP concentrations where CRP is activated, the cooperative interplay between DNA looping and CRP-mediated DNA bending further enhances the looping-induced cooperativity to the observed effective Hill coefficient of $2.5 \sim 3$.

As shown by Novick & Weiner and by Ozbudak et al ^{16,17}, the hyper-sensitivity of *Plac* activity to the inducer of LacR is crucial for the bistable (i.e., the all-or-none) nature of *E. coli*'s lactose utilization strategy ^{16,17}. The latter is the molecular basis of important physiological effects such as the diauxie shift ⁷². It is interesting that hyper-sensitivity resulted in this case from an agglomeration of weak interactions. This distributed way of implementing an important molecular function (instead of, e.g., a highly cooperative LacR-inducer interaction) may reflect the serendipity of the evolutionary dynamics, or alternatively a robust evolutionary strategy to preserve important system-level functions.

1.11 The importance of being quantitative

Critical to this study was the recognition of various “artifacts” produced by processes not directly related to the regulation of the *lac* promoter. The key discriminating feature we used was the quantitative comparison of the observed

responses to the expectations of the thermodynamic model, in light of the known biochemical processes. Although fitting the experimental data to a Hill form is in itself not a very discriminating task, we find that the “reasonableness” of the numerical values of the fitting parameters, especially the apparent Hill coefficients of the responses, can be quite revealing. Nevertheless, the power of the quantitative analysis is by itself limited, as it can only suggest the existence of problems but does not identify the sources. It is through the quantitative comparison of the characteristics of a series of key mutants that the major conclusions of this study are established. It is important to note that some of our most discriminating mutants did not display much difference in the high/low states of expression, but exhibited clear differences in the abruptness of the transition between states.

The *lac* promoter is one of the prototypical model systems of gene regulation. Classic studies on the regulation of this promoter have established numerous fundamental concepts as well as laying down the appropriate methodology for studying the molecular biology of gene regulation. Quantitative studies of the *lac* promoter can again play important roles in laying down the foundation of quantitative systems biology, whose goal is to understand the behavior of a “system” in terms of the relevant properties of its components. The *lac* promoter is admittedly a rather simple system. Nevertheless, we see from this study that system-level properties such as the sensitivity of the IPTG response resulted from a closely intertwined set of interactions among the molecular constituents. We demonstrated how this system can be dissected by careful quantitative characterization and targeted genetic manipulations, along with guidance from quantitative modeling and the knowledge of the biochemistry of the molecular

constituents. The experience gained here may be of value to the study of other more complex biological systems.

1.12 Acknowledgement

We are indebted to Bob Schleif and Bill Loomis for pointing out conceptual flaws in early versions of our experiments. We also benefited from valuable comments from many other colleagues throughout the course of this work, including Nicolas Buchler, Antoine Danchin, Hernan Garcia, Kathleen Matthews, Stefan Oehler, Rob Philips, and Agnes Ullmann. This research was supported by the DOE (Grant No. DE-FG02-03-ER63691), the NSF (Grant No. MCB-0417721), and by the NIH (Grant No. GM070915). TK and TH additionally acknowledge support through the NSF PFC-sponsored Center for Theoretical Biological Physics (Grants No. PHY-0216576 and PHY-0225630).

Chapter 1 contains material in Kuhlman T, Zhang Z, Saier MH, Jr., Hwa T. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc Natl Acad Sci* 2007; 104(14): 6043-6048., used with permission. The dissertation author was the primary investigator and author of this work.

1.13 References

1. Monod J. The phenomenon of enzymatic adaptation. *Growth*. 1947;11:223-89.
2. Beckwith J. The lactose operon. In: Neihardt JL, Ingraham KB, Low B, Magasanik B, Umberger HE (eds). *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. Volume 2. Washington, D.C.: American Society for Microbiology, 1987:1439-43.
3. Beckwith J, Grodzicker T, Arditti R. Evidence for two sites in the lac promoter region. *J Mol Biol*. 1972 Aug 14;69(1):155-60.
4. Gilbert W, Muller-Hill B. The Lactose Repressor. In: Beckwith J, Zipser D (eds). *The Lactose Operon*. Cold Spring Harbor NJ: Cold Spring Harbor Press, 1970:93-109.
5. Muller-Hill B. *The lac Operon*. Berlin: de Gruyter; 1996.
6. Oehler S, Amouyal M, Kokhof P, von Wilcken-Bergmann B, Muller-Hill B. Quality and position of three lac operators of *E. coli* define efficiency of repression. *EMBO*. 1994;13(14):973-9.
7. Davidson EH, Rast JP, Oliveri P, et al. A genomic regulatory network for development. *Science*. 2002 Mar 1;295(5560):1669-78.
8. Buchler NE, Gerland U, Hwa T. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A*. 2003 Apr 29;100(9):5136-41.
9. Epstein W, Rothman-Denes LB, Hesse J. Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 1975 Jun;72(6):2300-4.
10. Saier MH, Jr. Protein phosphorylation and allosteric control of inducer exclusion and catabolite repression by the bacterial phosphoenolpyruvate: sugar phosphotransferase system. *Microbiol Rev*. 1989 Mar;53(1):109-20.
11. Saier MH, Jr., Feucht BU, McCaman MT. Regulation of intracellular adenosine cyclic 3':5'-monophosphate levels in *Escherichia coli* and *Salmonella typhimurium*. Evidence for energy-dependent excretion of the cyclic nucleotide. *J Biol Chem*. 1975 Oct 10;250(19):7593-601.
12. Aboud M, Burger M. The effect of catabolite repression and of cyclic 3',5' adenosine monophosphate on the translation of the lactose messenger RNA in *Escherichia coli*. *Biochem Biophys Res Commun*. 1970 Mar 27;38(6):1023-32.

13. Contesse M, Crepin M, Gros F, Ullmann A, Monod J. On the mechanism of catabolite repression. In: Beckwith JR, Zipser D (eds). *The Lactose Operon*. Cold Spring Harbor, NJ: Cold Spring Harbor Press, 1970:401-16.
14. Pastan I, Perlman R. Cyclic adenosine monophosphate in bacteria. *Science*. 1970 Jul 24;169(943):339-44.
15. Makman RS, Sutherland EW. Adenosine 3',5'-Phosphate in *Escherichia Coli*. *J Biol Chem*. 1965 Mar;240:1309-14.
16. Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci U S A*. 1957;43:553-66.
17. Ozbudak EM, Thattai M, Lim HN, Shraiman BI, Van Oudenaarden A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*. 2004 Feb 19;427(6976):737-40.
18. Eismann ER, Muller-Hill B. lac repressor forms stable loops in vitro with supercoiled wild-type lac DNA containing all three natural lac operators. *J Mol Biol*. 1990 Jun 20;213(4):763-75.
19. Muller J, Oehler S, Muller-Hill B. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol*. 1996 Mar 22;257(1):21-9.
20. Oehler S, Eismann ER, Kramer H, Muller-Hill B. The three operators of the lac operon cooperate in repression. *EMBO*. 1990;9(4):973-9.
21. Vilar JM, Leibler S. DNA looping and physical constraints on transcription regulation. *J Mol Biol*. 2003 Aug 29;331(5):981-9.
22. Friedman BE, Olson JS, Matthews KS. Interaction of lac repressor with inducer, kinetic and equilibrium measurements. *J Mol Biol*. 1977 Mar 25;111(1):27-39.
23. Chen J, Alberti S, Matthews KS. Wild-type operator binding and altered cooperativity for inducer binding of lac repressor dimer mutant R3. *J Biol Chem*. 1994 Apr 29;269(17):12482-7.
24. O'Gorman RB, Rosenberg JM, Kallai OB, et al. Equilibrium binding of inducer to lac repressor.operator DNA complex. *J Biol Chem*. 1980 Nov 10;255(21):10107-14.
25. Daly TJ, Matthews KS. Allosteric regulation of inducer and operator binding to the lactose repressor. *Biochemistry*. 1986 Sep 23;25(19):5479-84.

26. Yagil G, Yagil E. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys J*. 1971 Jan;11(1):11-27.
27. Setty Y, Mayo AE, Surette MG, Alon U. Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A*. 2003 Jun 24;100(13):7702-7.
28. Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). *J Mol Biol*. 1999 Oct 22;293(2):199-213.
29. Lewis M, Chang G, Horton NC, et al. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*. 1996 Mar 1;271(5253):1247-54.
30. Balaeff A, Mahadevan L, Schulten K. Structural basis for cooperative DNA binding by CAP and Lac repressor. *Structure*. 2004;12(1):123-32.
31. Fried MG, Hudson JM. DNA looping and lac repressor-CAP interaction. *Science*. 1996 Dec 13;274(5294):1930-1; author reply 1-2.
32. Hudson JM, Fried MG. Co-operative interactions between the catabolite gene activator protein and the lac repressor at the lactose promoter. *J Mol Biol*. 1990 Jul 20;214(2):381-96.
33. Vossen KM, Stickle DF, Fried MG. The mechanism of CAP-lac repressor binding cooperativity at the *E. coli* lactose promoter. *J Mol Biol*. 1996 Jan 12;255(1):44-54.
34. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev*. 2005 Apr;15(2):125-35.
35. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*. 2005 Apr;15(2):116-24.
36. Ackers GK, Johnson AD, Shea MA. Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci U S A*. 1982 Feb;79(4):1129-33.
37. Monod J. Remarks on the mechanism of enzyme induction. *Enzymes: Units of Biological Structure and Function*, Henry Ford Hospital International Symposium, Detroit. New York: Academic Press, 1956:7-28.
38. Zubay G, Schwartz D, Beckwith J. Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proc Natl Acad Sci U S A*. 1970 May;66(1):104-10.

39. Eismann E, von Wilcken-Bergmann B, Muller-Hill B. Specific destruction of the second lac operator decreases repression of the lac operon in *Escherichia coli* fivefold. *J Mol Biol.* 1987 Jun 20;195(4):949-52.
40. Jensen PR, Westerhoff HV, Michelsen O. The use of lac-type promoters in control analysis. *Eur J Biochem.* 1993 Jan 15;211(1-2):181-91.
41. Brana H. Adenyl cyclase in cell-free extracts of *Escherichia coli*. *Folia Microbiol (Praha).* 1969;14(3):185-9.
42. Ide M. Adenyl cyclase of *Escherichia coli*. *Biochem Biophys Res Commun.* 1969 Jul 7;36(1):42-6.
43. Perlman RL, Pastan I. Pleiotropic deficiency of carbohydrate utilization in an adenyl cyclase deficient mutant of *Escherichia coli*. *Biochem Biophys Res Commun.* 1969 Sep 24;37(1):151-7.
44. Tao M, Lipmann F. Isolation of adenyl cyclase from *Escherichia coli*. *Proc Natl Acad Sci U S A.* 1969 May;63(1):86-92.
45. Magasanik B. Glucose effects: inducer exclusion and repression. In: Beckwith J, Zipser D (eds). *The Lactose Operon*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1970.
46. Magasanik B, Neidhardt F. Regulation of carbon and nitrogen utilization. In: Neidhardt F (ed). *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. Volume 2. Washington, D.C.: American Society for Microbiology, 1987.
47. Ullmann A, Danchin A. Role of cyclic AMP in bacteria. *Adv Cyclic Nucleotide Res.* 1983;15:1-53.
48. Crasnier-Mednansky M, Park MC, Studley WK, Saier MH, Jr. Cra-mediated regulation of *Escherichia coli* adenylate cyclase. *Microbiology.* 1997 Mar;143 (Pt 3):785-92.
49. Krin E, Sismeiro O, Danchin A, Bertin PN. The regulation of Enzyme IIA(Glc) expression controls adenylate cyclase activity in *Escherichia coli*. *Microbiology.* 2002 May;148(Pt 5):1553-9.
50. Takahashi M, Blazy B, Baudras A, Hillen W. Ligand-modulated binding of a gene regulatory protein to DNA. Quantitative analysis of cyclic-AMP induced binding of CRP from *Escherichia coli* to non-specific and specific DNA targets. *J Mol Biol.* 1989 Jun 20;207(4):783-96.

51. Fried MG, Crothers DM. Equilibrium studies of the cyclic AMP receptor protein-DNA interaction. *J Mol Biol.* 1984 Jan 25;172(3):241-62.
52. Garner MM, Revzin A. Stoichiometry of catabolite activator protein/adenosine cyclic 3',5'-monophosphate interactions at the lac promoter of *Escherichia coli*. *Biochemistry.* 1982 Nov 23;21(24):6032-6.
53. Harman JG. Allosteric regulation of the cAMP receptor protein. *Biochim Biophys Acta.* 2001 May 5;1547(1):1-17.
54. Aboud M, Burger M. Cyclic 3',5' adenosine monophosphate-phosphodiesterase and the release of catabolite repression of beta-galactosidase by exogenous cyclic 3',5' adenosine monophosphate in *Escherichia coli*. *Biochem Biophys Res Commun.* 1971 Apr 2;43(1):174-82.
55. Brana H, Chytil F. Splitting of the cyclic 3',5'-adenosine monophosphate in a cell-free system of *Escherichia coli*. *Folia Microbiol.* 1965;11:43-6.
56. Monard D, Janecek J, Rickenberg HV. The enzymic degradation of 3',5' cyclic AMP in strains of *E. coli* sensitive and resistant to catabolite repression. *Biochem Biophys Res Commun.* 1969 May 22;35(4):584-91.
57. Nielsen LD, Monard D, Rickenberg HV. Cyclic 3',5'-adenosine monophosphate phosphodiesterase of *Escherichia coli*. *J Bacteriol.* 1973 Nov;116(2):857-66.
58. Nielsen LD, Rickenberg HV. Cyclic AMP phosphodiesterase of *Escherichia coli*. *Methods Enzymol.* 1974;38:249-56.
59. Imamura R, Yamanaka K, Ogura T, et al. Identification of the *cpdA* gene encoding cyclic 3',5'-adenosine monophosphate phosphodiesterase in *Escherichia coli*. *J Biol Chem.* 1996 Oct 11;271(41):25423-9.
60. Gerland U, Moroz JD, Hwa T. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc Natl Acad Sci U S A.* 2002 Sep 17;99(19):12015-20.
61. Straney DC, Straney SB, Crothers DM. Synergy between *Escherichia coli* CAP protein and RNA polymerase in the lac promoter open complex. *J Mol Biol.* 1989 Mar 5;206(1):41-57.
62. Malan TP, Kolb A, Buc H, McClure WR. Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. *J Mol Biol.* 1984 Dec 25;180(4):881-909.

63. Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev.* 2005 Apr;15(2):136-44.
64. Oehler S, Alberti S, Muller-Hill B. Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Res.* 2006;34(2):606-12.
65. Nichols JC, Matthews KS. Combinatorial mutations of lac repressor. Stability of monomer-monomer interface is increased by apolar substitution at position 84. *J Biol Chem.* 1997 Jul 25;272(30):18550-7.
66. Jobe A, Bourgeois S. lac Repressor-operator interaction. VI. The natural inducer of the lac operon. *J Mol Biol.* 1972 Aug 28;69(3):397-408.
67. Fried MG, Daugherty MA. In vitro interaction of the Escherichia coli cyclic AMP receptor protein with the lactose repressor. *J Biol Chem.* 2001 Apr 6;276(14):11226-9.
68. Perros M, Steitz TA. DNA looping and lac repressor-CAP interaction. *Science.* 1996 Dec 13;274(5294):1929-30; author reply 31-2.
69. Borowiec JA, Zhang L, Sasse-Dwight S, Gralla JD. DNA supercoiling promotes formation of a bent repression loop in lac DNA. *J Mol Biol.* 1987 Jul 5;196(1):101-11.
70. de Crombrughe B, Busby S, Buc H. Cyclic AMP receptor protein: role in transcription activation. *Science.* 1984 May 25;224(4651):831-8.
71. Becker NA, Kahn JD, Maher LJ, 3rd. Bacterial repression loops require enhanced DNA flexibility. *J Mol Biol.* 2005 Jun 17;349(4):716-30.
72. Monod J. *Récherches sur la croissance des cultures bactériennes.* Paris: Hermann et Compagnie Ed., 1942.
73. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A.* 2000 Jun 6;97(12):6640-5.

CHAPTER 2

Combinatorial Transcriptional Control of the Lactose Operon of *Escherichia coli*:

Experimental and Theoretical Methods

2.1 Plasmids and Strains

E. coli K-12 MG1655 and derived mutants were used in all experiments reported in Chapter 1. Chromosomal deletions were performed using the method of Datsenko & Wanner¹. For each deletion the kanamycin resistance gene *kan* was amplified from the template plasmid pKD4 using primer sites P0 and P2 including 45 bp homology extensions. The PCR products were then transformed into electrocompetent MG1655 cells, and *kan* was eliminated using plasmid pCP20 bearing FLP recombinase¹. The resulting strains TK110, TK120, TK130, TK140, and TK150 are listed in Table 1.1 together with the start and end coordinates of the respective deletions. All mutations were verified using PCR. Multiple deletion strains were constructed by serially transferring and eliminating the TK100 series of *kan* insertion mutants into the target strain using P1 transduction in the order indicated in column 3 of Table 1.1, with the exception of strain TK200 carrying $\Delta lacI \Delta lacY$ double mutation. The close proximity of *lacI* to *lacY* necessitated the *de novo* synthesis of both mutations in the same host strain.

Measurements of the variation of CRP levels were performed on strain TK500. This strain was derived from strain TK400, which was derived in turn from *E. coli* BW25113¹, with *cyaA* deletion performed as explained above. The *crp* promoter *Pcrp* was amplified from MG1655 genomic DNA from start coordinate 3483424 to end

coordinate 3483721 and inserted into the EcoRI and BamHI sites of PRS551². The resulting *Pcrp-lacZ* fusion was inserted into the chromosome of TK400 using the method of Simons *et al.*². Measurements of the effect of the CRP operator on the *lac* promoter were performed using strain TK600 which is also derived from strain TK400. The *lac* promoter was amplified from start coordinate 365438 to end coordinate 365669 and cloned into the EcoRI and BamHI sites of pRS551. The CRP operator site at position -61.5 was altered from TGTGAGTTAGCTCACT to CAGACGTTAGCTCACT using site-directed mutagenesis (Stratagene) and the resulting construct was inserted into the chromosome of strain TK400 using the method of Simons *et al.*².

2.2 β -galactosidase assay

Assays of β -galactosidase activity were performed according to Miller³ and Griffith and Wolf⁴ with changes as follows. Overnight cultures of bacterial strains were grown in 3 ml M9 minimal medium (M9, 1 mM MgSO₄, 0.1 mM CaCl₂) containing 0.5% glucose as the sole carbon source and the standard concentrations of any necessary antibiotics in a 37°C water bath until saturation. This culture was diluted[§] into 24-well plates (Costar) containing 1 ml M9 medium in each well with glucose, antibiotics, and varying concentrations of 3',5'-cyclic adenosine monophosphate (cAMP) (0-10 mM) and isopropyl β -D thiogalactopyranoside (IPTG) (0-1 mM). The plates were then grown with

[§] In M9 medium, the growth rate of $\Delta cyaA$ strains increases from 0.4 doubling/hour at 0 cAMP to a maximum of 0.6 doubling/hour at ~1 mM cAMP; see Fig. 1.3a. As such, the initial inoculation of samples was performed at different dilutions (~1000x dilution for samples in 1 mM cAMP and ~100x dilution for samples at lower cAMP concentrations) such that all samples would reach OD₆₀₀ = 0.2-0.4 in

vigorous shaking in a humidity-controlled incubator, with OD₆₀₀ measurements taken every two hours in a Tecan Genios Pro plate reader. The cell-doubling rate ($\lambda_{1/2}$) for each sample was calculated as the slope of $\log_2(\text{OD}_{600})$ vs. time plot via linear regression analysis.

When OD₆₀₀ of a sample reaches 0.2-0.4, 0.5 ml of the sample was transferred to a 2-ml 96 well polypropylene block containing 0.5 ml Z-buffer³, 20 μl 0.1% SDS, and 40 μl chloroform. All samples were thoroughly disrupted by repeated agitation with a multichannel pipettor. 200 μl of each sample was transferred to a flat-bottom transparent 96 well plate (Costar). 40 μl of Z-buffer containing 4 mg/ml *o*-nitrophenyl- β -D-galactopyranoside (ONPG) was added to each well and OD₄₂₀ measurements were performed in a Tecan Genios Pro plate reader, at 1-minute intervals for the first 10 minutes, and at increasing intervals (1 hour, 5 hour, etc.) thereafter** until sufficient color had developed in all samples for measurement, and the measured OD₄₂₀ showed an extended regime of linear dependence on time (data not shown). Samples were maintained at 25°C throughout incubation. All assays were performed in triplicate or more.

2.3 Promoter activity

In the regime where the OD₄₂₀ readings were sufficiently large to ensure a linear dependence on time of measurements but below 1.0 (above which the OD₄₂₀ readings

approximately 12-15 hours. We verified that samples assayed with various initial dilutions in the given range gave no measurable systematic variations in results (data not shown).

saturate), the slope (s) of OD_{420} versus time (in minutes) was calculated via linear regression. β -galactosidase activity (A) was expressed in Miller Units (MU) according to the formula $A = (1000 \cdot s) / (0.5 \cdot OD_{600})$. To verify the reliability of the deduced activity for the weakly expressed promoters, we performed a serial dilution experiment, using MG1655 grown in M9 medium with 0.5% glucose and 1 mM IPTG, which yielded a β -galactosidase activity of ~ 1000 MU. We mixed this strain with strain BW25113, deleted of the *lac* operon, in varying proportions (up to 10^6 -fold dilution), and measured the apparent β -galactosidase activity of the mixture using the procedure specified above. We found a nearly perfect inverse dependence of the observed β -galactosidase activity with the fold-dilution applied down to ~ 0.03 MU (Fig. 2.1). As the lowest β -galactosidase activity encountered in the experiments was ~ 0.1 MU, our data lay completely in the responsive regime of the measurements.

In the text, we report the “promoter activity” (α) as

$$\alpha = A \cdot \lambda_{\frac{1}{2}}, \quad [2.3.1]$$

where $\lambda_{\frac{1}{2}}$ is the cell-doubling rate defined above (in unit of 1/hr). This measure of the promoter activity is motivated by the fact that the enzyme β -galactosidase is very stable⁵ so that in the balanced exponential growth phase, its “turnover” is governed by dilution due to cell growth; see discussion below. Note that with the definition Eq. [2.3.1], we take into account some straightforward growth-dependent effects on β -galactosidase activity as previously reported^{6,7}. There are of course other residual effects, e.g.,

** For the lowest expression encountered this took up to 50 hours; see Fig. 2.1.

variation in the cellular levels of RNA polymerase, ribosomes, etc at the different growth rates. However, these effects are apparently quite limited in magnitude and will be neglected in our analysis: As seen in Fig. 1.3*b* (triangles), the growth-rate adjusted activity of β -galactosidase expressed by the (constitutive) *Plac Δ crp* promoter in *cyaA*-cells changed very little in growth medium with various cAMP concentrations, even though the growth rate itself changed appreciably over this range of cAMP concentrations (Fig. 1.3*a*).

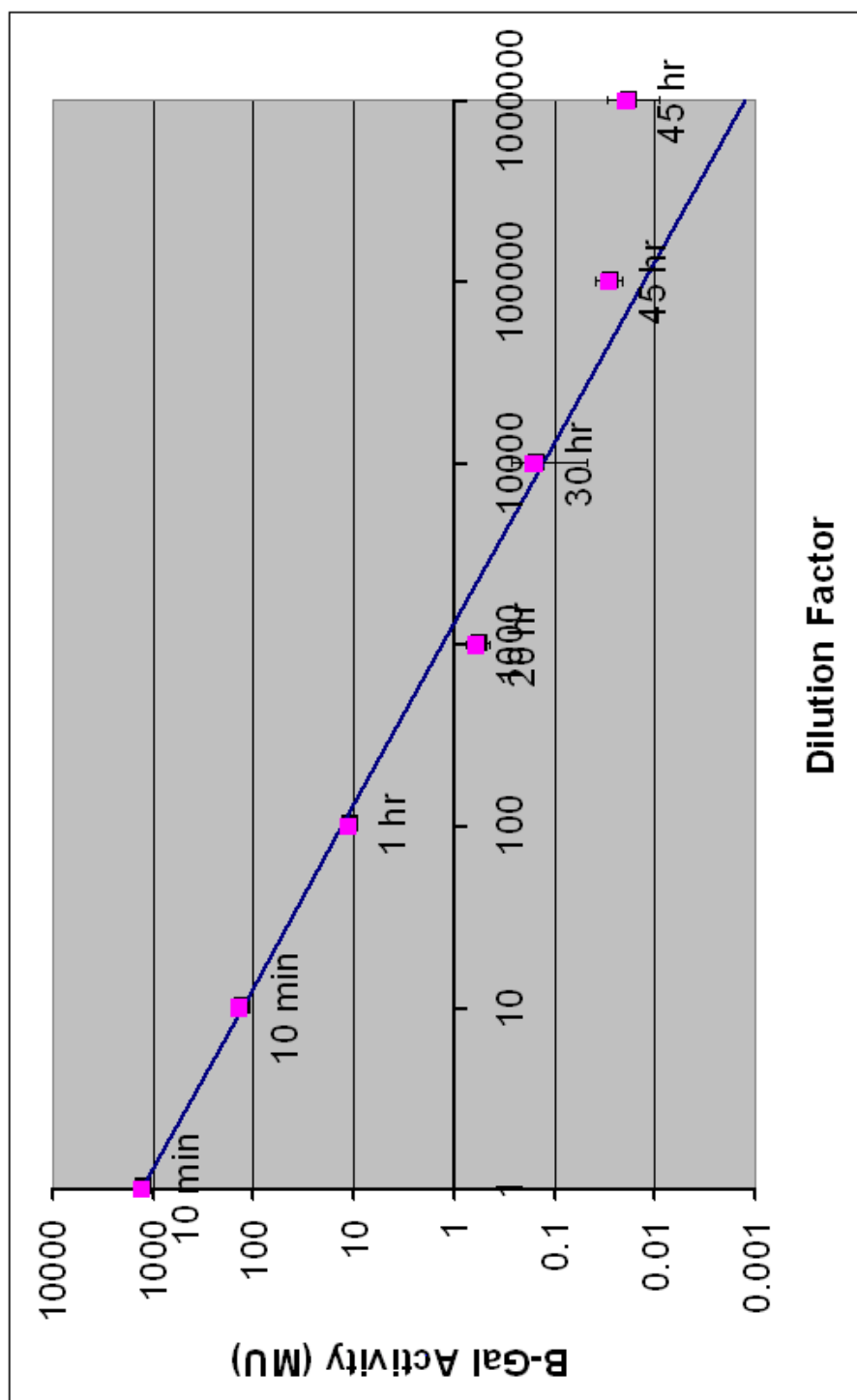


Figure 2.1: Determination of the limit of β -Gal activity detection. A fully induced culture of MG1655 was serially diluted 10 fold at a time to determine the lower limit of detection. The measurements (magenta points) are seen to agree well with the expected activity (blue line) down to ~ 0.05 Miller Units. The reaction incubation time necessary to allow sufficient color to develop for measurement is indicated below each data point.

2.4 Activation of the regulatory proteins

The concentrations of the active regulators are controlled by the total concentrations of the regulators, the intracellular inducer concentrations, and the biochemistry of inducer-regulator interaction. The interaction of CRP with cAMP is quite straightforward⁸⁻¹¹ and will be discussed first. The active component is the CRP dimer associated with one cAMP molecule¹⁰⁻¹². For simplicity, we will assume that all the CRP monomers associate in the dimer form, as justified by the very small dimer dissociation constant [0.1 – 1 nM; ¹³] The concentration of the activated CRP dimer (denoted by $[CRP^*]$) is then given by

$$[CRP^*] \equiv [cAMP - CRP] = [CRP] \cdot \frac{[cAMP]_{in}^*}{[cAMP]_{in}^* + K_{cAMP}}, \quad [2.4.1]$$

where $[CRP]$ denotes the total CRP dimer concentration, $K_{cAMP} = 10 \mu\text{M}$ is the relevant dissociation constant¹⁰, and $[cAMP]_{in}^*$ is the intracellular concentration of cAMP. Intracellular cAMP is believed to be rapidly exported by PMF pumps^{14,15}. For *cyaA* strains which cannot synthesize cAMP endogenously, the high export/import ratio leads to a much smaller intracellular cAMP concentration compared to the extracellular concentration. It is this intracellular cAMP concentration which dictates the degree of CRP activation (assuming that cAMP-CRP equilibration is much faster than cAMP transport.) The details of cAMP transport are not understood; even the cAMP pump has not been identified despite many years of study. Here, we appeal to the results of Epstein *et al.*¹⁶ and assume that the intracellular cAMP concentration is linearly related to the extracellular concentration across the range of concentrations used, i.e.,

$[cAMP]_{in}^* = \gamma_{cAMP} \cdot [cAMP]$, where $[cAMP]$ is the cAMP level of the medium. Since $\gamma_{cAMP} \approx 10^{-3}$ according to ref. 16, we have $[cAMP]_{in}^* = \gamma_{cAMP} \cdot [cAMP] = K_{cAMP}$ for most of the range of cAMP used in our experiment (0 – 10 mM). Hence Eq. [2.4.1] is simplified to

$$[CRP^*] \approx [CRP] \cdot [cAMP] \cdot \gamma_{cAMP} / K_{cAMP} . \quad [2.4.2]$$

Obtaining the active concentration of the Lac repressor is more involved. The Lac repressor is a dimer of dimers ¹⁷. Co-crystal structure of LacR and operator DNA indicates that the dimer binds specifically to the operator sequence as a unit in the absence of IPTG, while the IPTG-bound dimers do not have the required structure to bind specifically to operator DNA. Additionally, since the two dimeric units of the LacR tetramer interact very weakly ¹⁸, we do not expect the IPTG binding by one dimeric unit to affect the specific DNA binding of the other dimeric unit. Finally, we do not expect a dimeric unit to be able to bind specifically to operator sites if either one of its monomeric subunit is bound to IPTG. This assumption is based on the work of Winter, Berg, and von Hippel ¹⁹, who showed that sequence specific binding would only occur if the energy gained from sequence-specific binding more than compensates the loss of electrostatic energy the repressor molecule could gain in a conformation that does not allow specific DNA binding. When one of the monomers bind to IPTG, that monomer could not contribute to specific DNA binding, and the other monomer would not be able to contribute enough to compensate given what we know about the energetics of LacR-DNA binding ¹⁹. The last feature described above, that both monomers of a dimeric unit need to be free of IPTG binding in order for the dimeric unit to bind specifically to

operator DNA, is a crucial source of cooperativity in our model. While there is no direct biochemical proof of this feature, a MWC model incorporating this feature of LacR-DNA binding is able to reproduce the effective Hill coefficient of $1.4 \sim 1.6$ which describes the IPTG dependence of specific operator binding²⁰. It then follows that there are two species of “active repressors” in our model: (i) Only one of the two dimeric units is not bound at all by IPTG (denoted by LacR*), and (ii) neither of the dimeric units is bound by IPTG (denoted by LacR**).

To find the concentrations of these two active species (denoted by $[LacR^*]$ and $[LacR^{**}]$ respectively) in terms of the total LacR tetramer concentration $[LacR]$ and the IPTG concentration $[IPTG]$, we need to keep track of all possible molecular species of LacR-IPTG complex. As illustrated in Fig. 2.2, there is one species with no IPTG (denoted by LacR-0), four sub-species with one IPTG bound (LacR-1), two sub-species with two IPTG bound to a single dimer (LacR-2a), four sub-species with two IPTG bound to two different dimers (LacR-2b), four subspecies with three IPTG bound (LacR-3) and one subspecies with four IPTG bound (LacR-4). Since the IPTG-LacR binding is non-cooperative²⁰, the equilibrium between any two subspecies linked by a line in Fig. 2.2 is described by the same dissociation constant K_{IPTG} . Taking into account of the indistinguishability of the subspecies, we arrive at the following equilibrium relations among the species of LacR-IPTG complex:

$$\frac{[LacR-1]}{[LacR-0]} = \frac{4[IPTG]}{K_{IPTG}}, \quad \frac{[LacR-2a]}{[LacR-0]} = 2 \cdot \left(\frac{[IPTG]}{K_{IPTG}} \right)^2; \quad \frac{[LacR-2b]}{[LacR-0]} = 4 \cdot \left(\frac{[IPTG]}{K_{IPTG}} \right)^2;$$

$$\frac{[LacR-3]}{[LacR-0]} = 4 \cdot \left(\frac{[IPTG]}{K_{IPTG}} \right)^3; \quad \frac{[LacR-4]}{[LacR-0]} = \left(\frac{[IPTG]}{K_{IPTG}} \right)^4.$$

Since the total LacR concentration is the sum of the concentrations of the individual species, we have

$$[LacR^{**}] \equiv [LacR-0] = [LacR] / \left(1 + [IPTG] / K_{IPTG} \right)^4 \quad [2.4.3]$$

and

$$\begin{aligned} [LacR^*] &\equiv [LacR-1] + [LacR-2a] \\ &= [LacR] \cdot \frac{4([IPTG] / K_{IPTG}) + 2([IPTG] / K_{IPTG})^2}{(1 + [IPTG] / K_{IPTG})^4}. \end{aligned} \quad [2.4.4]$$

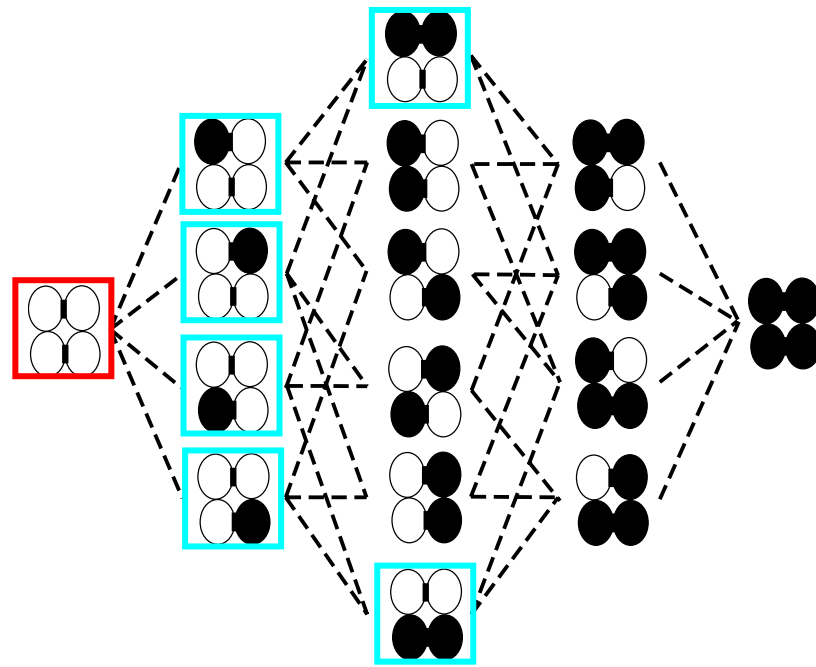


Figure 2.2: Different states of LacR-IPTG binding. The ellipses represent the LacR monomers. Two monomers form a dimeric unit (joined by the thick solid line). Two dimers form a tetramer. The open ellipses indicate LacR monomers not bound to IPTG; filled ellipses indicate LacR monomers bound to IPTG. Each line indicates an allowed direct transition between two states. From left to right, the different columns represent subspecies denoted as LacR-0, LacR-1, LacR-2, LacR-3, and LacR-4, respectively. The active species LacR** and LacR* are indicated by the red and cyan boxes respectively.

2.5 Thermodynamic modeling

The transcriptional activity of a given promoter controlled by the binding and interaction of transcription factors (TF) can be modeled using a generalized Shea-Ackers like thermodynamic model ²¹ as described in Buchler *et al.* ²² and Bintu *et al.* ^{23,24}. Briefly, we assume that the rate of protein expression, G , is proportional to the equilibrium probability P of the RNA polymerase (RNAP) binding to its DNA target, the promoter, given the cellular concentrations of all of the relevant transcript factors. The probability P can be written as

$$P = \frac{Z_{ON}}{Z_{ON} + Z_{OFF}},$$

where Z_{ON} and Z_{OFF} are the partition sums of the Boltzmann weights over all states of transcription factor binding for the promoter bound and not bound, respectively, by the RNA polymerase. Given the knowledge of the cellular concentrations of the TFs, the partition function of the TFs themselves (Z_{OFF}) is simply

$$Z_{OFF} = \sum_{all \ \sigma_i = \{0,1\}} \prod_{(i,j)} \left(\frac{[TF_{s(i)}]}{K_i} \right)^{\sigma_i} \cdot \left(\frac{[TF_{s(j)}]}{K_j} \right)^{\sigma_j} \cdot \omega_{i,j}^{\sigma_i \sigma_j}, \quad [2.5.1]$$

where K_i refers to the dissociation constant of operator site i with its TF [of type s as indicated by subscript $s(i)$], and $\omega_{i,j}$ is the cooperativity factor between the two TFs bound to operator sites i and j . It is important to note that K_i is the effective *in vivo* binding constant including the effect of binding of its TF to all other possible binding sites on the chromosome, and is typically not the value of *in vitro* measurement ²⁵. The values of the $\omega_{i,j}$ depend on the spacing between sites i and j . If two sites are

overlapping, $\omega_{i,j} = 0$, reflecting the fact that they cannot be simultaneously occupied. If two sites are positioned such that the two bound TFs cannot contact each other, then $\omega_{i,j} = 1$, indicating the lack of interactions between the bound TFs. For two TFs that can bind cooperatively and their binding sites are positioned such that the bound TFs do interact, then the value of $\omega_{i,j}$ is the cooperativity factor of that interaction, and is typically of the order 10~100²². The expression for Z_{ON} is similar to [2.5.1], but with one of the sites being the promoter, and one of the [TF_s] being [RNAP].

As an example, consider a promoter with a single binding site for an activator protein A , which binds to its site with dissociation constant K_A and interacts with RNAP via a cooperativity factor $\omega > 1$. The equilibrium promoter occupation probability P can then be written as

$$P = \frac{[RNAP]/K_{RNAP} \cdot (1 + \omega \cdot [A]/K_A)}{1 + [A]/K_A + [RNAP]/K_{RNAP} \cdot (1 + \omega \cdot [A]/K_A)} \simeq \frac{[RNAP]}{K_{RNAP}} \cdot \frac{1 + \omega \cdot [A]/K_A}{1 + [A]/K_A}$$

where the approximation above is justified by the fact that typical core promoters are very weak such that $[RNAP]/K_{RNAP} = 1$ ²²⁻²⁴. The rate of protein expression, $G \propto P$, can then be written as

$$G \approx G_0 \cdot \frac{1 + \omega \cdot [A]/K_A}{1 + [A]/K_A}. \quad [2.5.2]$$

where G_0 is a basal protein expression rate, involving the rate of transcription, translation as well as mRNA turnover, in the absence of any regulatory proteins. It is not a number computable from the thermodynamic model; on the other hand, the knowledge of this

overall prefactor is not necessary for characterizing the relative effect of transcriptional control.

In the case of a repressor protein R which binds to its operator with dissociation constant K_R and precludes RNAP binding at the core promoter, the transcription rate can be written as

$$P = \frac{[RNAP] / K_{RNAP}}{1 + [R] / K_R + [RNAP] / K_{RNAP}} ; \frac{[RNAP]}{K_{RNAP}} \cdot \frac{1}{1 + [R] / K_R}$$

and the rate of protein expression as

$$G \approx G_0 \cdot \frac{1}{1 + [R] / K_R} . \quad [2.5.3]$$

For a promoter which is controlled by both a repressor and an activator, the simplest case would be that the two exert their influence on the RNAP independently and do not interact with each other. In this case, the combined response would simply be the product of the two expressions given above, i.e.,

$$G \approx G_0 \cdot \left(\frac{1 + \omega \cdot [A] / K_A}{1 + [A] / K_A} \right) \cdot \left(\frac{1}{1 + [R] / K_R} \right) . \quad [2.5.4]$$

Below, we apply the general thermodynamic to the *lac* promoter.

2.6 Activation by cAMP-CRP

We first address activation by cAMP-CRP, whose concentration is $[CRP^*]$ as given by Eq. [2.4.2]. cAMP-CRP binds to the operator site at position -61.5 (see Fig. 1.7) and recruits the RNA polymerase to the core promoter. In the absence of LacR, this is

just the case of “simple activation” considered in Eq. [2.5.2]. Denoting the *in vivo* binding constant²⁵ of cAMP-CRP to the operator site by K_{CRP} , and the cooperativity factor of CRP-RNAP interaction by ω , then the rate of LacZ synthesis is

$$G = G_0 \cdot \frac{1 + \omega \cdot [CRP^*] / K_{CRP}}{1 + [CRP^*] / K_{CRP}}. \quad [2.6.1]$$

where G_0 is the same basal rate introduced in Eq. [2.5.2]. Eq. [2.6.1] describes the rate of LacZ synthesis. What we measure is the enzymatic activity (A) of LacZ (see Section 2.2, *β -galactosidase assay*) in the steady state of balanced exponential growth. We take A to be proportional to the steady-state LacZ concentration $[LacZ]^*$. The latter is given through the kinetic equation

$$\frac{d}{dt}[LacZ] = G - \lambda \cdot [LacZ],$$

with the steady-state solution $G = \lambda \cdot [LacZ]^* \propto \lambda \cdot A$, where λ is the cell growth rate. In our study, we defined the promoter activity to be $\alpha \equiv A \cdot \lambda_{\frac{1}{2}}$ where $\lambda_{\frac{1}{2}} = \lambda / \ln 2$ is the doubling rate of the culture; see Eq. [2.3.1]. Hence, $G \propto \alpha$ as expected. Inserting the cAMP-dependence of $[CRP^*]$ as given by Eq. [2.4.2], we have

$$\alpha_{cAMP} = b_0 \cdot \frac{1 + \omega \cdot [cAMP] / \kappa_{cAMP}}{1 + [cAMP] / \kappa_{cAMP}}, \quad [2.6.2]$$

where $b_0 \propto G_0 \cdot \lambda$ and

$$\kappa_{cAMP} = K_{cAMP} \cdot K_{CRP} / [CRP] \cdot \gamma_{cAMP}. \quad [2.6.3]$$

Eq. [2.6.2] is the Hill form used in chapter one for the cAMP response (Eq. [1.3.1]), with $f_{cAMP} = \omega$, and $C_{cAMP} = \kappa_{cAMP}$. The predicted range for C_{cAMP} listed in Table 1.3 is obtained by using Eq. [2.6.3] with $K_{cAMP} = 10 \mu\text{M}$ ¹⁰, $K_{CRP} = 5 - 50 \text{ nM}$ ¹⁰, $\gamma_{cAMP} = 0.001$ (see discussion above), and $[CRP] = 1500 \text{ nM}$ ²⁶.

2.7 Simple repression by LacR

Let us first consider the case of “simple repression” by the operator O1 alone (see Fig. 1.7), without the involvement of DNA looping and without activation by cAMP-CRP. In this case, repression is due solely to the direct steric interference between the binding of LacR to O1 and RNAP to the core promoter^{17,27,28}. Specifically, both activated species LacR* and LacR** can bind to O1 and block the promoter. Denoting the *in vivo* binding constant between either activated species and O1 by K_1 , the rate of LacZ synthesis is given by Eq. [2.5.3] with $[R] = [LacR^*] + 2[LacR^{**}]$, with the factor of 2 in the second term arising from the fact that there are two ways for LacR** to bind to O1. Thus,

$$G = G_0 / \left(1 + [LacR^*] / K_1 + 2 [LacR^{**}] / K_1 \right). \quad [2.7.1]$$

Inserting into Eq. [2.7.1] the IPTG dependences of $[LacR^{**}]$ and $[LacR^*]$ as given by Eqs. [2.4.3] and [2.4.4], we obtain the result

$$G = G_0 / \left(1 + \frac{2 [LacR] / K_1}{(1 + [IPTG] / K_{IPTG})^2} \right). \quad [2.7.2]$$

Note that the factor $2[LacR]/(1+[IPTG]/K_{IPTG})^2$ in the denominator of Eq. [2.7.2] simply gives the concentration of the dimeric units of LacR tetramers which are not bound by IPTG.

The value of the factor $2[LacR]/K_1$ (denoted by R in Ch. 1) can be inferred by noting that the degree of repression f , defined as the ratio of LacZ expression for cells grown in medium with saturating and zero IPTG, is simply given by $1 + R$ accord to Eq. [2.7.2]. Oehler *et al.* reported in a 1990 paper²⁹ that the degree of repression was ~ 20 for a mutant *lac* promoter containing only the operator O1. In a later paper³⁰, they reported degrees of repression of 200 and 4700 for the same mutant *lac* promoter in cells expressing 5x and 90x respectively of the amount of Lac tetramers of wild-type cells. These results yield an estimate of $R \approx 20$ or $R \approx 50$ for wild-type cells. In the following analysis we use $R \approx 50$, although results of similar quality are obtained for $R \approx 20$.

2.8 LacR-mediated DNA looping

To include repression due to DNA looping, we take into account the additional possibility that a completely uninduced LacR tetramer (i.e., LacR**) may bind to two operator sites, the main operator O1 and one of the two auxiliary operator sites O2 or O3; see Fig. 1.7. This is incorporated into the thermodynamic model by adding two looping terms to the denominator of the expression in Eq. [2.7.1], i.e.,

$$G = G_0 / \left(1 + \frac{[LacR^*]}{K_1} + \frac{2[LacR^{**}]}{K_1} + \frac{[LacR^{**}][L_{12}]}{K_1 K_2} + \frac{[LacR^{**}][L_{13}]}{K_1 K_3} \right) \quad [2.8.1]$$

with K_2 and K_3 being the dissociation constant of the LacR tetramer with O2 and O3 respectively, and $[L_{12}]$ and $[L_{13}]$ describing the effective “local concentration” of the repressor at O2 and O3 respectively given that the tetramer is bound at O1; see refs. 22-24. In the full thermodynamic treatment, each of the two looping terms, $[LacR^{**}][L_{1i}]/K_1K_i$, where $i = 2$ or 3 , is further divided by the factor $(1+[LacR^*]/K_1 + 2[LacR^{**}]/K_i)$. This factor takes into account of the fact that looping of O₁ with the auxiliary operator O_i is not possible if O_i is occupied itself (referred to as the “squelching” effect). We neglected this effect here as the LacR concentration is sufficiently low in wild-type cells. [Note: the squelching term for O1-O3 looping is completely negligible as $2[LacR] \ll K_3$. O1-O2 looping is actually negligible only for $[IPTG] > K_{IPTG}$. This is indeed the regime where most of the data were taken, i.e., most of the data in Fig. 1.8a (squares) are larger than $K_{IPTG} \sim 10 \mu\text{M}$. However, the fact that the data point for $[IPTG] = 0$ also falls on the same curve (bottom red) suggests that $[L_{12}]/2K_2 \ll [L_{13}]/2K_3$.]

The effective concentrations $[L_{12}]$ and $[L_{13}]$ involve the free energy cost of DNA looping and have not been directly characterized^{††} by biochemical or biophysical methods for the *lac* promoter in the absence of CRP binding. Inserting again the IPTG-dependence of $[LacR^{**}]$, we obtain

$$G = G_0 / \left(1 + \frac{2 [LacR] / K_1}{(1 + [IPTG] / K_{IPTG})^2} + \frac{([LacR] / K_1) \cdot ([L_{12}] / K_2 + [L_{13}] / K_3)}{(1 + [IPTG] / K_{IPTG})^4} \right). \quad [2.8.2]$$

As the *Plac* activity α is proportional to G (see above), we have

$$\alpha_{IPTG} = b_0 / \left(1 + \frac{R}{(1 + [IPTG] / K_{IPTG})^2} + \frac{R \cdot L_0}{(1 + [IPTG] / K_{IPTG})^4} \right) \quad [2.8.3]$$

as quoted in Eq. [1.7.1] for the IPTG response, with $R = 2 [LacR] / K_1$ and $L_0 = \frac{1}{2} ([L_{12}] / K_2 + [L_{13}] / K_3)$. With DNA looping, the degree of repression becomes

$$f = 1 + R \cdot (L_0 + 1). \quad [2.8.4]$$

From Eq. [2.8.3], it is also straightforward to compute the dependence of sensitivity (defined e.g., as the maximum logarithmic derivative of the transition region) on the loop parameter L_0 . We find a monotonic function increasing from ~ 1.2 for small values of L_0 to ~ 2.5 at $L_0 = 50$.

In a recent study, Oehler *et al.*³² characterized the IPTG response for Lac systems incapable of DNA looping (either by the removal of the auxiliary operator O3 or by using mutant LacR incapable of tetramerization) resulted in a noticeable decrease in the cooperativity of the IPTG response of the lac promoter. We analyzed the data published in ref. 31 quantitatively and found those results to match very well with the expected behavior developed in this work: As shown in Fig. 1.6a, the IPTG responses for the two systems incapable of DNA looping (red triangles and blue squares) are well-fitted by Eq. [2.8.3] with the looping term L_0 set to zero (solid line). The two fitting parameters $R \sim 200$ and $K_{IPTG} \sim 6 \mu\text{M}$ obtained are consistent with that expected for strains

^{††} Their values can be indirectly inferred from the experiments of Oehler *et al.*³⁰ and Muller *et al.*³⁸ as was done by Vilar and Leibler³¹. However the estimates from the two experiments differ by a factor of ~ 5 .

overproducing LacR 5-10x. The ~1000-fold change of the wild-type strain is a consequence of residual affinity of overexpressed LacR for the operator sequences³³; a previous study by Oehler in which the fold change is determined by LacR+/- comparison shows an 8000-fold change as predicted by the model³⁰.

2.9 Combinatorial control by LacR and cAMP-CRP

We now consider the combined control of the *lac* promoter by both LacR and CRP. According to the thermodynamic model, if there is no interaction between CRP-cAMP mediated activation and LacR mediated repression, then the joint dependence would simply be the algebraic product of the activation and repression functions (Eqs. [2.6.1] and [2.8.1]) in a way analogous to Eq. [2.5.4], i.e.,

$$G = G_0 \cdot \frac{\left(\frac{1 + \omega \cdot [CRP^*] / K_{CRP}}{1 + [CRP^*] / K_{CRP}} \right)}{\left(1 + \frac{[LacR^*]}{K_1} + \frac{2 [LacR^{**}]}{K_1} + \frac{[LacR^{**}][L_{12}]}{K_1 K_2} + \frac{[LacR^{**}][L_{13}]}{K_1 K_3} \right)} \quad [2.9.1]$$

Alternatively, the co-dependence of the promoter activity on cAMP and IPTG can be obtained from the product of Eq. [2.6.2] and Eq. [2.8.3],

$$\alpha = b_0 \cdot \frac{\left(\frac{1 + \omega \cdot [cAMP] / \kappa_{cAMP}}{1 + [cAMP] / \kappa_{cAMP}} \right)}{\left(1 + \frac{R}{(1 + [IPTG] / K_{IPTG})^2} + \frac{R \cdot L_0}{(1 + [IPTG] / K_{IPTG})^4} \right)} \quad [2.9.2]$$

Note that Eq. [2.9.2] is of the same form as the IPTG response of Eq. [2.8.3], but with the basal activity b_0 replaced by α_{cAMP} of Eq. [2.6.2]. This corresponds to a simple cAMP-dependent vertical shift of the IPTG response in the log-log plot.

As the data shown in Fig. 1.8b clearly deviated from the above product form (see below), we examined the consequence of a previously proposed interaction between CRP and DNA looping^{18,34-36}. As proposed by Lewis *et al.*¹⁸, upon binding to its operator, CRP induces a 90-130° bend in the DNA, bringing O1 and O3 into closer proximity. This is thought to increase the local concentration $[L_{13}]$, thereby increasing the strength of the looping interaction between them. Experimental evidence for such CRP enhancement of O1-O3 looping has been found *in vitro*³⁴⁻³⁶, but no supporting evidence has been reported so far *in vivo*.

This effect can be incorporated into the thermodynamic model developed above simply by including a cooperativity factor $\Omega_0 > 1$ into the term in the denominator of Eq. [2.9.1] involving the product of $[CRP^*]$ and $[L_{13}]$. This amounts to changing the factor $[L_{13}]$ appearing in Eq. [2.9.1] to

$$[L_{13}] \cdot \left(1 + \Omega_0 \cdot [CRP^*] / K_{CRP}\right) / \left(1 + [CRP^*] / K_{CRP}\right). \quad [2.9.3]$$

It then follows that the cAMP and IPTG dependence of the promoter activity can be written in the same form as Eq. [2.9.2], but with the looping factor L_0 changed to

$$L = L_0 \cdot \frac{1 + \Omega \cdot [cAMP] / \kappa_{cAMP}}{1 + [cAMP] / \kappa_{cAMP}}, \quad [2.9.4]$$

where Ω is an effective cooperativity factor given by

$$\Omega = \left(\frac{[L_{12}]}{K_2} + \Omega_0 \cdot \frac{[L_{13}]}{K_3} \right) / \left(\frac{[L_{12}]}{K_2} + \frac{[L_{13}]}{K_3} \right). \quad [2.9.5]$$

The parameters $[L_{12}]$, $[L_{13}]$, and Ω_0 have not been determined from direct biochemical studies. However, Fried *et al.*^{34,36} have measured an effective cooperativity factor defined as the ratio of concentration of the cAMP-CRP-LacR-DNA ternary product with the product of the concentrations of cAMP-CRP-DNA and LacR-DNA binary products. In terms of the parameters defined above, this effective cooperativity factor can be written as

$$\frac{\left(1 + \frac{[CRP^*]}{K_{CRP}}\right)(1 + R + R \cdot L) + (\Omega_0 - 1) \cdot \frac{[CRP^*]}{K_{CRP}} R \cdot \frac{[L_{13}]}{2K_3}}{\left(1 + \frac{[CRP^*]}{K_{CRP}}\right)(1 + R + R \cdot L)}. \quad [2.9.6]$$

For saturating amounts of cAMP-CRP and LacR such that $[CRP] \gg K_{CRP}$ and $R \gg 1$, the expression in Eq. [2.9.6] is identical to the right hand side of Eq. [2.9.5]. Hence, the effective cooperativity factor Ω introduced in Eq. [2.9.4] is just the cooperativity factor of the range 4–12 determined experimentally by Fried *et al.*^{34,36}.

Finally, we consider the case where $[CRP^*]$ depends nonlinearly on the cAMP concentration of the medium such that the observed cAMP response is described by the general Hill form

$$\alpha_{cAMP} = b_0 \cdot \frac{1 + \omega \cdot ([cAMP] / C_{cAMP})^m}{1 + ([cAMP] / C_{cAMP})^m} \quad [2.9.7]$$

with Hill coefficient $m > 1$. This is for example the situation for TK250 cells bearing the *cpdA* gene (see blue symbols in Fig. 1.9b). Working backward from Eq. [2.6.1], we have

$$[CRP^*] = K_{CRP} ([cAMP] / C_{cAMP})^m \quad [2.9.8]$$

Using Eq. [2.9.8] instead of Eq. [2.4.2] in the combinatorial control function (Eq. [2.9.1] with $[L_{13}]$ replaced by the expression in Eq. [2.9.3], we obtain the following for the co-dependence of the promoter activity:

$$\alpha = b_0 \cdot \frac{\left(\frac{1 + \omega \cdot ([cAMP] / C_{cAMP})^m}{1 + ([cAMP] / C_{cAMP})^m} \right)}{\left(1 + \frac{R}{(1 + [IPTG] / K_{IPTG})^2} + \frac{R \cdot L}{(1 + [IPTG] / K_{IPTG})^4} \right)} \quad [2.9.9]$$

with the loop parameter L in Eq. [2.9.9] being

$$L = L_0 \cdot \frac{1 + \Omega \cdot ([cAMP] / C_{cAMP})^m}{1 + ([cAMP] / C_{cAMP})^m}, \quad [2.9.10]$$

and the cooperativity factor Ω in Eq. [2.9.10] still given by Eq. [2.9.5].

2.10 Comparison to alternative models

While our model described by Eqs. [2.9.2] and [2.9.4] appears to accurately describe the data (see Fig. 1.8b), one might question whether conceptually simpler models (e.g., generalized Hill functions) might explain the data equally well. Below, we construct several such models and fit to the data surface from strain TK310 using non-linear least squares minimization in MATLAB (Mathworks). The points at $[IPTG] = 0$ are weighted to increase their significance (to simulate the wide range (in log scale) of low IPTG concentrations ($< 5 \mu M$) over which the baseline expression from the promoter

does not significantly increase). For each model the specified parameters are allowed to vary freely to obtain the optimum fit.

Model a. A naïve 6-parameter model based on generic Hill functions and assuming no interaction between LacR and CRP:

$$\alpha = b_0 \cdot \left(\frac{1 + \omega([cAMP]/C_{cAMP})^{m_{cAMP}}}{1 + ([cAMP]/C_{cAMP})^{m_{cAMP}}} \right) \left/ \left(1 + \frac{R}{1 + ([IPTG]/C_{IPTG})^{m_{IPTG}}} \right) \right. \quad [2.10.1]$$

As can be seen from Fig. 2.3a, neither the fold change nor the sensitivity to IPTG response is properly captured by this model for the range of cAMP concentrations studied.

Model b. A 7-parameter model in which the coefficient of the IPTG dependent term, i.e., R in Eq. [2.10.1] above, is a four-parameter cAMP-dependent Hill function,

$$R = R_0 \cdot \frac{1 + \omega'([cAMP]/C'_{cAMP})^{m'_{cAMP}}}{1 + ([cAMP]/C'_{cAMP})^{m'_{cAMP}}}, \quad [2.10.2]$$

where we assume $C'_{cAMP} = C_{cAMP}$, and $m'_{cAMP} = m_{cAMP}$, but allow R_0 and ω' to vary freely.

From Fig. 2.3b, we see that the fold change of the IPTG response is better reproduced than Model a, but the overall correspondence between data and model prediction is still poor, particularly for intermediate IPTG concentrations (10-500 μ M) for both the high and low cAMP levels.

Model c. The same as model Model b, but with C'_{cAMP} , and m'_{cAMP} both varying freely.

From Fig. 2.3c it is seen that the fit to the data is further improved. However the number of parameters is increased to nine while the fit to data points at intermediate IPTG concentrations is still problematic, especially at low cAMP levels.

Model d. The model Eq. [2.10.1] with the modification to R made in Eq. [2.10.2], and with the Hill coefficient m_{IPTG} in Eq. [2.10.1] also cAMP dependent. The latter is specified by another Hill function

$$m_{IPTG} = m_0 \cdot \frac{1 + \omega''([cAMP]/C''_{cAMP})^{m''_{cAMP}}}{1 + ([cAMP]/C''_{cAMP})^{m''_{cAMP}}} . \quad [2.10.3]$$

Here, we use $C''_{cAMP} = C'_{cAMP} = C_{cAMP}$, $\hat{m}''_{cAMP} = m'_{cAMP} = m_{cAMP}$ but allow both ω'' and ω' to vary freely. This 8-parameter model is able to describe the full data set well; see Fig. 2.3d. This finding is not surprising given the result in Fig. 1.4a, that the effective Hill coefficient of the IPTG response is strongly cAMP dependent.

We see that while both our model (i.e. Eqs. [2.9.2] and [2.9.4]) and Model d above (i.e., Eqs. [2.10.1]-[2.10.3]) adequately describe the experimental data, our model requires seven parameters while Model d requires eight. More importantly, the form used in our model is well founded on known biochemical facts as described above, and the values of fitted parameters compare favorably with the known range in all cases when comparison can be made. (Note: The loop cost L_0 is the only parameter where independent estimate is not available.) On the other hand, it is difficult to justify biochemically the cAMP-dependence of the Hill coefficient in Model d. The above

results therefore demonstrate that our model based on the thermodynamic treatment of DNA looping and its interaction with CRP is the minimal model describing the activity of the *lac* promoter.

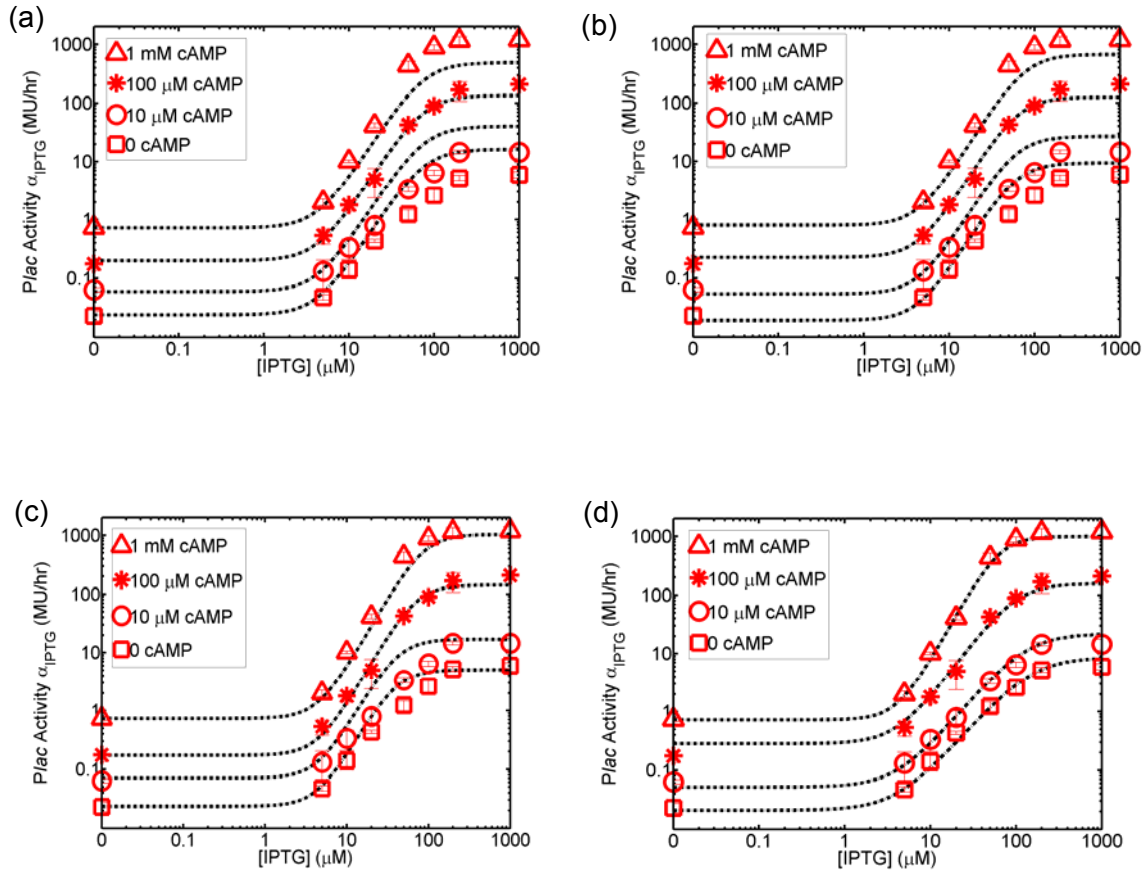


Figure 2.3: Comparison to alternative models. (a) A naïve model assuming no interaction between CRP and LacR. (b) A model assuming cAMP dependent fold change with $C'_{cAMP} = C_{cAMP}$, $m'_{cAMP} = m_{cAMP}$ (see Eq. [2.10.2]). (c) Same as b, but with freely varying cAMP related parameters. (d) Same as b, but assuming a LacR-IPTG Hill coefficient, m_{IPTG} , governed by a Hill function (Eq. [2.10.3]).

2.11 Data Analysis

For each strain and medium used, the promoter activity α was computed from the raw data (i.e., β -galactosidase activity and growth curves) according to [2.3.1]. Its dependence on either the concentrations of cAMP or IPTG in the medium was then fitted to an effective Hill form

$$\alpha_X = b_X \cdot \frac{1 + f_X ([X] / C_X)^{m_X}}{1 + ([X] / C_X)^{m_X}} \quad [2.11.1]$$

where $[X]$ refers to the concentration of the inducer X (cAMP or IPTG) in the medium. In the fits, the basal activity b_X was directly fixed by the promoter activity obtained for $[X] = 0$. The other three Hill parameters were obtained by fitting to Eq. [2.11.1] using nonlinear least-square minimization (MATLAB 7). Reported errors are the 68.3% (1 standard deviation) confidence intervals for the given parameters.

Because DNA looping is involved in repression of the *lac* promoter by the Lac tetramer^{17,29,30,37,38}, the IPTG responses of strains TK230 and TK310 (in medium with no cAMP added) were also fitted to the form Eq. [2.8.3] predicted by the DNA looping model above. In the fit, we fixed the value of R to 50 as discussed above. We then used the experimentally determined promoter activity in saturating 1 mM IPTG condition to fix the coefficient b_0 , and used the remaining data to fit the two parameters K_{IPTG} , and L_0 according the procedure described above. The results are listed in Table 1.3 (rows 1 and 2). [The values of the IPTG-LacR dissociation constant K_{IPTG} obtained are comparable to the results of *in vitro* biochemical studies: In the absence of DNA, the LacR-IPTG

interaction was characterized by $K_{\text{IPTG}} = 0.8 - 4 \mu\text{M}$ at physiologically relevant pH values^{33,39,40}. In the presence of a saturating amount of operator DNA however, a much weaker LacR-IPTG affinity characterized by $K_{\text{IPTG}} \approx 80 \mu\text{M}$ was found, although with non-specific DNA fragments the smaller value of K_{IPTG} was obtained as before²⁰. Our results with $K_{\text{IPTG}} = 15 \sim 20 \mu\text{M}$ lie in between this range, presumably reflecting the interaction of LacR with both specific and non-specific sequences *in vivo*.]

For the IPTG response of strain TK310 and TK250 grown in medium with various cAMP concentrations (the data in Fig. 1.8*b* and Fig. 1.9*b*), we determined the looping parameter L for the IPTG dependence at each cAMP concentration by first computing the degree of repression f as the ratio of b_0 (determined as above) and the value of the promoter activity at $[\text{IPTG}] = 0$, and then extracting the loop parameter from f using Eq. [2.8.4]. The values of L obtained were plotted in Fig. 1.8*a* and Fig. 1.9*a* for strain TK310 and TK250 respectively. Then for the IPTG response of strain TK310, the predicted form Eq. [2.9.2] was plotted as the dashed line in Fig. 1.8*b* using the values of L shown in Fig. 1.8*a* and other parameter values as listed in Table 1.3 (row 2) and Table 1.4 (row 2). Similarly, for the IPTG response of strain TK250, the predicted form Eq. [2.9.9] was plotted as the dashed line in Fig. 1.9*b* using the values of L shown in Fig. 1.9*a* and other parameter values as listed in Table 1.3 (row 1) and Table 1.4 (row 3).

Finally, the cAMP dependence of the loop parameter L was analyzed by fitting their values to the expected form according to the extended thermodynamic model, Eq. [2.9.4] for strain TK310 and Eq. [2.9.10] for strain TK250. In these fits, we varied only the effective cooperativity factor Ω , taking the values of the other parameters from

Table 1.3 (row 2) and Table 1.4 (row 2) for strain TK310, and Table 1.3 (row 1) and Table 1.4 (row 3) for strain TK250. Estimate of error on Ω was again obtained using the procedure described above.

2.12 Acknowledgement

Chapter 2 contains material from the supplementary information contained in Kuhlman T, Zhang Z, Saier MH, Jr., Hwa T. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc Natl Acad Sci* 2007; 104(14): 6043-6048., used with permission. The dissertation author was the primary investigator and author of this work.

2.13 References

1. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*. 2000 Jun 6;97(12):6640-5.
2. Simons RW, Houman F, Kleckner N. Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene*. 1987;53(1):85-96.
3. Miller JH. *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1972. 352-5 p.
4. Griffith KL, Wolf RE, Jr. Measuring beta-galactosidase activity in bacteria: cell growth, permeabilization, and enzyme assays in 96-well arrays. *Biochem Biophys Res Commun*. 2002 Jan 11;290(1):397-402.
5. Mandelstam J. Turnover of protein in starved bacteria and its relationship to the induced synthesis of enzyme. *Nature*. 1957 Jun 8;179(4571):1179-81.
6. Wanner BL, Kodaira R, Neidhardt FC. Regulation of lac operon expression: reappraisal of the theory of catabolite repression. *J Bacteriol*. 1978 Dec;136(3):947-54.
7. Wanner BL, Kodaira R, Neidhardt FC. Physiological regulation of a decontrolled lac operon. *J Bacteriol*. 1977 Apr;130(1):212-22.
8. Malan TP, Kolb A, Buc H, McClure WR. Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. *J Mol Biol*. 1984 Dec 25;180(4):881-909.
9. Takahashi M, Blazy B, Baudras A. An equilibrium study of the cooperative binding of adenosine cyclic 3',5'-monophosphate and guanosine cyclic 3',5'-monophosphate to the adenosine cyclic 3',5'-monophosphate receptor protein from *Escherichia coli*. *Biochemistry*. 1980 Oct 28;19(22):5124-30.
10. Takahashi M, Blazy B, Baudras A, Hillen W. Ligand-modulated binding of a gene regulatory protein to DNA. Quantitative analysis of cyclic-AMP induced binding of CRP from *Escherichia coli* to non-specific and specific DNA targets. *J Mol Biol*. 1989 Jun 20;207(4):783-96.
11. Fried MG, Crothers DM. Equilibrium studies of the cyclic AMP receptor protein-DNA interaction. *J Mol Biol*. 1984 Jan 25;172(3):241-62.
12. Garner MM, Revzin A. Stoichiometry of catabolite activator protein/adenosine cyclic 3',5'-monophosphate interactions at the lac promoter of *Escherichia coli*. *Biochemistry*. 1982 Nov 23;21(24):6032-6.

13. Brown AM, Crothers DM. Modulation of the stability of a gene-regulatory protein dimer by DNA and cAMP. *Proc Natl Acad Sci U S A*. 1989 Oct;86(19):7387-91.
14. Saier MH, Jr., Feucht BU, McCaman MT. Regulation of intracellular adenosine cyclic 3':5'-monophosphate levels in *Escherichia coli* and *Salmonella typhimurium*. Evidence for energy-dependent excretion of the cyclic nucleotide. *J Biol Chem*. 1975 Oct 10;250(19):7593-601.
15. Goldenbaum PE, Hall GA. Transport of cyclic adenosine 3',5'-monophosphate across *Escherichia coli* vesicle membranes. *J Bacteriol*. 1979 Nov;140(2):459-67.
16. Epstein W, Rothman-Denes LB, Hesse J. Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 1975 Jun;72(6):2300-4.
17. Muller-Hill B. *The lac Operon*. Berlin: de Gruyter; 1996.
18. Lewis M, Chang G, Horton NC, et al. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*. 1996 Mar 1;271(5253):1247-54.
19. Winter RB, Berg OG, von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor--operator interaction: kinetic measurements and conclusions. *Biochemistry*. 1981 Nov 24;20(24):6961-77.
20. O'Gorman RB, Rosenberg JM, Kallai OB, et al. Equilibrium binding of inducer to lac repressor.operator DNA complex. *J Biol Chem*. 1980 Nov 10;255(21):10107-14.
21. Ackers GK, Johnson AD, Shea MA. Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci U S A*. 1982 Feb;79(4):1129-33.
22. Buchler NE, Gerland U, Hwa T. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A*. 2003 Apr 29;100(9):5136-41.
23. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev*. 2005 Apr;15(2):125-35.
24. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*. 2005 Apr;15(2):116-24.

25. Gerland U, Moroz JD, Hwa T. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc Natl Acad Sci U S A*. 2002 Sep 17;99(19):12015-20.
26. Cossart P, Gicquel-Sanzey B. Regulation of expression of the *crp* gene of *Escherichia coli* K-12: in vivo study. *J Bacteriol*. 1985 Jan;161(1):454-7.
27. Schmitz A, Galas DJ. The interaction of RNA polymerase and lac repressor with the lac control region. *Nucleic Acids Res*. 1979 Jan;6(1):111-37.
28. Gilbert W, Gralla JD, Majors J, Maxam A. Lactose operator sequences and the action of lac repressor. In: Sund H, Blauer R (eds). *Protein-Ligand Interactions*. Berlin-New York: de Gruyter, 1975:193-206.
29. Oehler S, Eismann ER, Kramer H, Muller-Hill B. The three operators of the lac operon cooperate in repression. *EMBO*. 1990;9(4):973-9.
30. Oehler S, Amouyal M, Kokhof P, von Wilcken-Bergmann B, Muller-Hill B. Quality and position of three lac operators of *E. coli* define efficiency of repression. *EMBO*. 1994;13(14):973-9.
31. Vilar JM, Leibler S. DNA looping and physical constraints on transcription regulation. *J Mol Biol*. 2003 Aug 29;331(5):981-9.
32. Oehler S, Alberti S, Muller-Hill B. Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Res*. 2006;34(2):606-12.
33. Gilbert W, Muller-Hill B. The Lactose Repressor. In: Beckwith J, Zipser D (eds). *The Lactose Operon*. Cold Spring Harbor NJ: Cold Spring Harbor Press, 1970:93-109.
34. Fried MG, Hudson JM. DNA looping and lac repressor-CAP interaction. *Science*. 1996 Dec 13;274(5294):1930-1; author reply 1-2.
35. Hudson JM, Fried MG. Co-operative interactions between the catabolite gene activator protein and the lac repressor at the lactose promoter. *J Mol Biol*. 1990 Jul 20;214(2):381-96.
36. Vossen KM, Stickle DF, Fried MG. The mechanism of CAP-lac repressor binding cooperativity at the *E. coli* lactose promoter. *J Mol Biol*. 1996 Jan 12;255(1):44-54.

37. Eismann ER, Muller-Hill B. lac repressor forms stable loops in vitro with supercoiled wild-type lac DNA containing all three natural lac operators. *J Mol Biol.* 1990 Jun 20;213(4):763-75.
38. Muller J, Oehler S, Muller-Hill B. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol.* 1996 Mar 22;257(1):21-9.
39. Nichols JC, Matthews KS. Combinatorial mutations of lac repressor. Stability of monomer-monomer interface is increased by apolar substitution at position 84. *J Biol Chem.* 1997 Jul 25;272(30):18550-7.
40. Jobe A, Bourgeois S. lac Repressor-operator interaction. VI. The natural inducer of the lac operon. *J Mol Biol.* 1972 Aug 28;69(3):397-408.

CHAPTER 3

Autoregulation as a Strategy in Synthetic Reporter Constructs

3.1 Introduction

Synthetic constructs able to express a gene of interest at easily controllable levels from inducible promoters in *E. coli* are some of the most useful tools of experimental quantitative biology. These inducible promoters are generally regulated by a transcription factor, such as LacR¹ or AraC^{2,3}, whose ability to enhance or repress transcription from a promoter is either significantly reduced or enhanced when the transcription factor is induced via some mechanism, such as heat shock or binding to a small molecule. As such, using modern molecular cloning techniques it is a simple matter to place a gene to be studied under the regulation of an inducible promoter, and then characterize the response of a system to increasing or decreasing amounts of the gene as a function of induction. For example, these constructs can potentially be used as controllable nodes in a network⁴ or to quantitatively explore the response of the cell or network to varying amounts of the gene (Ref. 5 and Chapters 4 and 5), among myriad other uses.

Unfortunately, the number of simple, effective promoters able to drive the expression of genes in response to easily achieved induction conditions is rather limited. There are a number of inducible promoters available in *E. coli* which respond to damage to the cell⁶, change in temperature^{7,8}, or introduction of compounds which may or may not affect the physiology or growth rate of the cell⁹⁻¹¹ (see also Fig. 1.3a). However, in many of these instances the changes in conditions required to induce expression from the

promoter lead to extensive physiological changes resulting in altered growth conditions or rates, making comparison of the effect of gene expression difficult or impossible at different levels. It is therefore advantageous to use an inducer with limited toxicity and/or physiological effects such that the effects of expression of the gene of interest can be easily compared under identical circumstances across a wide dynamic range. In *E. coli*, the most popular and readily available promoters inducible by small molecules with limited side-effects are usually regulated by the transcription factors LacR¹², TetR¹², and AraC^{12,13}, inducible by isopropyl- β -D-1-thiogalactopyranoside (IPTG), anhydrotetracycline (aTc), and L-arabinose respectively.

Of these, when the response of the promoter to varying concentrations of the small molecule inducer is assayed, it is generally seen that the response is very sensitive; expression from the promoter is either completely on or completely off, with only a very narrow range of inducer concentrations able to give intermediate values of promoter expression¹² (for example, see Fig. 3.3, blue circles). While in many cases a degree of this sensitivity can be attributed to feedback effects from membrane bound transporters enhancing uptake of the inducer in a nonlinear fashion^{14,15} or other effects such as DNA looping (Refs. 16,17 and Chapters 1 and 2), removal or modification of the responsible system often leads to only a small decrease in the sensitivity of the response or a corresponding decrease in the dynamic range of the promoter^{13,16,18}. It is therefore currently extremely difficult to accurately measure the effect of expression of the inducible promoter at intermediate expression values.

In this chapter we propose a scheme to elicit a more gradual response from inducible promoters while maintaining a wide dynamic range based on the concept of

autoregulation, in which the repressor proteins regulate their own expression as well that of the reporter. First, it is shown using a thermodynamic model of genetic regulation by transcription factors (see Refs. 19-21 and Chapters 1 and 2) that autoregulation of the transcription factor generally leads to less sensitive expression of the reporter than an equivalent system in which the transcription factor is supplied in constant amounts by a constitutive source. We then experimentally demonstrate this concept using the synthetic inducible promoters $P_{\text{LtetO-1}}$ and $P_{\text{LlacO-1}}$ of Lutz and Bujard¹². With the theory used to describe the system in hand, we are then able to discuss ways to further improve the performance of the system. We note that the strategy of autoregulation is widely used in nature^{22,23} and suggest the potential added control over gene expression levels in response to environmental conditions afforded by autoregulation as a motivation for its widespread adoption.

3.2 Sensitivity and dynamic range of a system with constitutive expression of repressor

Consider a system composed of two genes, one encoding a transcription factor R and the other a gene encoding a reporter protein E such as GFP or LacZ, or any other protein of interest. We will assume that the expression of R is constitutive and that the promoter for the gene E , P_E , contains binding sites for the transcription factor R . When the transcription factor R binds to any of the operators contained within P_E it will repress transcription from that gene, i.e. R is a transcriptional repressor protein (Fig. 3.1*a*). We will throughout refer to this system as the “constitutive system”.

We further assume that, like other transcription factors such as LacR and the tet repressor TetR, that the operator binding specificity of the repressor R is relieved upon binding to a small inducer molecule A (such as IPTG or aTc). The cooperative binding of the inducer molecule to m binding sites of the transcription factor can be modeled with a Hill function (see Chapter 2):

$$\begin{aligned} [R_0] &= [R] \cdot \frac{1}{1 + ([A]/K_A)^m} \\ [R_m] &= [R] \cdot \frac{([A]/K_A)^m}{1 + ([A]/K_A)^m}, \end{aligned} \quad [3.2.1]$$

where $[R]$ is the total concentration of transcription factor in the cell, $[R_0]$ is the concentration of transcription factor not bound to inducer, and $[R_m]$ is the concentration of transcription factor bound to m inducer molecules.

We now wish to investigate the expression profile of this system as a function of the inducer concentration. The system is modeled with a simple mass-action model consisting of the regulatory transcription factor (R) and the protein (E):

$$\begin{aligned} \frac{d[R]}{dt} &= \gamma_R - \beta_R[R] \\ \frac{d[E]}{dt} &= \gamma_E \cdot g([R_0]/K_E) - \beta_E[E], \end{aligned} \quad [3.2.2]$$

where the basal synthesis and degradation rates are given by the parameters γ_i and β_i respectively. Assuming that P_E contains n_E identical, independent binding sites for the repressor (as is the case for $P_{\text{Ltet-O1}}$ and $P_{\text{Llac-O1}}$; Ref 12 and Fig. 3.5), the regulatory function $g([R_0]/K_E)$ is given by

$$g([R_0]/K_E) = \frac{1}{1 + \sum_{j=1}^{n_E} \binom{n}{j} \left(\frac{[R_0]}{K_E}\right)^j} \quad [3.2.3]$$

$$\simeq \frac{1}{1 + ([R_0]/K_E)^{n_E}}.$$

Using Eq. [3.2.1], this can be written in terms of the total repressor concentration $[R]$ and inducer concentration $[A]$ as

$$g([R]/K_E, [A]) = \frac{1}{1 + \left(\frac{[R]}{K_E} \cdot \frac{1}{1 + ([A]/K_A)^m} \right)^{n_E}}. \quad [3.2.4]$$

The steady state concentrations of R and E are then obtained by setting Eqs. [3.2.2] to zero:

$$[R^*]_{const} = \frac{\gamma_R}{\beta_R}$$

$$[E^*]_{const} = \frac{\gamma_E}{\beta_E} \cdot \frac{1}{1 + \left(\frac{[R^*]_{const}}{K_E} \cdot \frac{1}{1 + ([A]/K_A)^m} \right)^{n_E}}. \quad [3.2.5]$$

Examining the behavior of this function in the exponentially increasing regime $[A] \gg K_A$, we see that

$$[E^*]_{const} \simeq \frac{\gamma_E \gamma_R}{\beta_E \beta_R K_E} \cdot \left(\frac{[A]}{K_A} \right)^{m \cdot n_E}, \quad [3.2.6]$$

i.e. that the sensitivity of the system is characterized by an effective Hill coefficient $m_{const} = m \cdot n_E$. Additionally, exploring expression levels in the limits $[A] \rightarrow 0$ and $[A] \rightarrow \infty$, we see that the fold change is given by

$$\text{fold change} = \frac{\lim_{[A] \rightarrow \infty} [E^*]_{const}}{\lim_{[A] \rightarrow 0} [E^*]_{const}} \simeq \left(\frac{[R^*]_{const}}{K_E} \right)^{n_E} \quad [3.2.7]$$

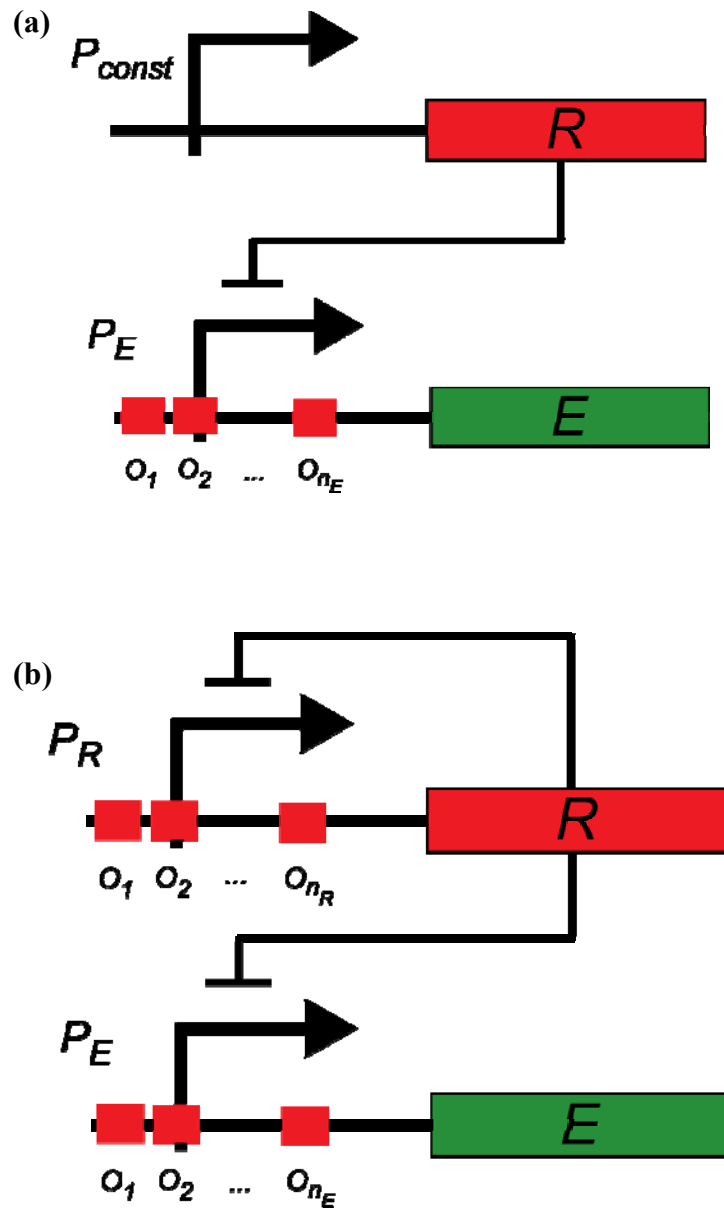


Figure 3.1: Architecture of constitutive and autoregulatory expression systems. (a) The constitutive expression system, with repressor R being constitutively expressed and regulating the promoter of the reporter gene, P_E , containing n_E operators O_i . **(b)** The autoregulated expression system, with the repressor R regulating its own expression from promoter P_R (containing n_R operators) as well as the reporter expression from P_E .

3.3 Sensitivity and dynamic range of a system with autoregulatory control

To examine the effects of autoregulation on the behavior of the system as a function of applied inducer concentration, we now place the repressor R under the control of a self-inhibitory feedback loop by replacing the constitutive promoter by a promoter P_R repressible by R . Like P_E , we again assume n_R identical, independent binding sites for the repressor (See Fig 3.1b). We will throughout refer to this system as the “autoregulated system”. Eqs. [3.2.2] are then modified according to

$$\begin{aligned}\frac{d[R]}{dt} &= \gamma_R \cdot g([R_0]/K_R) - \beta_R[R] \\ \frac{d[E]}{dt} &= \gamma_E \cdot g([R_0]/K_E) - \beta_E[E],\end{aligned}\tag{3.3.1}$$

with $g([R_0]/K_i)$ again given by Eq. [3.2.4] with the appropriate parameters. We further assume that the operator sites are sufficiently strong that the total amount of repressor is at all times sufficient to efficiently repress the system, i.e. that $[R] \gg K_i$.

We again inspect the behavior of this function in the regime of exponentially increasing expression, when $[A] \gg K_A$. In this regime, the regulatory function $g([R]/K_i, [A])$ can be approximated as

$$g([R]/K_i, [A]) \approx \left(\frac{[R]}{K_i}\right)^{-n_i} \cdot \left(\frac{[A]}{K_A}\right)^{m \cdot n_i}.\tag{3.3.2}$$

The steady state concentrations for R and E are found to be

$$\begin{aligned}
\frac{[R^*]_{auto}}{K_R} &= \left(\frac{\gamma_R}{\beta_R K_R} \right)^{\frac{1}{n_R+1}} \cdot \left(\frac{[A]}{K_A} \right)^{\frac{m \cdot n_R}{n_R+1}} \\
[E^*]_{auto} &= \left(\frac{K_E}{K_R} \right)^{n_E} \cdot \left(\frac{\gamma_E}{\beta_E} \right) \cdot \left(\frac{\beta_R K_R}{\gamma_R} \right)^{\frac{n_E}{n_R+1}} \cdot \left(\frac{[A]}{K_A} \right)^{\frac{m \cdot n_E}{n_R+1}}.
\end{aligned} \tag{3.3.3}$$

Comparing the result Eq. [3.3.3] to that of the constitutive system, Eq. [3.2.6], we see that the effect of autoregulation on the system is to decrease the effective Hill coefficient from $m_{const} = m \cdot n_E$ to $m_{auto} = m \cdot n_E / (n_R + 1)$, i.e. by a factor dependent upon the number of repressor binding sites at the promoter P_R . As the requirement for autoregulation is that there be at least one binding site for R within the promoter P_R , the effective Hill coefficient is at least half that of the constitutive system, with the sensitivity decreasing further with increasing n_R (see Fig. 3.2).

To determine the dynamic range of the system, we examine the behavior of the system in the limit $[R] \gg K_i$. Solving for the steady state of Eqs. [3.3.1] and taking the limits $[A] \rightarrow 0$ and $[A] \rightarrow \infty$, the fold change is given by

$$\begin{aligned}
\text{fold change} &= \left(\frac{[R^*]_{Min}}{K_E} \right)^{n_E} \\
[R^*]_{Min} &= \left(\frac{\gamma_R}{\beta_R} \cdot K_R^{n_R} \right)^{\frac{1}{n_R+1}}
\end{aligned} \tag{3.3.4}$$

If the repressed transcription rate of P_R is equivalent to the constitutive promoter rate in the constitutive system (i.e. $[R^*]_{Min} = [R^*]_{const}$), Eq. [3.3.4] is identical to Eq. [3.2.7].

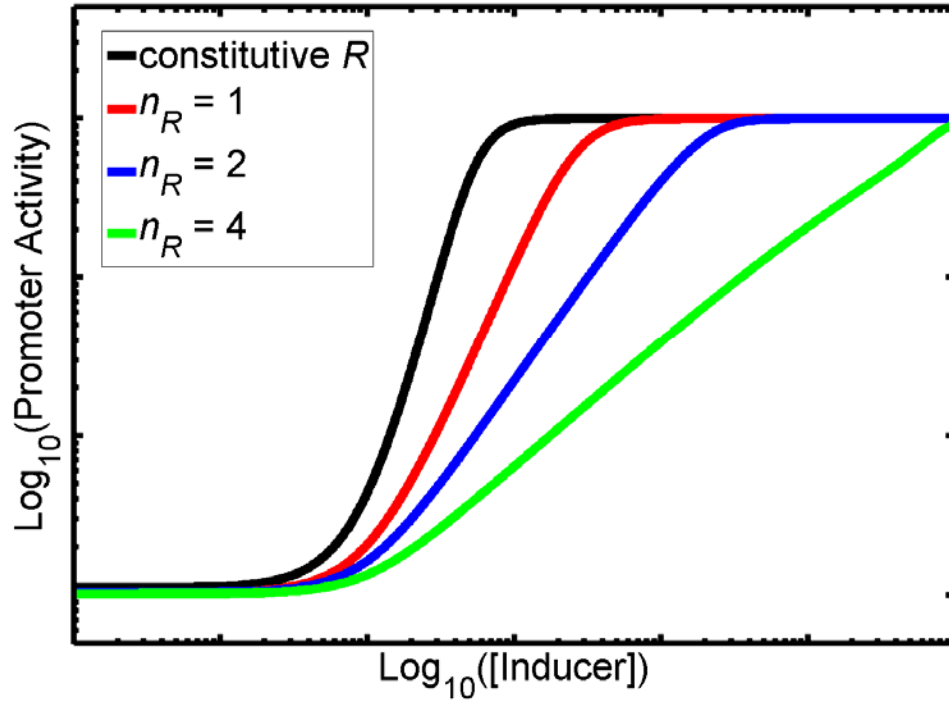


Figure 3.2: Effect of autoregulation on reporter expression. For all curves, $m = n_E = 2$, $\gamma_E = 10 \text{ nM} \cdot \text{s}^{-1}$, $\beta_E = \beta_R = 1 \text{ s}^{-1}$, $K_E = K_R = K_A = 0.01 \text{ nM}$; γ_R is adjusted in each case to yield identical fold changes for ease of comparison. The black curve is the situation with constitutive expression of R , i.e. Eq. [3.2.5]. The other three curves describe increasing degrees of autoregulation. The curves plotted are numerical steady state solutions of Eqs. [3.3.1] with $n_R = 1$ (red), $n_R = 2$ (blue), and $n_R = 4$ (green).

3.4 Experimental implementation of autoregulatory control

In the previous section, it was shown that the sensitivity of an expression system composed of a reporter whose expression is controlled by an autoregulated transcription factor can theoretically be greatly reduced compared to that of a simpler system in which the transcription factor is expressed constitutively. To test this hypothesis, we constructed the reporter plasmid pZE12T-E using the synthetic promoter $P_{\text{Ltet-O1}}$ ¹², which is regulated by TetR, driving the reporter *gfp*. Additionally, plasmids pZA31T-R and pZA31L-R were constructed which express TetR from either the promoter $P_{\text{Ltet-O1}}$ or the promoter $P_{\text{Llac-O1}}$ ¹² respectively.

We first characterized GFP expression from the constitutive expression system by transforming strain TK200 (MG1655 ΔlacI ΔlacY ; see Table 1.1) with the plasmids pZE12T-E and pZA31L-R. Expression of the target was assayed by measuring GFP fluorescence in the resulting cells grown in minimal M63 medium with various amounts of aTc (0 – 100 ng/ml). The resulting GFP expression is plotted as the blue circles in Fig. 3.3a. The resulting expression pattern from the constitutive system is very sensitive, with a Hill coefficient $m_{\text{tet}} = 5.9 \pm 0.3$ and a fold change of $f_{\text{const}} = 194 \pm 12$.

We next characterized expression from the autoregulated expression system by transforming strain TK200 with the plasmids pZE12T-E and pZA31T-R. GFP expression was again measured as a function of aTc, and the result is plotted as the red squares in Fig. 3.3a. The abruptness of the transition is significantly reduced, with a fitted effective Hill coefficient of 1.7 ± 0.2 and a fold change of $f_{\text{auto}} = 52.2 \pm 6$.

To test the generality of this result, we repeated these experiments with plasmids constructed using LacR as the repressor [expressed from either $P_{\text{Ltet-O1}}$ (plasmid

pZA31T-L) or P_{Lac-O1} (plasmid pZA31L-L)] and replacing the promoter driving *gfp* to the LacR repressible promoter P_{Lac-O1} (plasmid pZE12L-E). After transformation into strain TK200, the resulting cells were again assayed in minimal M63 medium with various amounts of IPTG (0 – 2 mM). The result is shown in Fig. 3.3b: again, the sensitivity of the response of the autoregulated system with LacR expressed from P_{Lac-O1} (red squares; $m = 1.8 \pm 0.1$) is significantly reduced from that of the constitutive system with LacR expressed from $P_{Ltet-O1}$ (blue circles; $m = 2.4 \pm 0.2$). In this case, however, the fold change of the autoregulated response is also reduced $\sim 30\times$ compared to that of the constitutive system, from $f_{const} = 275 \pm 15$ to only $f_{auto} = 9.6 \pm 1.5$.

It is a possibility that the high sensitivity observed in the response of the constitutive system is the result of averaging over a bimodal population exhibiting autocatalytic all-or-none expression rather than arising as a result of the cooperative interactions of the components of the system. Such behavior has been documented and extensively studied in both the *lac*^{14,24} and *ara*^{13,15,18} systems. In such a situation, populations of induced cells exhibit a bimodal distribution of cells exhibiting low and high expression, with the response to inducer merely corresponding to repartitioning of the population into the two expression states^{13,18}. To verify that the population of cells shows a gradual unimodal increase of fluorescence in response to inducer, we performed flow cytometry on populations of cells harboring the $P_{Ltet-O1}$ autoregulatory expression system to determine the distribution. As seen in Fig. 3.4, the population shows a unimodal distribution of cells homogeneously increasing fluorescence in response to increasing concentrations of inducer.

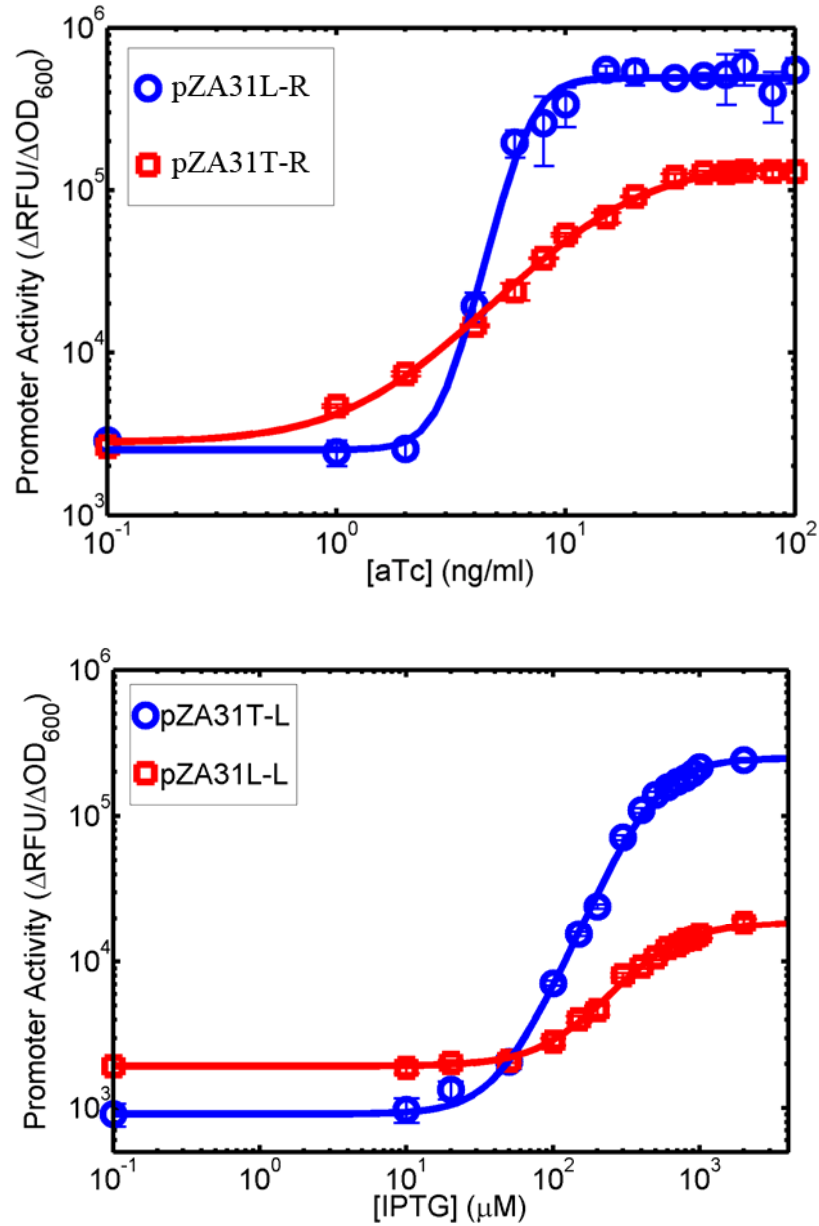


Figure 3.3: Expression profiles of experimental constitutive and autoregulated expression systems. Plasmids expressing GFP from inducible promoters were transformed into TK200 (MG1655 ΔlacI ΔlacY) with plasmids expressing the corresponding repressor protein from either a constitutive source (blue circles) or an autoregulated source (red squares). **(a)** GFP expressed from $P_{\text{Ltet-O1}}$, regulated by TetR. Hill coefficient of constitutive system $m_{\text{const}} = 5.9 \pm 0.3$; autoregulated system $m_{\text{auto}} = 1.7 \pm 0.2$. Fold change of constitutive system $f_{\text{const}} = 194 \pm 12$; autoregulated system $f_{\text{auto}} = 52.2 \pm 6$. **(b)** GFP expressed from $P_{\text{Llac-O1}}$, regulated by LacR. Hill coefficient of constitutive system $m_{\text{const}} = 2.4 \pm 0.2$; autoregulated system $m_{\text{auto}} = 1.8 \pm 0.1$. Fold change of constitutive system $f_{\text{const}} = 275 \pm 15$; autoregulated system $f_{\text{auto}} = 9.6 \pm 1.5$.

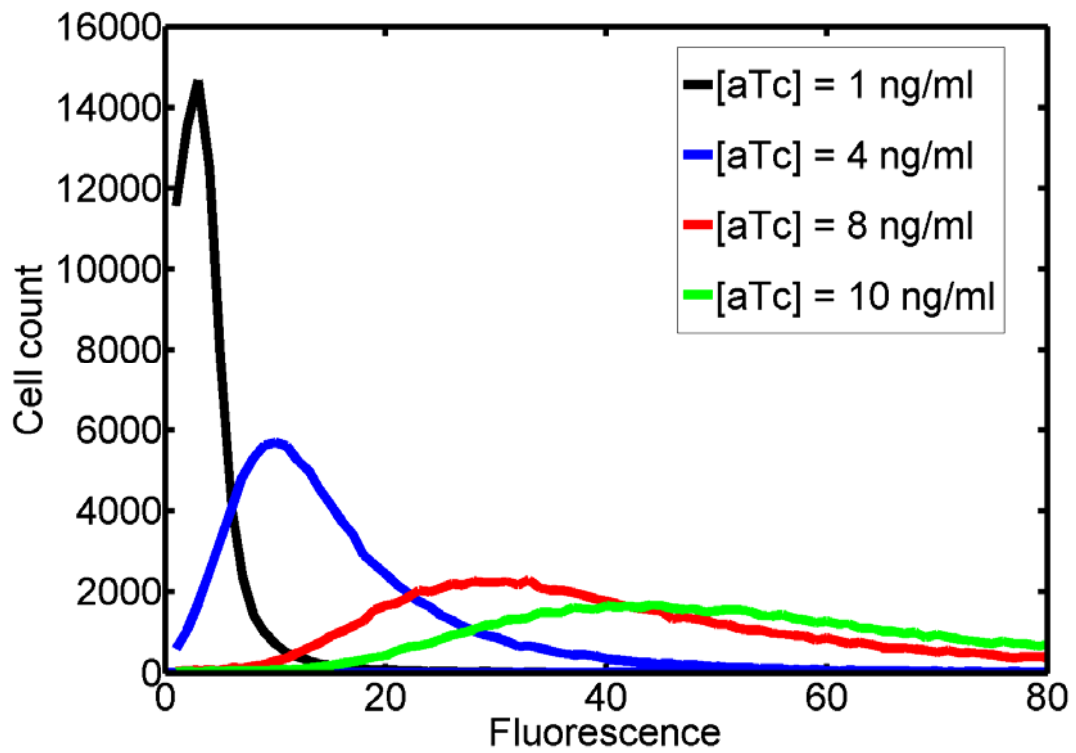


Figure 3.4: Population distribution of $P_{\text{Ltet-O1}}$ autoregulatory expression system as determined by flow cytometry. The population distribution of cells was determined across a range of expression values at representative aTc concentrations (see Fig. 3.3a). Curves correspond to [aTc] = 1 ng/ml (black), 4 ng/ml (blue), 8 ng/ml (red), 10 ng/ml (green).

3.5 Autoregulation decreases the effective sensitivity of P_L promoters

As shown in Sections 3.2-3.3, thermodynamic modeling theoretically predicts that autoregulation leads to a reporter response with reduced sensitivity compared to an equivalent system in which the repressor is supplied in constant concentration by a constitutive promoter (Figs. 3.1 and 3.2). In the previous section, we showed that simple experimental systems constructed using the synthetic inducible promoters $P_{Ltet-O1}$ and $P_{Llac-O1}$ of Lutz and Bujard¹² appear to support this theoretical prediction; we here explore this in detail.

The architectures of the promoters $P_{Ltet-O1}$ and $P_{Llac-O1}$ are very similar: both are constructed with a core promoter consisting of -10 and -35 sequences flanked by two operator sequences which bind either TetR or LacR (see Figs. 3.5), repressing expression of the driven gene. Comparison to Figs. 3.1 shows that these promoters are exactly as analyzed in Sections 3.2-3.3 with the number of operators $n_E = n_R = 2$. From Eqs. [3.2.6] and [3.3.3], we therefore expect that the sensitivity of the autoregulated system should be reduced 3 fold compared to that of the constitutive system. Indeed, the sensitivity of the autoregulated system utilizing TetR is seen to be reduced to $m_{auto} = 1.7 \pm 0.2$, a 3.5 fold reduction compared to the sensitivity of $m_{const} = 5.9 \pm 0.3$ of the constitutive system. The sensitivity of the LacR system is much less reduced: an autoregulated Hill coefficient $m_{auto} = 1.8 \pm 0.1$ reduced ~ 1.3 fold compared to the constitutive Hill coefficient $m_{const} = 2.4 \pm 0.2$. The reason for this lack of reduction in the sensitivity is unclear; however, other considerations, such as comparatively large residual affinity of induced LacR for its operators²⁵ and especially toxic effects of high over-expression of LacR (data not shown) make the LacR system generally less appealing than the TetR system.

3.6 Fine tuning for enhanced performance

In general, the ideal expression system would have low sensitivity and a large dynamic range to allow a wide range of expression values to be thoroughly and accurately measured. As seen in Fig. 3.3, the constitutive and autoregulated expression systems constructed with the “stock” $P_{\text{Ltet-O1}}$ and $P_{\text{Llac-O1}}$ promoters do indeed show reduced sensitivity; in the case of $P_{\text{Ltet-O1}}$ the sensitivity is reduced to the degree predicted by Eq. [3.3.3]. However, in both the TetR and LacR systems, the autoregulated systems show reduced fold changes ($< 100x$) with respect to the constitutive systems ($\sim 200x$ fold change), perhaps due to residual affinity of the induced repressors for their operators^{25,26}. However, as the core promoters of both $P_{\text{Ltet-O1}}$ and $P_{\text{Llac-O1}}$ are identical, the maximum amount of repressor in both the autoregulated and constitutive systems should be identical. Therefore, regardless of the residual affinity of induced repressor, both constructs should be affected to equal degrees. Another possibility is that, especially in the case of the LacR system, that the highest concentration of inducer studied was not sufficient to saturate all of the repressor in the cell (red squares, Fig. 3.3*b*). To study the full range of the LacR regulated system may require exposure to IPTG concentrations up to $> 1M$. However, the response of the TetR autoregulated system (Fig. 3.3*a*, red squares) is clearly saturated at a lower maximum expression level than that of the constitutive system (Fig. 3.3*b*, blue circles) for unknown reasons.

It would be desirable to increase the dynamic range to at least 1000x while maintaining the decreased sensitivity afforded by autoregulation. This can potentially be accomplished by modification of the promoters to adjust the parameters of the model to more favorable values. Four parameters of the system are potentially tunable through

mutagenic techniques: the basal transcription rates of the promoters, γ_E and γ_R , which can be adjusted via modification of the core promoter and/or the ribosomal binding sites, and the repressor-operator dissociation constants K_E and K_R , which can be adjusted via modification of the sequence of the operators themselves.

Of these, modification of the basal transcription rate of the reporter γ_E would result in a trivial overall vertical shift in the response curve. However, modification of the other parameters can yield more interesting results. Eq. [3.3.4] suggests that several approaches can be taken in order to increase the dynamic range: either R_{Min}^* can be increased by increasing the basal transcription rate γ_R or decreasing the strength of the operators within P_R , i.e. increasing K_R . Additionally, this goal can also be achieved by increasing the strength of the operators contained within P_E , i.e. decreasing K_E . However, Eqs. [3.3.2] - [3.3.4] describing the behavior of the autoregulatory system were derived under the assumption $[R]/K_i \gg 1$. Increasing the dissociation constant K_R could violate this assumption and will in general weaken the effect of autoregulation, potentially increasing the sensitivity. Thus, the most appealing strategies for increasing the dynamic range of these systems are to increase γ_R or decrease K_E . Examining the sequences of $P_{Ltet-O1}$ and $P_{Llac-O1}$ (Figs. 3.5), there are a number of mismatches in both the operators and the core promoter itself. Therefore, with a degree of fine tuning, the behavior of these systems can potentially be improved.

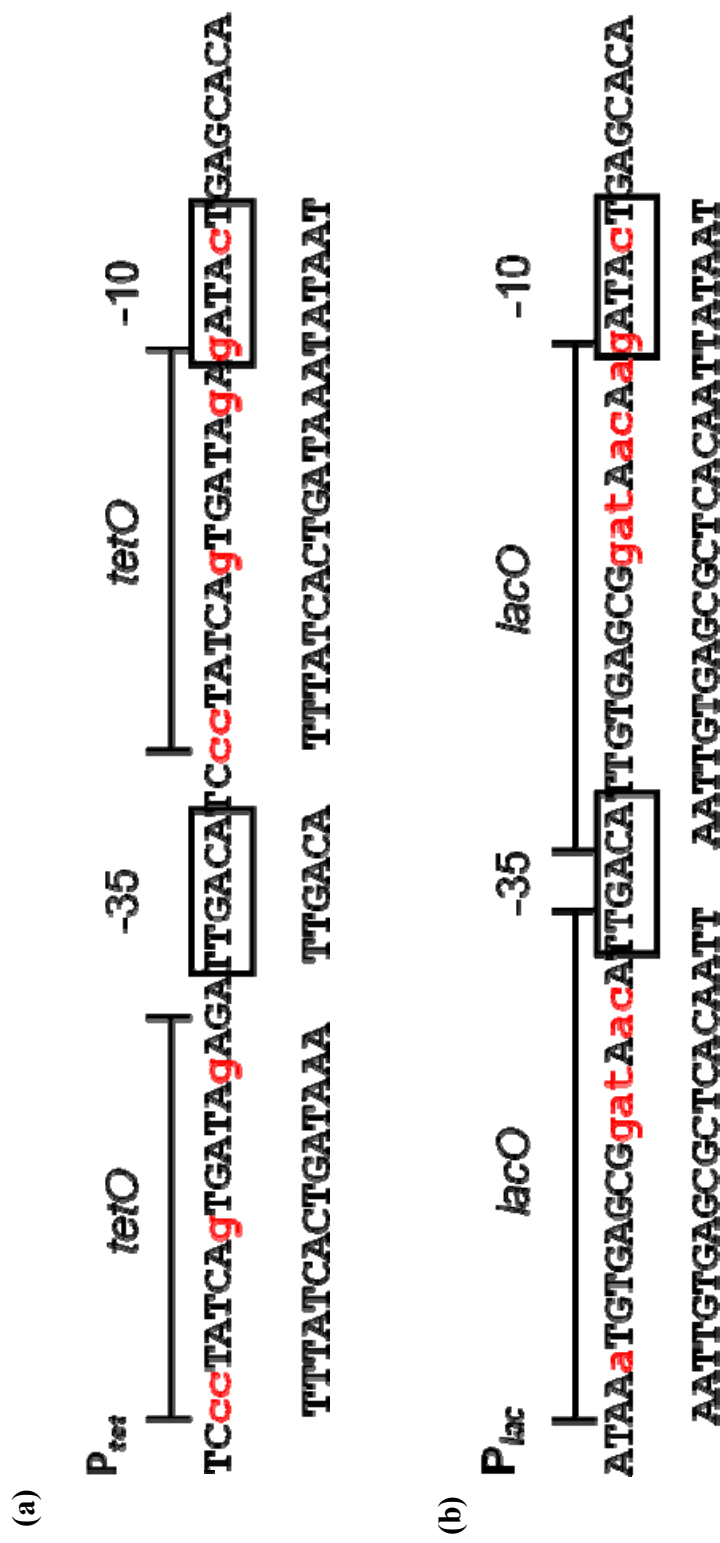


Figure 3.5: Sequences of P_{L_{tet}-O1} and P_{L_{lac}-O1} compared to consensus sequences. (a) The sequence of the P_{L_{tet}-O1} promoter (top). TetR binding operator sequences are labeled *tetO*; the core promoter -10 and -35 are boxed. (b) The sequence of the P_{L_{lac}-O1} promoter (top). LacR binding operator sequences are labeled *lacO*; the core promoter -10 and -35 are boxed. Comparison to the consensus sequences of the TetR/LacR operators and -10 and -35 sequences (bottom) reveals several mismatches (red lowercase).

3.7 Autoregulation in nature

Autoregulation is a strategy widely used in nature to regulate many transcription factors, including AraC²⁷, Crp²⁸, and TetR²⁹, among others. In particular, autoregulation is a common theme in amino acid biosynthesis. The repressors of leucine and serine^{30,31}, cysteine^{32,33}, arginine³⁴⁻³⁶, methionine^{37,38}, tryptophan³⁹, and phenylalanine, tyrosine, and tryptophan⁴⁰ biosynthesis pathways are all autoregulated to some degree. In all, roughly 40% of known transcription factors in *E. coli* negatively autoregulate themselves^{22,23}.

Why is autoregulation such a common theme in metabolic pathways? It has been suggested that autoregulation may speed the kinetics of the response to induction⁴¹. While simple transcription units composed of long-lived transcription factors require times on the order of the doubling time to reach steady state⁴², a system implementing autoregulation is able to reduce this rise-time to a fraction of the doubling time⁴¹. In addition, previous studies^{22,43} showed that autoregulation affords greater control over fluctuations in the levels of network components. As the networks involved must withstand exposure to considerable variations and random perturbations from the environment, this added stability can protect the cell against small, transient fluctuations in the environment of the cell leading to large and potentially detrimental fluctuations in levels of various components.

The present study also suggests a further advantage of autoregulation in metabolic pathways. When exposed to changes in the level of an extraneous environmental factor or potential nutrient source, it is advantageous to the cell to be able to adapt to and take advantage of or protect itself against this change in its environment. Furthermore, it is

advantageous to be able to respond to the stimulus in a subtle, graded manner; if the fluctuation is small, it would be a great waste of resources and potentially harmful to the cell to produce the proteins necessary in inappropriately large or small amounts in an extremely sensitive, all-or-none fashion. By being able to respond to greater or lesser degrees based on the level of stimulus in the environment, the cell affords itself greater control over its use of resources. However, at this point this is merely speculative; a more thorough analysis of the parameter values involved in natural autoregulatory pathways is required to determine if nature does indeed take advantage of this potential benefit.

3.8 Materials and Methods: Plasmids

pZE12-*luc*¹² was used to make the reporter plasmids pZE12L-E and pZE12T-E. Using site-directed mutagenesis, an *EcoRI* site was created by adding GAAT immediately downstream of +1 of the P_{Lac-O1} promoter. The region between the newly created *EcoRI* site and the *EcoRI* site 6 bp upstream of RBS was then deleted by *EcoRI* digestion and subsequent religation, yielding pZE12-lucM. The *KpnI*-*XbaI* flanking *luc* gene in pZE12-lucM was replaced by the *gfpmut3b* structure gene⁴⁴. This yields pZE12L-E, which harbors the P_{Lac-O1}:*gfpmut3b* construct with a 5'-UTR defined by an *EcoRI* site immediately downstream of +1 and *KpnI* site immediately upstream of the translation start of *gfpmut3b*. The 15-base sequence sandwiched by the *EcoRI* and *KpnI* sites, ATTAAAGAGGAGAAA, contains an RBS indicated by the underlined bases. pZE12S therefore contains the ColE1 *ori*, the P_{Lac-O1} promoter¹², and the coding sequence of the *gfpmut3b* gene.

To construct pZE12T-E, plasmid pZE12L-E was digested with *XhoI* and *KpnI* and the resulting plasmid fragment was purified via electrophoresis, removing the P_{Llac-O1} promoter. The *XhoI/KpnI* P_{Ltet-O1} promoter was then cloned from the plasmid pZA31¹² and religated into the purified plasmid fragment, yielding the reporter plasmid pZE12T-E.

The repressor source plasmids pZA31T-R and pZA31T-L were constructed by cloning the *tetR* or *lacI* gene into the *KpnI* and *BamHI* sites of plasmid pZA31. *lacI* was amplified using a 5' primer altered to adjust the first codon from GTG to ATG. The corresponding plasmids pZA31L-R and pZA31L-L were constructed by first removing the P_{Ltet-O1} promoter by digestion with *XhoI* and *KpnI* and purified by electrophoresis. The promoter P_{Llac-O1} was cloned from pZE12-*luc* and religated into the *KpnI* and *BamHI* sites of pZA31, yielding plasmids pZA31T-R and pZA31T-L.

3.9 Medium, growth, and measurements

TK200 (MG1655 $\Delta lacI \Delta lacY$) strains each containing the reporter and repressor plasmids were grown in M63 minimal media with standard concentrations of the appropriate antibiotics. The overnight cultures were diluted into fresh M63 media ($OD_{600} \approx 0.002$) containing the appropriate antibiotics as well as varying amounts of the inducer (aTc or IPTG) in the wells of 48-well plates. The plates were incubated with shaking at 37C and taken for OD_{600} and fluorescence measurements every hour for up to 15 hours (until a final OD_{600} of 0.2 - 0.3) using a TECAN Genios-Pro plate reader. Each measurement was repeated 3 times and the data was analyzed as discussed below.

To obtain single-cell data, 1.7 ml of each culture was spun down and resuspended in 1 ml PBS. Expression data were collected using a Becton-Dickinson FACS Calibur flow cytometer with a 488-nm argon excitation laser and a 515- to 545-nm emission filter (FL1) at a low flow rate. Photomultiplier tube (PMT) voltage was set to 950V, and a linear amplifier was set at 9.5X. Forward scatter and fluorescence values were collected for 50,000 cells.

3.10 Data analysis

In order to obtain expression patterns for the different strains, we averaged at each time point the data obtained from the different repeats for each combination of strain and inducer. Fitting $\log_2(OD_{600})$ vs. time to a linear function yielded straight lines (data not shown) whose slope gave the doubling rate (μ) for each strain and condition (yielding doubling time ~ 2 hours for most strains.) Next, for all of the time points concerning each strain and condition, we plotted the average fluorescence vs. average OD_{600} on linear-linear plot and extracted the slope (f). Each slope gave the average fluorescence per growing cell [in unit of RFU/OD] for that strain and the corresponding growth condition. The raw fluorescence production rate per cell was computed as $f\mu(1 + \mu\tau)^{45}$, taking into account the maturation kinetics of GFPmut3 (maturation half-life of ~ 30 min)⁴⁴. We then subtracted away from this raw rate the background fluorescence production rate, obtained in the same way from data collected from our negative control (TK200). This yielded the rate of GFP production synthesis and is referred to as the GFP expression.

The dependence of the GFP expression on either the concentrations of aTc or IPTG in the medium was then fitted to an effective Hill form

$$\alpha_X = b_X \cdot \frac{1 + f_X ([X] / C_X)^{m_X}}{1 + ([X] / C_X)^{m_X}} \quad [3.10.1]$$

where $[X]$ refers to the concentration of the inducer X (cAMP or IPTG) in the medium. In the fits, the basal activity b_X was directly fixed by the promoter activity obtained for $[X] = 0$. The other three Hill parameters were obtained by fitting to Eq. [3.10.1] using nonlinear least-square minimization (MATLAB 7). Reported errors are the 68.3% (1 standard deviation) confidence intervals for the given parameters.

3.11 References

1. Gilbert W, Muller-Hill B. Isolation of the Lac Repressor. *Proc Natl Acad Sci U S A*. 1966 Dec;56(6):1891-8.
2. Schleif R. Induction of the L-arabinose operon. *J Mol Biol*. 1969 Nov 28;46(1):197-9.
3. Schleif R. An L-arabinose binding protein and arabinose permeation in *Escherichia coli*. *J Mol Biol*. 1969 Nov 28;46(1):185-96.
4. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*. 2000 Jan 20;403(6767):335-8.
5. Masse E, Escorcia FE, Gottesman S. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev*. 2003;17:2374-83.
6. Lewis LK, Jenkins ME, Mount DW. Isolation of DNA damage-inducible promoters in *Escherichia coli*: regulation of polB (dinA), dinG, and dinH by LexA repressor. *J Bacteriol*. 1992 May;174(10):3377-85.
7. Lelivelt MJ, Kawula TH. Hsc66, an Hsp70 homolog in *Escherichia coli*, is induced by cold shock but not by heat shock. *J Bacteriol*. 1995 Sep;177(17):4900-7.
8. Taylor WE, Straus DB, Grossman AD, Burton ZF, Gross CA, Burgess RR. Transcription from a heat-inducible promoter causes heat shock regulation of the sigma subunit of *E. coli* RNA polymerase. *Cell*. 1984 Sep;38(2):371-81.
9. Epstein W, Rothman-Denes LB, Hesse J. Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 1975 Jun;72(6):2300-4.
10. Masse E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2002 Apr 2;99(7):4620-5.
11. Rouch DA, Brown NL. Copper-inducible transcriptional regulation at two promoters in the *Escherichia coli* copper resistance determinant *pco*. *Microbiology*. 1997 Apr;143 (Pt 4):1191-202.
12. Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*. 1997 Mar 15;25(6):1203-10.

13. Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD. Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology*. 2001 Dec;147(Pt 12):3241-7.
14. Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci U S A*. 1957;43:553-66.
15. Siegele DA, Hu JC. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc Natl Acad Sci U S A*. 1997 Jul 22;94(15):8168-72.
16. Oehler S, Alberti S, Muller-Hill B. Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Res*. 2006;34(2):606-12.
17. Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev*. 2005 Apr;15(2):136-44.
18. Khlebnikov A, Risa O, Skaug T, Carrier TA, Keasling JD. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J Bacteriol*. 2000 Dec;182(24):7029-34.
19. Buchler NE, Gerland U, Hwa T. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A*. 2003 Apr 29;100(9):5136-41.
20. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev*. 2005 Apr;15(2):125-35.
21. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*. 2005 Apr;15(2):116-24.
22. Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. *Nature*. 2000 Jun 1;405(6786):590-3.
23. Thieffry D, Huerta AM, Perez-Rueda E, Collado-Vides J. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays*. 1998 May;20(5):433-40.
24. Ozbudak EM, Thattai M, Lim HN, Shraiman BI, Van Oudenaarden A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*. 2004 Feb 19;427(6976):737-40.

25. Gilbert W, Muller-Hill B. The Lactose Repressor. In: Beckwith J, Zipser D (eds). The Lactose Operon. Cold Spring Harbor NJ: Cold Spring Harbor Press, 1970:93-109.
26. Kamionka A, Bogdanska-Urbaniak J, Scholz O, Hillen W. Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Res.* 2004;32(2):842-7.
27. Lee NL, Gielow WO, Wallace RG. Mechanism of *araC* autoregulation and the domains of two overlapping promoters, *P_c* and *P_{BAD}*, in the L-arabinose regulatory region of *Escherichia coli*. *Proc Natl Acad Sci U S A.* 1981 Feb;78(2):752-6.
28. Aiba H. Autoregulation of the *Escherichia coli* *crp* gene: CRP is a transcriptional repressor for its own gene. *Cell.* 1983 Jan;32(1):141-9.
29. Saenger W, Orth P, Kisker C, Hillen W, Hinrichs W. The Tetracycline Repressor-A Paradigm for a Biological Switch. *Angew Chem Int Ed Engl.* 2000 Jun 16;39(12):2042-52.
30. Lin R, D'Ari R, Newman EB. Lambda *placMu* insertions in genes of the leucine regulon: extension of the regulon to genes not regulated by leucine. *J Bacteriol.* 1992 Mar;174(6):1948-55.
31. Wang Q, Wu J, Friedberg D, Plakto J, Calvo JM. Regulation of the *Escherichia coli* *lrp* gene. *J Bacteriol.* 1994 Apr;176(7):1831-9.
32. Bielinska A, Hulanicka D. Regulation of the *cysB* gene expression in *Escherichia coli*. *Acta biochimica Polonica.* 1986;33(2):133-7.
33. Kredich NM. The molecular basis for positive regulation of *cys* promoters in *Salmonella typhimurium* and *Escherichia coli*. *Molecular microbiology.* 1992 Oct;6(19):2747-53.
34. Lim DB, Oppenheim JD, Eckhardt T, Maas WK. Nucleotide sequence of the *argR* gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor. *Proc Natl Acad Sci U S A.* 1987 Oct;84(19):6697-701.
35. Szwajkajzer D, Dai L, Fukayama JW, Abramczyk B, Fairman R, Carey J. Quantitative analysis of DNA binding by the *Escherichia coli* arginine repressor. *J Mol Biol.* 2001 Oct 5;312(5):949-62.
36. Maas WK. The arginine repressor of *Escherichia coli*. *Microbiol Rev.* 1994 Dec;58(4):631-40.

37. Saint-Girons I, Duchange N, Cohen GN, Zakin MM. Structure and autoregulation of the metJ regulatory gene in *Escherichia coli*. *J Biol Chem*. 1984 Nov 25;259(22):14282-5.
38. Urbanowski ML, Stauffer GV. Autoregulation by tandem promoters of the *Salmonella typhimurium* LT2 metJ gene. *J Bacteriol*. 1986 Mar;165(3):740-5.
39. Klig LS, Carey J, Yanofsky C. trp repressor interactions with the trp aroH and trpR operators. Comparison of repressor binding in vitro and repression in vivo. *J Mol Biol*. 1988 Aug 20;202(4):769-77.
40. Camakaris H, Pittard J. Autoregulation of the tyrR gene. *J Bacteriol*. 1982 Apr;150(1):70-5.
41. Rosenfeld N, Elowitz MB, Alon U. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol*. 2002 Nov 8;323(5):785-93.
42. Monod J, Pappenheimer AM, Jr., Cohen-Bazire G. [The kinetics of the biosynthesis of beta-galactosidase in *Escherichia coli* as a function of growth.]. *Biochim Biophys Acta*. 1952 Dec;9(6):648-60.
43. Savageau MA. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature*. 1974 Dec 13;252(5484):546-9.
44. Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*. 1996;173(1 Spec No):33-8.
45. Leveau JH, Lindow SE. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J Bacteriol*. 2001;183:6752-62.

CHAPTER 4

Quantitative Characteristics of Gene Regulation by Small RNA

4.1 Introduction

Small non-coding RNAs (sRNA) have been demonstrated in recent years to play central regulatory roles in prokaryotes and eukaryotes¹⁻⁴. Organisms that employ sRNA in post-transcriptional regulation range from bacteria to mammals. Interestingly, they are predominantly implicated in regulating critical pathways, such as stress responses in bacteria⁵⁻⁸, or developmental timing and cell differentiation in plants and metazoans^{9,10}. What makes these regulators unique and indispensable is only starting to be unveiled.

The regulatory role of small RNAs in bacteria has been a subject of research for the last several decades. Early on, sRNAs were mainly recognized for their specialized roles in controlling the transposition of insertion elements^{11,12}, in regulating plasmid copy number during plasmid replication¹³⁻¹⁶, and in mediating plasmid maintenance through the toxin-antidote system¹⁷. These antisense RNAs are typically encoded on the opposite strand, *in cis* with their targets^{16,18}. In accord with their biological function¹⁸, some of these RNAs are metabolically stable (*e.g.* the ones controlling transposition) while others are very unstable (such as the ones controlling plasmid copy number). For the latter, it has been demonstrated that the strength of inhibition is strongly related with the binding rate of the antisense and its target, rather than the binding affinity^{19,20}.

Until recently, only a few cases involving chromosomally-encoded sRNA were known^{21,22}. The advent of large-scale experimental techniques²³⁻²⁵ and bioinformatic

methods ²⁵⁻²⁸ has led to the identification and the subsequent verification of numerous chromosomal sRNAs in a variety of bacterial species in the past five years. Currently, there are over 70 identified sRNAs in *E. coli* ^{6,8,29}. Like regulatory proteins, many of these sRNAs act *in trans*, regulate the expression of multiple target genes, and are themselves regulated by one or more factors. They have been found to play important roles in the regulation of important pathways including oxidative response ³⁰, osmotic response ^{22,31}, acid response ^{32,33}, quorum sensing ⁷, and more ^{5,6,8}.

The mechanisms by which sRNAs exert their effect are diverse. Most known bacterially encoded sRNAs act *in trans* by binding at the 5' untranslated region (UTR) of a target mRNA ^{2,3,6}, with specificity achieved through base-pairing between the two RNA molecules. Upon binding, the small RNA can reduce the efficiency of translation initiation – *e.g.* by interfering with ribosomal binding - or the stability of the target mRNA. This class of sRNAs include RyhB (regulator of iron metabolism) ^{34,35,36}, OxyS (oxidative stress) ³⁰, MicC and MicF (osmotic stress) ^{22,31}, and many others. Conversely, RprA and DsrA promote translation of their target, *rpoS* (encoding the stationary phase sigma factor σ^s) ⁵, while GadY – the only sRNA known to bind the 3'-UTR of its target – stabilizes its target ³³.

A large class of sRNAs bind tightly to the RNA chaperone Hfq, a highly abundant protein which also binds the target mRNA in a number of cases studied ^{30,37-39}. Binding to Hfq may protect these RNA molecules from degradation ^{40,41}. While most mRNAs in bacteria are short-lived, Hfq-bound mRNAs are atypically stable, thereby allowing for the regulation of their turnover rates. Hfq has also been shown to facilitate the pairing of the sRNA with its target mRNA ³⁶, leading to the inhibition of translational initiation. In

turn, pairing of the two RNA molecules exposes both molecules to rapid degradation. Importantly, the interaction between the sRNA and its target is non-catalytic in nature, since a given sRNA molecule may be degraded along with its target, and not recycled to regulate other targets³⁵ (Fig. 4.1).

Are there special features of sRNA-mediated regulation that make it especially suitable for intricate regulation? We address this question in the context of two well-characterized bacterial sRNAs in *E. coli*. Using a combination of experimental and theoretical approaches, we demonstrate that sRNA provides a novel mode of gene regulation, featuring a tunable threshold-linear expression pattern, a robust noise resistance characteristic, and a built-in capability for hierarchical cross talk. These properties are attributed to the non-catalytic nature of sRNA-mediated regulation, which make it qualitatively distinct from the common protein-mediated gene regulation.

4.2 Coupled degradation leads to the threshold-linear response.

In this study we focus on small RNAs which regulate gene expression through base-pairing with mRNA molecules of their target genes. The formation of RNA duplex may lead to inhibition of translation and accelerated degradation of the target. In addition, the best studied class of small RNAs in *E. coli* were shown to be degraded along with their target mRNA³⁵, i.e., the degradation of the target mRNA is *coupled* to the degradation its regulator.

This non-catalytic nature of sRNA-mediated regulation suggests a novel “threshold-linear” mode of action, by which the expression of, e.g., a stress-response gene is silenced below a *threshold*, and is gradually activated above it. As suggested recently by Lenz et al.⁷ and illustrated in Fig. 4.1a, if the transcription rate for the target mRNA (α_m) is below that for the sRNA (α_s), then most of the targets are expected to pair with the sRNAs and be rapidly degraded. Conversely, if the transcription rate of the mRNA exceeds that of the sRNA (Fig. 4.1b), then most of the sRNAs are expected to turnover, while the unconsumed mRNAs are free to be translated into proteins. In this latter case, the expressed protein level would reflect the difference between the two transcription rates. This scenario is summarized by the blue line in Fig 4.1c, where the steady state mRNA level of the target gene (m) is plotted against its transcription rate. Messenger RNAs start to accumulate only when the target transcription rate exceeds the threshold, given by the transcription rate of the sRNA (vertical dashed line). More generally, it is possible that the degradation of the mRNA in the complex does not always lead to the co-degradation of the sRNA. Suppose the probability that the sRNA is degraded upon the degradation of the mRNA is $p < 1$. Then as discussed next, a target transcription rate that is larger by a factor $1/p$ would be required to switch from the silenced to the linear range.

The above qualitative model can be formulated quantitatively via a simple kinetic model for sRNA-mediated gene silencing. The model is cast in terms of two mass-action equations^{7,42} for the cellular concentrations of the sRNA (s) and its target mRNA (m):

$$\begin{aligned}\frac{dm}{dt} &= \alpha_m - \beta_m m - kms \\ \frac{ds}{dt} &= \alpha_s - \beta_s s - pkms\end{aligned}\tag{4.2.1}$$

In this model, transcription of mRNAs and sRNAs are characterized by the rates α_m and α_s , and their turnover by rates β_m and β_s respectively. The coupled degradation between sRNA and its target is assumed to be stoichiometric, and is thus described through a second-order kinetic constant, k , and the probability p that the sRNA is also lost in the process. The levels of Hfq and any endoribonuclease involved are assumed to be at saturation and are not tracked explicitly.

The predicted pattern of gene expression is obtained by solving Eq. [4.2.1] in the steady state,

$$m = \frac{1}{2\beta_m} \left[(\alpha_m - \alpha_s / p - \lambda) + \sqrt{(\alpha_m - \alpha_s / p - \lambda)^2 + 4\alpha_m \lambda} \right], \tag{4.2.2}$$

expressed in terms of the two control variables, α_m and α_s / p , and an effective parameter $\lambda = (\beta_m \beta_s) / (kp)$ which balances the rate of mRNA turnover *not* due to the sRNA and the mutual degradation rate, and referred to as the “leakage rate”. This leakage rate is an (inverse) measure of the strength of sRNA-mRNA interaction. For strong, rapid sRNA-mRNA interactions, the leakage rate is small and the solution [4.2.2] is given

approximately by $m \approx \frac{1}{\beta_m} (\alpha_m - \alpha_s / p) + \frac{\alpha_s}{\beta_m} \frac{\lambda}{\alpha_m - \alpha_s / p}$ above the threshold $(\alpha_m > \alpha_s / p)$, and $m \approx \frac{\alpha_m}{\beta_m} \frac{\lambda}{\alpha_s / p - \alpha_m}$ below. With $\lambda \rightarrow 0$, this is just the threshold-

linear function depicted by the blue line of Fig 4.1*c*. For small but finite λ , the mRNA level is somewhat larger, especially near the threshold (where the denominators of the λ -terms become small). Thus, leakage serves to make the transition smoother, as illustrated by the red line of Fig. 4.1*c*, but does not change the qualitative feature of the threshold-linear form. We note that the value of the threshold (α_s/p) is set by the sRNA transcription rate and is hence a “dynamic variable” controllable by the genetic control circuit (rather than being a fixed quantity encoded by the genomic sequence.) In particular, the threshold value is not affected by the strength of the interaction parameter k (as long as the leakage is reasonably small to preserve the threshold-linear form). With $p < 1$, corresponding to the case where a portion of the sRNA may be recycled, the transition shifts to higher mRNA expression levels and becomes smoother (via increasing the leakage rate).

The kinetic model [4.2.1] may offer quantitative predictions, given the values of the different parameters. Below, we will apply the model to the small RNA RyhB^{34-36,43}, which is one of the best characterized sRNA and for which ample kinetic data exists for us to infer realistic values for all of the model parameters; the results are summarized in Table 4.1. From these parameter values, we estimate a leakage rate $\lambda \approx 0.1$ nM/min for RyhB. This is small but non-negligible compared with the relevant range of the transcription rates, α_m and α_s . In fact, the smoothened threshold-linear expression pattern plotted in red in Fig. 4.1*c* is the steady state solution Eq. [4.2.2] with the RyhB parameters listed in Table 4.1.

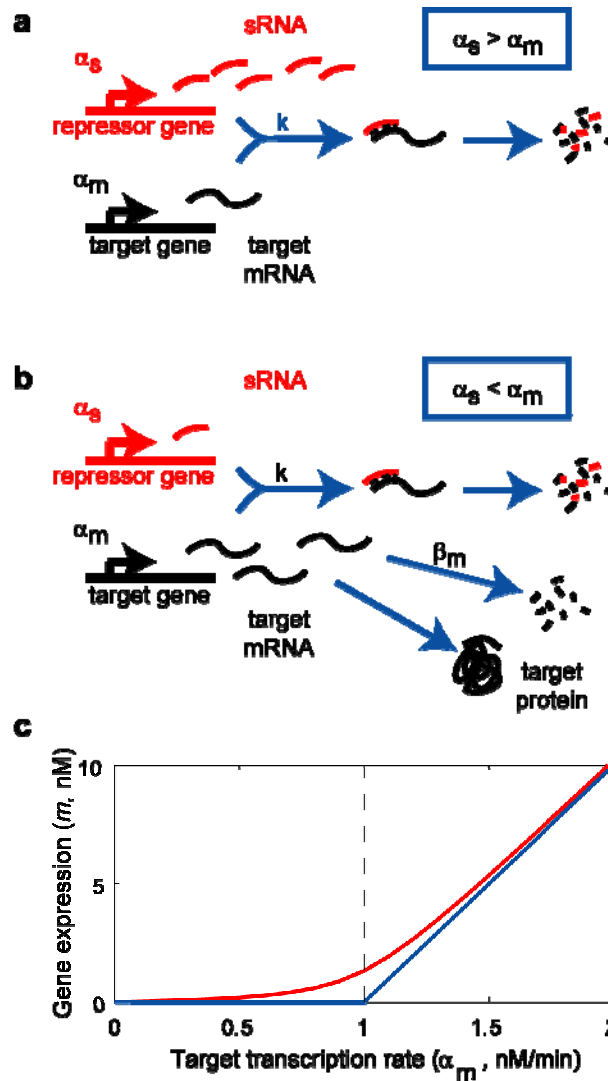


Figure 4.1. Threshold-linear response of a target gene. (a) and (b). Model for the interaction between mRNAs of a target gene and a small RNA. In the idealized scenario, when the small RNA transcription rate is larger than that of the target (a), gene expression is silenced, whereas if sRNA is transcribed less efficiently than its target (b), the unbound mRNAs code for proteins. (c). Response curve of a target gene. The blue line depicts the idealized threshold-linear mode of regulation in which gene expression is silenced if the target transcription rate is below a *threshold* set by the transcription rate of the sRNA (indicated by the dashed line). Above this threshold, gene expression increases linearly with the difference between the mRNA and sRNA transcription rates. The idealized scenario is expected when binding between sRNA and mRNA occurs extremely rapidly. The red line given by Eq. [4.2.2], with the estimated parameters of Table 4.1 ($\alpha_s = 1$ nM/min).

Table 4.1: Model Parameters

Parameter	Meaning	Estimated value	
		RyhB/sodB	RNA-IN/RNA-OUT
α_m	Transcription rate of target mRNA	$\frac{1}{10} - 1$ nM/min	$\frac{1}{250} - \frac{1}{25}$ nM/min
α_s	Transcription rate of sRNA	$\frac{1}{10} - 10$ nM/min	1-10 nM/min
β_m	Decay rate of free mRNA	$\frac{1}{10}$ (min) ⁻¹	$\frac{3}{10}$ (min) ⁻¹
β_s	Decay rate of free sRNA	$\frac{1}{50}$ (min) ⁻¹	$\frac{1}{50}$ (min) ⁻¹
k	Binding rate of sRNA:mRNA complex	$\frac{1}{50}$ (nM min) ⁻¹	$\frac{1}{50}$ (nM min) ⁻¹

4.3 Quantitative experimental characterization of RyhB

To validate model Eq. [4.2.1], we test experimentally its direct prediction, namely a smoothened threshold-linear response function. To this end, we characterized quantitatively the response of a target gene regulated by the small RNA RyhB. RyhB is involved in regulation of iron homeostasis in *E. coli*, and is expressed at low cellular iron levels. Its targets include iron-storage and oxidative response genes^{34,43} whose expressions are needed to combat problems associated with elevated iron levels, but not when iron is deficient⁴⁴.

To circumvent the complex regulation of the endogenous system, we constructed a synthetic target gene, consisting of the 5' control-region of *sodB* (which is the strongest known natural target of RyhB^{35,43}), transcriptionally fused to the coding sequence of the reporter *gfp*. The target gene, *crsodB-gfp*, was driven by an inducible *lac* promoter, P_{Lac-O1} ⁴⁵, and placed on the multicopy plasmid pZE12S; see Materials and Methods.

This construct allowed us to control the transcription rate of the target, α_m , by changing the concentration of the inducer IPTG in the medium. To quantify the relation between IPTG concentration and target transcription rate, we first characterized the bare target expression by transforming the pZE12S plasmid into a *ryhB*⁻ strain of *E. coli* BW-RI. Expression of the target was assayed by measuring GFP fluorescence in the resulting cells grown in minimal M63 medium with various amounts (0 – 0.5 mM) of IPTG (see Fig. 4.2 and its caption). For each concentration of IPTG, we use this fluorescence reading to define the “*promoter activity*”.

We then repeated these measurements in cells harboring pZE12S and either chromosomal or plasmid-encoded *ryhB*. The expression rate of the small RNA, α_s , was

controlled by a variety of means, as detailed below. The GFP expression obtained for each level of RyhB expression (Fig. 4.2) was then plotted against the promoter activity at the corresponding IPTG level; see Fig. 4.3*a*.

As a calibration, we first observe that GFP expressions in wild-type cells grown in media with 100 μ M FeSO₄ (Fig. 4.3*a*, red circles) are indistinguishable from those of the *ryhB*⁻ strain (Fig. 4.3*a*, dashed black line), indicating the complete repression of RyhB activity at such a high iron level as expected. For the same wild-type (*ryhB*⁺) cells grown in media with no added Fe (red crosses in Fig. 4.3*a*), GFP expressions were moderately reduced across all IPTG levels, and more so for strain carrying the pZA30R plasmid which harbors *ryhB* driven by its native promoter (Fig. 4.3*a*, green crosses). These results are qualitatively consistent with the expected increase of RyhB expression upon reducing the iron level and upon adding multi-copy plasmid-bound sources.

To see the effect of even higher RyhB expression, we used the pZA31R plasmid which harbors the *ryhB* gene driven by the strong synthetic P_{Ltet-O1} promoter⁴⁵ inducible by aTc. Reduction in GFP expression was seen already in medium with no aTc added (blue circles), due presumably to the leaky expression of this very small gene on a multi-copy plasmid. The addition of small amounts of aTc (blue crosses, squares, and asterisks) drastically reduced GFP expression (up to 30-fold reduction compared to the *ryhB*⁻ strain). In sum, using combination of chromosome and plasmid sources for RyhB, we measured the response of the target gene to its promoter activity in the presence of six different levels of expression of the small RNA.

In order to verify that indeed RyhB regulation is achieved mainly by changes in target mRNA level, we quantified the mRNA concentration directly at two levels of

sRNA expression using quantitative real-time RT-PCR (Fig. 4.3b). We find that reduction in mRNA level is consistent but smaller than the corresponding reduction in fluorescence, suggesting some level of translational repression (see below).

In strains deleted of *hfq*, GFP expression was reduced ~3-fold. Moreover, aTc dependence was completely abolished and GFP expression became the same in *ryhB*⁺ and *ryhB*⁻ strains (Fig. 4.3c), indicating that *hfq* is required for the repression effect observed here. This behavior is expected for a RyhB-target, since RyhB accumulation and RyhB-target interaction requires Hfq³⁴⁻³⁶. As a control, we show in Fig. 4.3c the GFP expression obtained for a *hfq*⁺ strain in which the plasmid pZE12S was replaced by pZE12G, harboring the same P_{Llac-O1}:*gfp* reporter construct except that the 5'-UTR of the *gfp* gene was a short 27-base segment contained in the UTR of the original pZE12 plasmid instead of from *sodB*; see Methods. The independence of the observed GFP activity to the different degrees of RyhB expression indicates that the *sodB* control region is required for interaction.

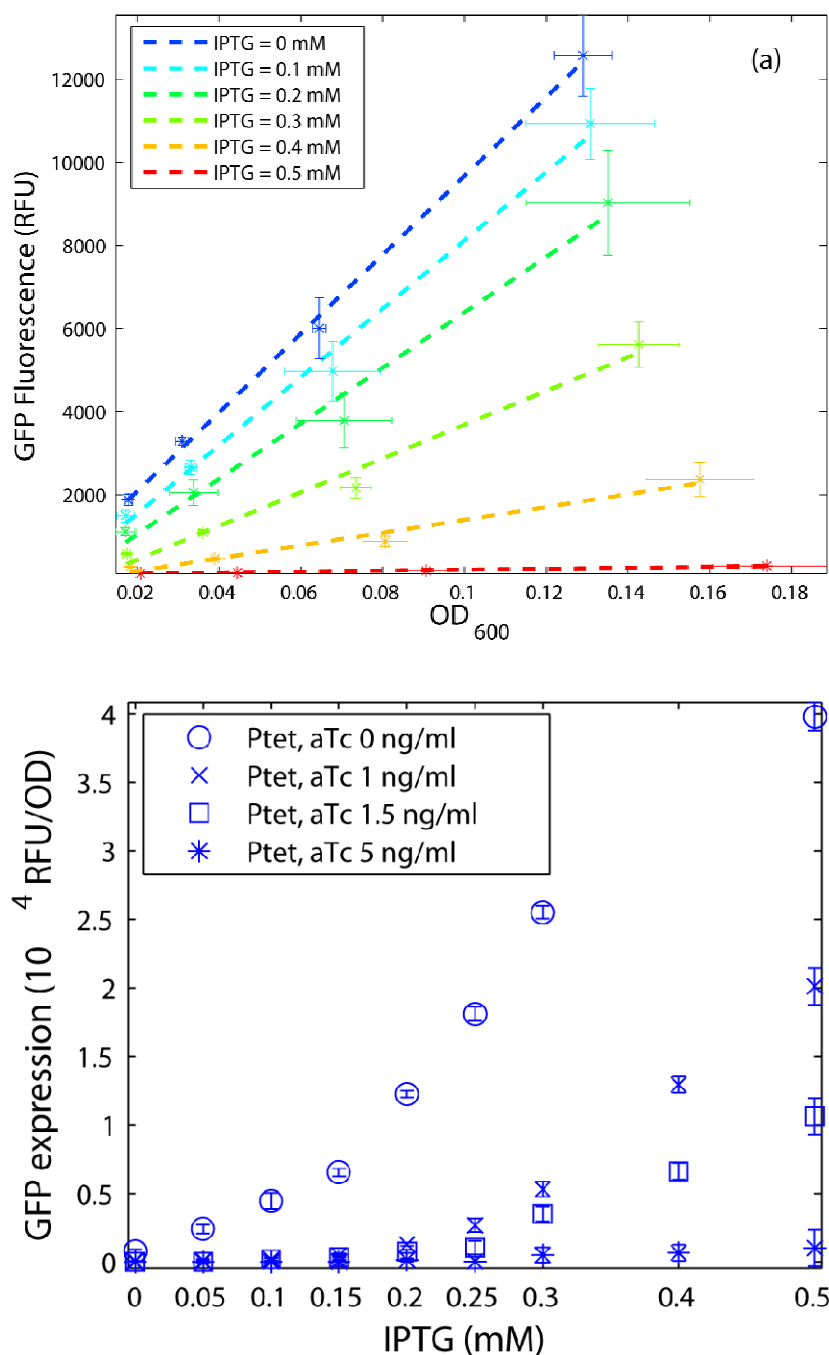


Figure 4.2: Example of raw data used to compile Fig. 4.3a. (a) GFP fluorescence is plotted against OD₆₀₀ for ryhB⁻ strain (ZZS21). Lines are given by a linear fit. **(b)** Rate of GFP production for the strains of Fig. 4.3a as a function of IPTG concentration. See Materials & Methods for details.

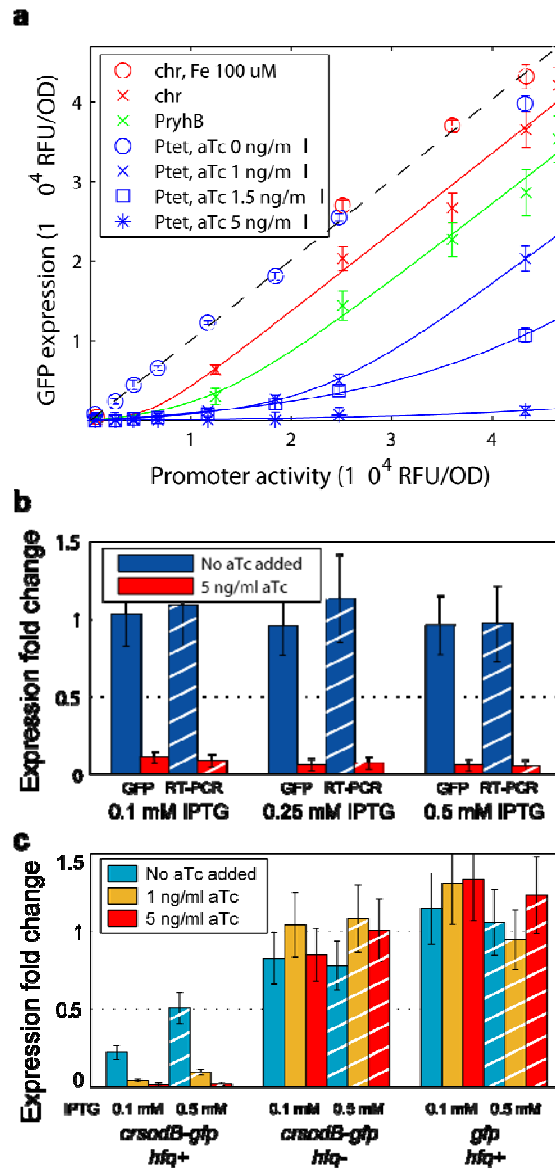


Figure 4.3. Threshold-linear response of a reporter target of RyhB. (a). GFP expressions of various *ryhB*⁺ strains (blue: strain ZGS22, green: ZGS24, red: ZGS23) are plotted against the “promoter activity”, defined as the GFP expression of the *ryhB*⁻ strain (ZGS21) grown in identical medium; see Table 4.4 for information on the strains. The curves are given by a *single* parameter fit of the data to the steady-state solution of Eq. [4.2.1], as explained in the text and Table 4.2. **(b).** Ratio of GFP expression in *ryhB*⁺ (ZGS23) and *ryhB*⁻ (ZGS21) strains measured through GFP fluorescence (dark bars) and RT-PCR (light bars). As expected, most of *ryhB* repression activity is obtained through reduction in mRNA level. **(c).** Different RyhB levels (synthesized from *Ptet-ryhB* driven by different levels of aTc) do not significantly change GFP expression in *hfq*⁻ strain (ZGS23q), or when *gfp* is placed immediately downstream of its promoter (ZGS13). Presented data were obtained using IPTG at 0.1 mM (solid bars) and 0.5 mM (striped bars).

Table 4.2: Best-fit parameters of the data in Fig. 4.3a: Parameters extracted from the data in Fig. 4.3a to model Eq. [4.2.1], given in terms of 50% confidence interval

RyhB – <i>crsodB-gfp</i> (Fig. 4.3a)	
Fitted values for α_s	
Chr- <i>ryhB</i>	0.7 ± 0.17
PryhB- <i>ryhB</i>	1.3 ± 0.14
Ptet- <i>ryhB</i> , aTc 1 ng/ml	2.5 ± 0.13
Ptet- <i>ryhB</i> , aTc 1.5 ng/ml	3.7 ± 0.19
Ptet- <i>ryhB</i> , aTc 5 ng/ml	10.4 ± 3.4
Fitted Value for a_λ	0.15 ± 0.03

4.4 Reporter target of RyhB exhibits threshold-linear expression

To make the comparison with model Eq. [4.2.1], the data in Fig. 4.3a were fitted to the steady-state solution of the model, Eq. [4.2.2]. This fit requires a single global parameter (associated with the leakage rate) and one additional free parameter (corresponding to the sRNA expression level, α_s/p) per curve. The latter characterizes the position of the softened threshold, and is listed in Table 4.2 for each RyhB source studied. The corresponding best-fit curves are shown as the colored lines in Fig. 4.3a.

The data of Fig. 4.3a revealed a spectrum of gene expression patterns: For low RyhB expressions (blue circles and red crosses), expression of the target gene was mainly controlled by its promoter activity, while for high RyhB expression (blue asterisks), target expression was greatly reduced regardless of the promoter activity. The threshold-linear mode of action is best seen for intermediate RyhB expressions (green and blue crosses, blue squares) – target expression was strongly repressed at low transcription levels, but increased linearly (and with the same slope) with promoter activity at high transcription levels for the different RyhB sources.

Our model predicts that the strength of repression exhibited at a given transcription rate of the small RNA, should decrease as the transcription rate of the target gets larger. This prediction is verified by the above results for the multicopy synthetic target gene. We have further verified that this behavior is retained when the synthetic target is encoded in a single copy on the chromosome (Fig. 4.4).

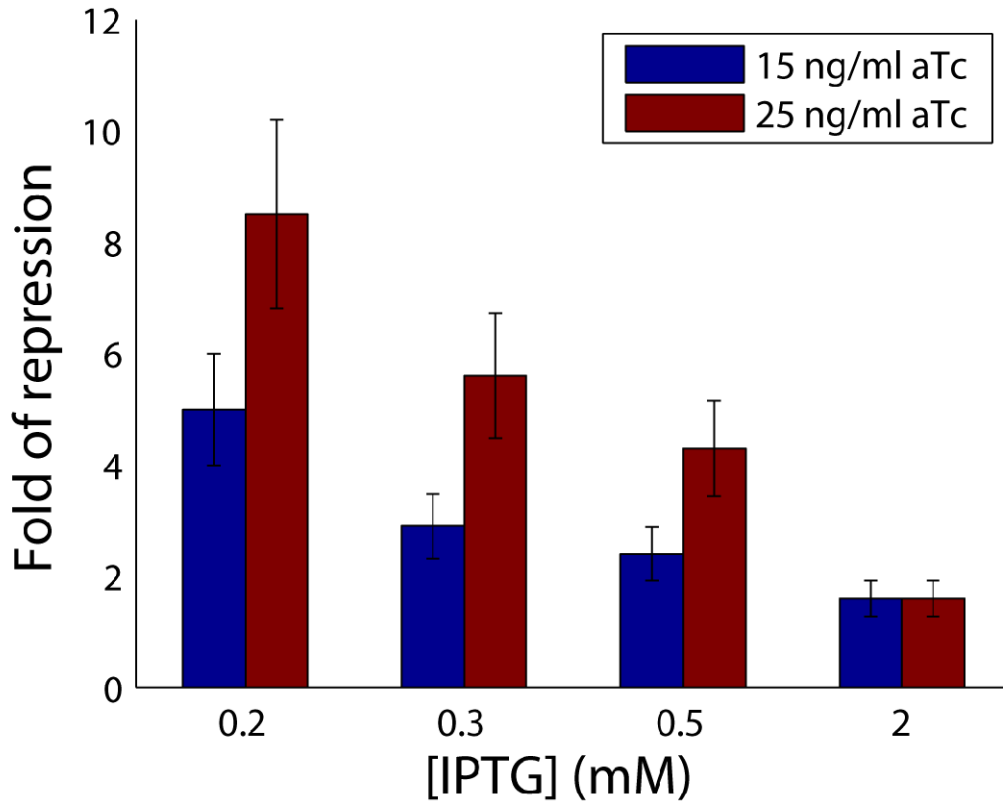


Figure 4.4. Activity-dependent repression of chromosomal target. Expression of a single-copy target gene, *crsodB-gfp*, inserted to the chromosome was measured by flow-cytometry. GFP expression of strains with (ZZS43) and without (ZZS41) plasmid-borne *ryhB* is compared. The regulatory effect of RyhB is characterized by the fold of repression, defined as

$$\frac{\text{GFP expression}(ryhB+) - \text{GFP background expression}}{\text{GFP expression}(ryhB-) - \text{GFP background expression}}$$

Here background expression is measured in cells without the target chromosomal insertion (ZZS01). These results show that the non-linear effect of RyhB can also be seen in for a single-copy target gene. See Section 4.11, *Materials & Methods* for details.

4.5 Irreversible binding suffices for threshold-linear response

Small RNAs do not need to promote directly degradation of their target mRNAs in order to silence gene expression. Binding of the sRNA at or near the ribosomal binding site may block translation initiation, making the RNA duplex inert. This binding step may be considered irreversible even if the formation of the RNA complex does not accelerate degradation, as long as the duplex does not dissociate into the two intact RNA components at relevant time scales. In this case, the system can still be described by the model Eq. [4.2.1], if we identify m and s in Eq. [4.2.1] as the *free* mRNA and sRNA concentrations, and set $p=1$. We thus expect the same smoothened threshold-linear response as described above. Here it is the irreversibility of binding – rather than coupled degradation – that implements the non-catalytic mode of regulation.

We tested this hypothesis using RNA-OUT, one of the best characterized antisense sRNA that regulates the transposition of the IS10 insertion element in *E. coli*^{11,12}. The transposase gene, which we refer to as *is10in*, is driven by the *pIN* promoter. Located only 35 bases downstream on the opposite strand is the *pOUT* promoter, which drives the transcription of the gene *is10out* encoding RNA-OUT. The perfect base pairing between the two RNA molecules at the 5'-UTR of *is10in* leads to a strong irreversible binding⁴⁶, which represses the translation of *is10in*⁴⁷ with only mild effect on its stability⁴⁸. Based on data from previous experiments, we estimated parameters for this sRNA-target pair – see Table 4.1. Most of these parameters are similar to those obtained for RyhB and its targets. However, the degradation rate of transposase mRNA, β_m , is larger than the corresponding rate of typical RyhB targets⁴⁹,

making the leakage rate leakage rate λ larger. We therefore expect *is10in* to exhibit a smoother threshold-linear expression pattern.

To measure the effect of repression by RNA-OUT, we constructed a synthetic target consisting of a modified *is10in* control region transcriptionally fused to *gfp*. The control region we use differs from that of the native *is10in* in two nucleotide positions, making its RBS stronger (see Ref. 49 and Section 4.11, *Materials & Methods*). The target gene *cris10-gfp* was inserted immediately downstream of the *lac* promoter in plasmid pZE12IN. Promoter activities at 8 levels of IPTG (0-0.75 mM) were established as before by measuring GFP fluorescence, in a strain transformed with pZE12IN and the control plasmid pZA31-*luc* which does not express RNA-OUT.

As a controlled source of RNA-OUT, we used the pZA31O plasmid which harbors the *is10out* gene driven by the strong synthetic $P_{\text{Ltet-O1}}$ promoter⁴⁵. We measured the response function at four different expression levels of RNA-OUT using different concentrations of the inducer aTc (0, 2, 6 and 10 ng/ml), and fitted the data to the steady state of model Eq. [4.2.1], as described above (best-fit parameters are given in Table 4.3). The data and the corresponding fitted curves are presented in Fig. 4.5a. In absence of the inducer, RNA-IN expression coincides with that of the corresponding strain with no RNA-OUT source. At higher levels of inducer the threshold-linear response is recovered, displaying a smooth transition as expected. These results validate the prediction that coupled degradation is not necessary for the threshold-linear form, as long as coupling between the sRNA and its target is irreversible.

To verify that RNA-OUT repressed translation of *cris10-gfp* mRNA without significantly altering its accumulation, we quantify the mRNA concentration of *cris10-*

gfp using RT-PCR. Indeed, with 0.3 and 0.75 mM IPTG mRNA levels did not change more than 2-fold under different levels of aTc (in *is10out+* strains) and in *is10out-* strain (Fig. 4.5*b*).

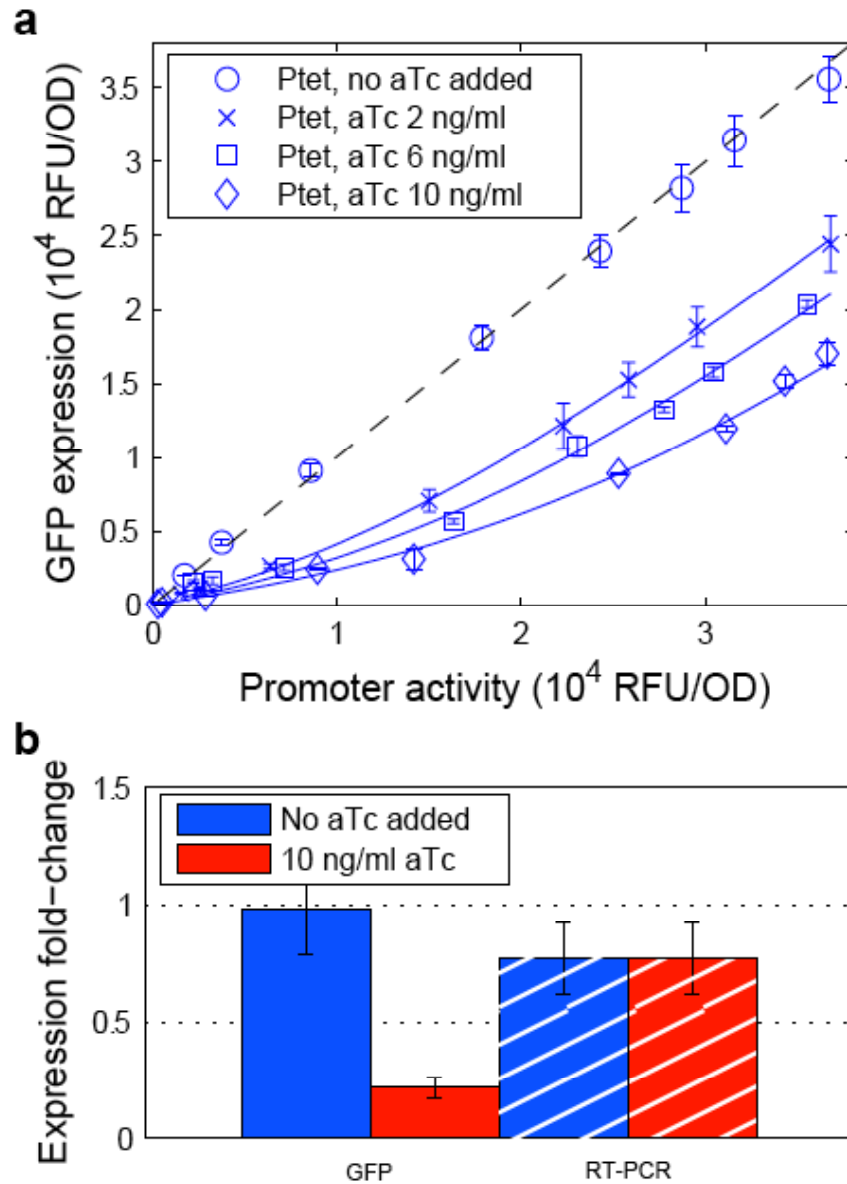


Figure 4.5. Threshold-linear response of a reporter target of RNA-OUT. (a). GFP expressions of *is10out*⁺ strain (ZZS35) are plotted against the “promoter activity”, defined as the GFP expression of the *is10out*⁻ strain (ZZS31) grown in identical medium; see caption of Fig. 4.3 for details. (b). Ratio of GFP expression in *is10out*⁺ (ZZS35) and *is10out*⁻ (ZZS31) strains measured through GFP fluorescence (solid bars) and RT-PCR (striped bars). As expected, changes in mRNA level of *cris10-gfp* are insignificant as compared with GFP protein level.

Table 4.3: Best-fit parameters of the data in Fig. 4.5a: Parameters extracted from the data in Fig. 4.5a to model Eq. [4.2.1], given in terms of 50% confidence interval

RNA-OUT – <i>cris10-gfp</i> (Fig. 4.5a)	
Fitted values for α_s	
Ptet- <i>is10out</i> , aTc 2 ng/ml	1.52 ± 0.05
Ptet- <i>is10out</i> , aTc 6 ng/ml	2.06 ± 0.11
Ptet- <i>is10out</i> , aTc 10 ng/ml	2.85 ± 0.10
Fitted Value for a_λ	0.65 ± 0.07

4.6 Hierarchical cross-talk between sRNA targets

Most small RNAs (including RyhB) have multiple targets^{6,8,43}, whose expressions are independently regulated⁵⁰. As any one of the targets can reduce the level of sRNA, it is plausible for sRNA to mediate indirect interaction (*cross-talk*) between its different targets. It is straightforward to extend model Eq. [4.2.1] to the case of multiple targets and account for the indirect interaction between them. For example, for two target mRNAs, m_1 and m_2 , one has (assuming $p=1$ for simplicity)^{††}

$$\begin{aligned}\frac{d}{dt}s &= \alpha_s - \beta_s s - (k_1 m_1 + k_2 m_2)s \\ \frac{d}{dt}m_1 &= \alpha_1 - \beta_1 m_1 - k_1 m_1 s \\ \frac{d}{dt}m_2 &= \alpha_2 - \beta_2 m_2 - k_2 m_2 s\end{aligned}\tag{4.6.1}$$

We assume that the two genes are expressed from different promoters, and can thus be controlled by different signals. We therefore allow for α_1 and α_2 to change independently.

Consider an experiment in which the activity of the reporter gene $\alpha_1 < \alpha_s$ is held fixed, while the activity of another target, α_2 , is increased (say, by an external stimulus; see Fig. 4.6). In the strong interaction limit, $\lambda \ll 1$, we find

$$\begin{aligned}\text{For } (\alpha_1 + \alpha_2) \gg \alpha_s, \quad m_1^* &= \frac{\alpha_1}{\beta_1} \left(1 + \frac{\lambda_2}{\lambda_1} \frac{\alpha_s}{\alpha_2} \right)^{-1} \text{ and } m_2^* = (\alpha_2 - \alpha_s) / \beta_2; \\ \text{For } (\alpha_1 + \alpha_2) < \alpha_s, \quad m_1^* &= m_2^* = 0.\end{aligned}\tag{4.6.2}$$

^{††}More generally, our model is given by $\frac{d}{dt}s = \alpha_s - \beta_s s - (\sum_i k_i m_i)s$ and $\frac{d}{dt}m_i = \alpha_i - \beta_i m_i - k_i m_i s$, where i enumerates the different targets.

With α_1 fixed, one can think of the threshold for gene expression as being set by $\alpha_1 + \alpha_s$. As long as α_2 is below this threshold, both genes are silenced. Once α_2 exceeds the threshold, mRNA molecules of both species survive the sRNA repression, and both genes start being expressed. When gene 2 is highly expressed, its expression grows proportionally to α_2 , with the usual shift due to the sRNA effect. In this limit the expression of gene 2 is not affected by the presence of gene 1. At the same time, the expression of gene 2 is relieved from the sRNA effect ($m_1 \simeq \alpha_1 / \beta_1$), an effect that is amplified if $\lambda_2 \gg \lambda_1$. This can be easily explained, as high abundance of m_1 , and/or strong interaction with the sRNA, directs all sRNAs to annihilate mRNA molecules of gene 2 rather than of gene 1. Indirectly, gene 2 acts as an activator of gene 1.

In Fig. 4.7a we compare the expression of our reporter target, *crsodB-gfp*, in *sodB*⁺ and *sodB*⁻ cells. In *ryhB*⁺ cells (Fig. 4.7a), *sodB* expression affected the expression of the target reporter: at the same level of promoter activity, the reporter gene was repressed more in the absence of chromosomal *sodB* than in its presence. In comparison, no significant difference in expression was observed between *sodB*⁺ and *sodB*⁻ in *ryhB*⁻ mutants (data not shown). Fitting the data of Fig 4.7a to this model (black line) highlights the fact that the effect of *sodB* expression on the expression of the GFP reporter is peaked at intermediate levels of reporter promoter activity, close to the expression threshold of this strain (see the blue line of Fig. 4.3a) where target expression is predicted to be most sensitive to changes in sRNA levels.

Further evidence for a crosstalk between targets of a common sRNA is given in Fig. 4.7b. Here the mRNA levels of two chromosomal RyhB-targets (*sodB* and *fumA*), as

measured by RT-PCR, is shown for different expression levels of the synthetic target gene (*crsodB-gfp*) driven by the P_{Lac} promoter. At basal expression levels of the latter (no IPTG added), expression of the chromosomal targets is repressed by RyhB up to 10-fold (compare the blue and cyan symbols at “promoter activity” = 0). High expression of the plasmid target (0.25 mM IPTG) effectively rescues the chromosomal targets from repression, as seen by the collapse of all symbols at the highest level of “promoter activity”.

The crosstalk between different targets may allow for one target to relieve sRNA-mediated repression of another target. To explore this possibility, we used our model to calculate the expression level of a reporter target (R), in the presence and in the absence of another target gene (T). The ratio between the two levels of expression is displayed in Fig. 4.7c. We explore, for instance, the case where transcription of R is smaller than that of the small RNA, and therefore in the absence of T its expression level is small.

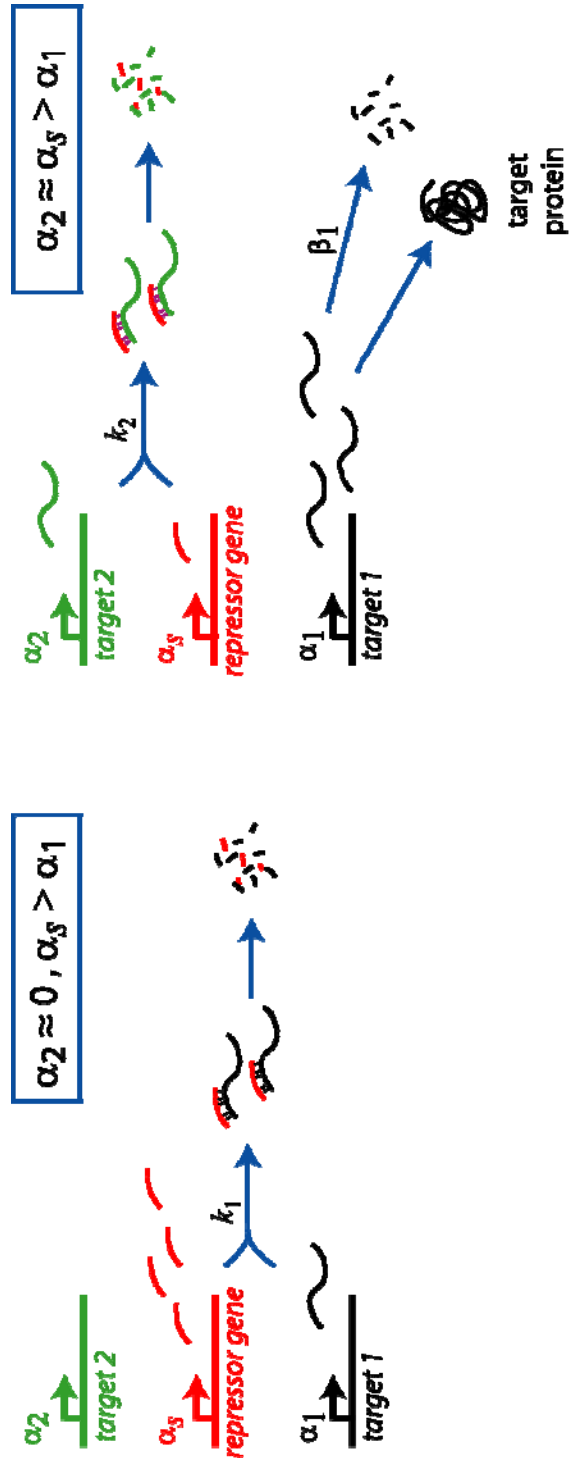


Figure 4.6. Model for indirect interaction between targets of a small RNA: in the case $k_2 \gg k_1$. When gene 2 is not expressed, the small RNA silences the expression of gene 1. When gene 2 is expressed, most sRNA molecules bind and degrade with mRNAs of gene 1, allowing mRNAs of gene 1 to be translated into proteins.

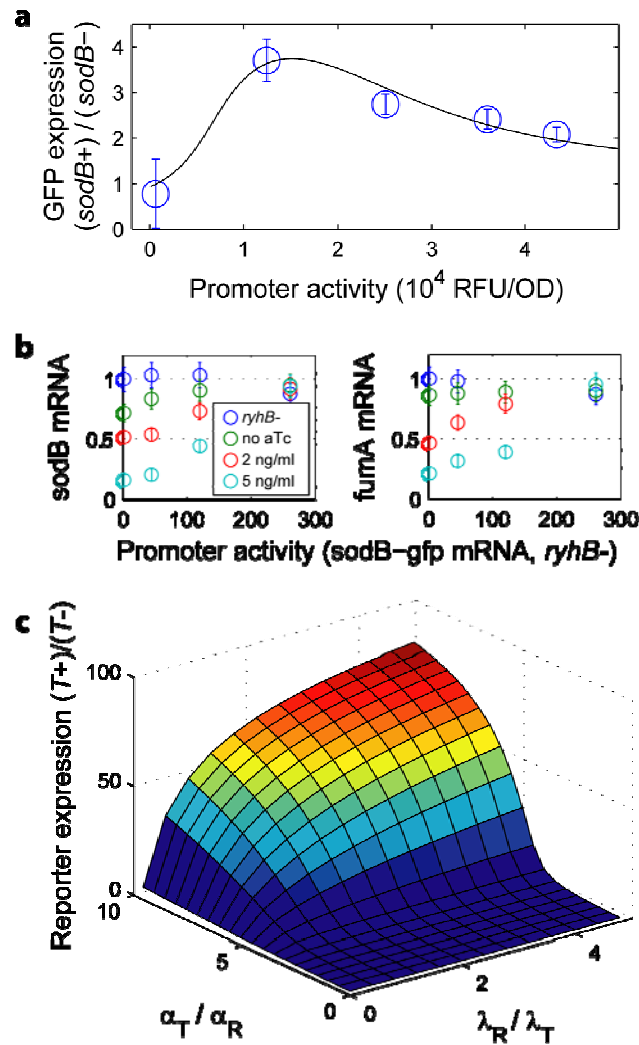


Figure 4.7. Cross talk between different targets of a common sRNA. (a). The fold-change between expression of the reporter target *crsodB-gfp* in strain ZZS22 (*sodB*⁺) and strain ZZS22s (*sodB*⁻) cells. Promoter activity of the reporter is defined as in Fig. 4.3. The black line depicts the steady-state solution of a coupled-degradation model which is a straightforward generalization of Eq. [4.2.1] to the case of two targets (see Section 4.6, *Hierarchical cross-talk between sRNA targets*). **(b).** The effect multicopy target expression on chromosomal expression of chromosomal targets. mRNA levels of two chromosomal RyhB-targets, *sodB* and *fumA*, were measured in cells with (ZZS23) and without (ZZS21) the *ryhB* plasmid. The repression effect of RyhB is measured as the ratio between mRNA levels in the two strains. As the expression level of the synthetic target, *crsodB-gfp*, is increased (by increasing IPTG concentration, 0, 0.1, 0.15, 0.25 mM), the repression effect on the chromosomal targets is reduced. **(c).** Predictions of the coupled degradation model for the expression of the reporter gene (gene R) to different transcription levels of another target (gene T). For generating the figure we arbitrarily choose $\alpha_s = 5\alpha_1$. The ratio between the leakage parameters of the two genes determined the strength of the interaction and the abruptness at the onset of gene R activity, as the expression of gene T crosses the threshold.

We mentioned in the context of a single target that the strength of the sRNA-target interaction (inversely proportional to the leakage parameter λ) is expected to influence only the smoothness of the transition but not the threshold value of the threshold-linear expression pattern. However, when multiple targets are expressed simultaneously, the different mRNA species compete for association with the same pool of sRNA, and the relative interaction strength becomes a key determinant of the complex interactions that ensue. Our model predicts a novel hierarchical cross talk effect: the expression of a weakly interacting target is highly affected by another target that is more strongly interacting (see the large λ_R / λ_T region of Fig. 4.7c), whereas the effect of a strongly interacting target is much less affected by a weakly interacting target (small λ_R / λ_T). Interestingly, for large λ_R / λ_T , the response of the reporter target to changes in the transcription rate of another target may be very sharp. For example, the data of Fig. 4.7c allows for effective Hill coefficient ~ 10 for $\lambda_R / \lambda_T \approx 5$. Thus, the sensitivity of the sRNA-mediated repression is translated into sensitivity in the indirect interaction between its targets.

Discussion

The “standard model” of gene regulation in bacteria involves primarily transcriptional initiation control by one or more regulatory proteins. Solid understanding of the key mechanistic ingredients of transcriptional regulation⁵¹, stemmed from decades of research in molecular biology, lead to a reasonable quantitative description^{52,53}. While such a framework for small RNA is still lacking, the successful description of our

experimental results by the simple kinetic model for sRNA-mediated regulation [Fig. 4.3 and Eq. [4.2.1]] prompted us to use this model to compare between the two modes of regulation.

4.7 Properties of sRNA repression vs. protein repression

The effect of a transcription factor, which represses transcription initiation by blocking binding of the RNA-polymerase has been shown^{52,53} to take the form

$$m^* = \frac{\alpha_m}{\beta_m} \frac{1}{1 + (r/K)^n} \quad [4.7.1]$$

Here r is the concentration of the protein repressor, K its binding affinity to the promoter, and n is the Hill coefficient which describes the effect of cooperativity (typically one has $n \leq 2$). The effect of the protein repressor is multiplicative. The fold of repression is determined solely by the properties of the repressor (its binding affinity to the promoter, its concentration, etc).

Proteins can also take part in post-transcriptional regulation. A protein regulator can act by destabilizing a target mRNA, either directly (*e.g.* by recruiting a ribonuclease) or indirectly (by blocking translation). For example, the protein CsrA inhibits translation of the *glgCAP* mRNA by binding to its 5'-UTR⁶⁴, thus exposing it to degradation. As a crucial difference from a small RNA, the protein is assumed to be unaffected by the degradation of the RNA molecule, and would be recycled to exert its effect on other molecules. In terms of mass-action equations, this can be modeled by

$$\begin{aligned}
\frac{d}{dt}R &= \alpha_R - \beta_R R - k_{+1}mR + k_{-1}C + \beta_C C \\
\frac{d}{dt}m &= \alpha_m - \beta_m m - k_{+1}mR + k_{-1}C \\
\frac{d}{dt}C &= k_{+1}mR - k_{-1}C - \beta_C C
\end{aligned}
\tag{4.7.2}$$

Here C is the concentration of mRNA:protein complex, k_{+1} and k_{-1} are the binding and unbinding rates of this complex, and β_c is the degradation rate of the mRNA when it is attached to the protein. α_R and β_R are, respectively, the production and decay rates of the protein.

Perhaps surprisingly, the solution of model Eq. [4.7.2] suggests that the steady state effect of this mode of repression is similar to that of a transcriptional repressor. In the case where the protein repressor inhibits translation, we have

$$\text{Protein production} \sim m^* = \frac{\alpha_m}{\beta_m} \frac{1}{1 + \frac{\beta_c}{\beta_m} \frac{r}{K}} \tag{4.7.3}$$

with $K^{-1} = k_{+1}/(k_{-1} + \beta_c)$. If the repressor does not interfere with translation, all mRNA are actively translated, and one has

$$\text{Protein production} \sim m^* + C^* = \frac{\alpha_m}{\beta_m} \frac{1 + \frac{r}{K}}{1 + \frac{\beta_c}{\beta_m} \frac{r}{K}} \tag{4.7.4}$$

More generally, it is possible that multiple proteins can bind cooperatively to the mRNA molecule to inhibit its translation or promote its degradation (as is the case of CsrA⁶⁴). It is simple to show that this would result in a Hill coefficient > 1 , as in Eq. [4.7.1].

In common to all the cases considered, the repressor reduces the abundance of mRNA by a constant multiplicative factor (“fold-change”), irrespective of the activity of the promoter (Fig. 4.8*b*). In other words, the effect of an enzymatic regulator is multiplicative, whereas the effect of sRNA is additive. However, if a certain target is always expressed less efficiently than its sRNA regulator ($\alpha_m \ll \alpha_s$), the effect of the latter becomes multiplicative, as can be seen from Eq. [4.6.2]. This is due to the fact that a weak target cannot change significantly the level of sRNA.

4.8 sRNA regulation is subject to dynamic control

Analysis of a simple model of protein-mediated gene regulation (Section 4.7, *Properties of sRNA repression vs. protein repression*) predicts that regardless of whether a protein regulator acts as a transcriptional repressor or as a catalyst of mRNA degradation, target expression always increases linearly with the promoter activity. The ratio between expression levels at different concentrations of the regulator is independent of the target activity – see Fig. 4.8*b*. Thus, one can safely talk about the “strength” of a repressor, or the “fold-change” in gene expression in the presence and absence of the repressor.

This is however not the case for the threshold-linear mode that characterizes sRNA-mediated regulation. Here the “fold-change” depends not only on the presence of the repressor, but also on the activity of the target, as illustrated by the arrows in Fig. 4.8*a*. This property may have functional consequences: Small RNA serves to tightly

shutdown a gene which is repressed by other means, while at circumstance that allow for high expression of the target, sRNA expression has virtually no effect. Moreover, in the threshold-linear mode of sRNA-mediated gene regulation, onset of repression is set by comparison of transcription rates between sRNA and its target. As a result, the threshold value is dynamically tunable through controlling the rate of sRNA transcription. In contrast, protein-operator binding affinity which controls the onset of repression in protein-mediated regulation is fixed genetically by the operator sequence. Dynamic control of the latter would require other co-factor(s) and auxiliary binding sites and become more elaborate to implement.

In some cases a small RNA serves as a node in a regulatory cascade. Expression of the small RNA is controlled by a protein regulator that senses (directly or indirectly) an environmental signal. For example, the Ferric Uptake Regulator (Fur) is activated by free Fe^{2+} molecules and negatively regulates transcription of RyhB, which in turn regulates targets required when Fe^{2+} is abundant in the medium^{34,43}. Our results suggest that small RNA is more than a simple inverter of such a protein regulator – rather, it acts as a “stress-relief” valve. For example, while Fur senses levels of Fe^{2+} continuously (through rapid equilibration between Fur and Fur-Fe^{2+}), we predict that targets of RyhB will only respond when these levels cross some threshold. This threshold is set dynamically for each target by its own synthesis rate.

Recently it has been suggested that target of microRNA regulation in eukaryotes may be classified as “switch”, “tuning” and “neutral” targets, depending on their response to microRNA level^{54,55}. In the framework presented here, these classes correspond to targets whose transcription rate is well below, near, or well above that of

the RNA regulator. We emphasize, however, that the threshold linear picture we draw is only valid for if the level of the RNA regulator is affected by its interaction with its targets. This is yet to be established for microRNA in eukaryotes.

4.9 sRNA may exhibit tight repression of fluctuations

Deep in the repressed state, sRNAs are predicted to strongly repress *variations* in protein expression. The effect of noise on gene expression is a subject of extensive current research⁵⁶⁻⁵⁹. We studied this effect theoretically by generalizing the model Eq. [4.2.1] to incorporate stochastic fluctuations. In Fig. 4.8c we compare results of stochastic simulations for two genes with the same low mean protein expression: Gene S is silenced by a small RNA, and gene R is repressed transcriptionally by a protein regulator. Transcriptional repression is mimicked by choosing the transcription rate of gene R to be such that the average copy number of mRNA molecules (and thus, of proteins) would be the same for the two genes. For the parameters used in our simulations with $\alpha_s = 1$ nM/ min, $k = 0.02$ (nM·min)⁻¹, and $\alpha_m = 0.1$ nM/ min for gene S, this corresponds to $\alpha_m = 0.0043$ nM/ min for gene R. Other parameters are as given in Table 4.1. Data were collected at $t \approx 1000$ min from 10^4 independent simulations.

Two major features arise from Fig. 4.8c. First, we find that the number of proteins/cell is much more peaked around the mean value (~ 2.15), i.e. the variance is much lower, for the sRNA regulated gene (S). Second, we find that the probability of having no proteins at all is much higher in the case of transcription repression.

To get a better understanding of these features, we assume a high level of sRNA (as expected in the repressed state), and thus neglect fluctuations in its copy number. We therefore take $s = (\alpha_s - \alpha_m) / \beta_s$. With this assumption, one can think of the target mRNA as having an effective degradation rate $\tilde{\beta}_m = \beta_m + k(\alpha_s - \alpha_m) / \beta_s$. Following Ref. 60, we expect

$$\langle p \rangle = \frac{\alpha_m}{\tilde{\beta}_m} \frac{\alpha_p}{\beta_p} \quad [4.9.1]$$

$$\frac{\langle \Delta p^2 \rangle}{\langle p \rangle} = 1 + \frac{\alpha_p / \tilde{\beta}_m}{1 + \beta_p / \tilde{\beta}_m} \simeq 1 + \alpha_p / \tilde{\beta}_m, \quad [4.9.2]$$

where $\langle p \rangle$ is the mean protein number, and $\langle \Delta p^2 \rangle$ is its variance. The expression in [4.9.2] is the Fano factor, a common measure for the magnitude of the noise. Importantly, we see that it is linear in the mRNA life-time, which is substantially reduced by the small RNA. Moreover, we note that the Fano factor is independent of the transcription rate α_m . The results [4.9.1]-[4.9.2], with $\tilde{\beta}_m$ as defined above, are compared with our Monte-Carlo simulations in Fig. 4.9. As expected, our approximation, which neglects fluctuations in the sRNA concentration, works quite well as long as the transcription rate α_s is not too small. We discuss the noise properties of sRNA regulation more extensively in Chapter 5, *Noise and Silence in sRNA Regulation*.

In general, we predict a much-reduced variance in protein level for sRNA-mediated regulation. This can be understood by inspecting the time courses of protein expression (Fig. 4.8c). With the protein regulator (red curve), any leakage in transcription

is amplified through translation, resulting in large bursts of protein expression, as was recently observed experimentally^{61,62}. With the sRNA (blue curve), gene expression is much smoother as mRNA molecules are rarely translated. This difference in the noise properties may be very important in situations where a large burst of proteins will switch a cell from one stable state to another. In cases such as stress responses where unintentional entry into the alternative state may be harmful and spontaneous switching is to be avoided, sRNA-mediated regulation might possess a distinct advantage. Attenuation of noise by decreased burst size may also be accomplished by eukaryotic microRNAs⁵⁵, through a decrease in mRNA stability or inhibition of translation.

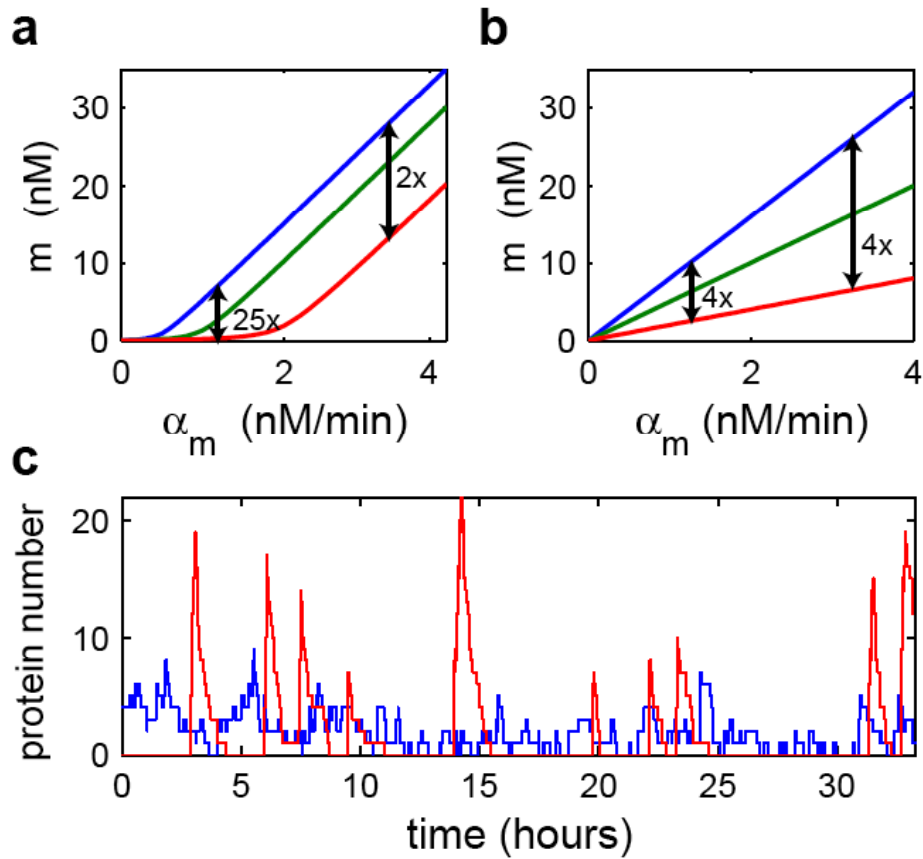


Figure 4.8. Comparison between sRNA- and protein-mediated repression. (a). Steady-state solution of model Eq. [4.2.1], with the estimated parameters of Table 4.1. The strength of sRNA repression decreases as the target transcription increases. (b). Steady-state solution of a model for protein regulators (Section 4.7, *Properties of sRNA repression vs. protein repression*), where the strength of repression is independent of target transcription rate. (c). Temporal behavior in a single stochastic simulation of the expression of two model genes S and R, regulated by sRNA and protein regulators respectively. The bursty nature of the noise for gene R is compared with the smooth fluctuations exhibited by gene S.

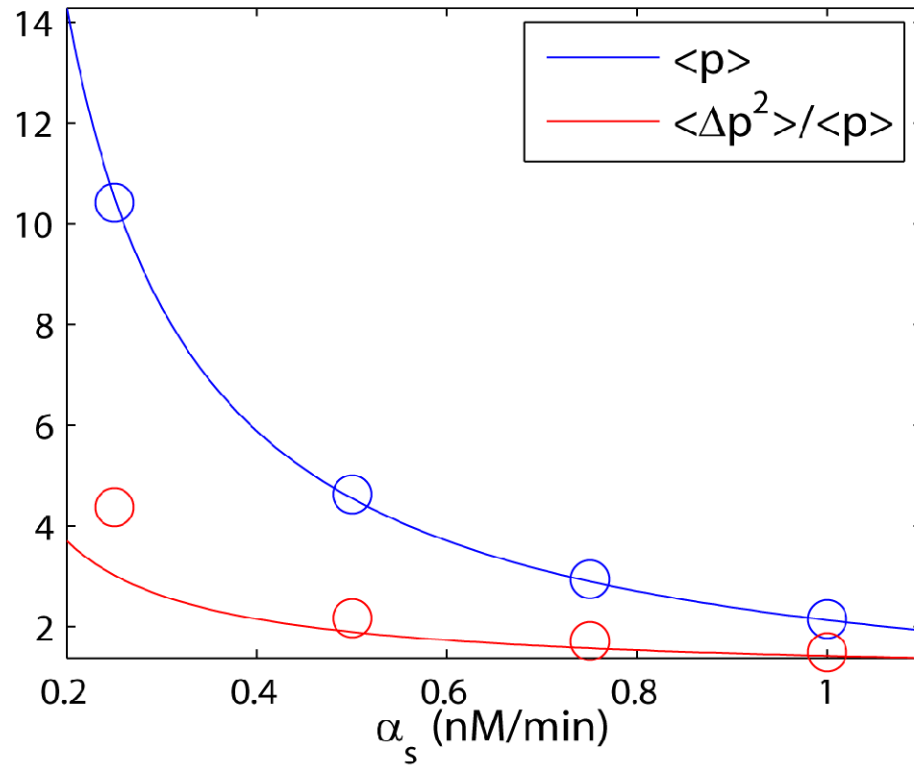


Figure 4.9. Mean and Fano factor of the protein level in the silenced state. Lines are given by [4.9.1]-[4.9.2], and symbols correspond to Monte-Carlo simulations. For gene S we take $\alpha_m = 0.1$ nM/sec and $\alpha_s = 1.0$ nM/sec. For gene R we take $\alpha_m = 0.0043$ nM/sec. Other parameters are as given in Table 4.1.

4.10 Small RNA regulation may be highly sensitive

Small RNA-mediated regulation was predicted to be ultrasensitive to small changes in sRNA expression near the threshold⁷. A common measure for the abruptness of a transition, or the *sensitivity*, is the maximal slope of the response curve, $m(\alpha_s)$, in a double-log scale. According to model [4.2.1], it is given by $\frac{1}{2}\sqrt{\alpha_m/\lambda}$, which quantifies our statement that lower leakage makes a sharper transition, and also predicts a sharper transition for highly expressed targets. With the parameters of Table 4.1, we find the sensitivity to be given approximately by 2.5 for $\alpha_m = 1$ nM/sec, and 4.3 for $\alpha_m = 3$ nM/sec. In comparison, for a protein repressor the sensitivity is bounded by the Hill coefficient, and is typically ≤ 2 . On the other hand we note that much higher sensitivity can be achieved, e.g., by zeroth-order kinetics⁶³.

Through quantitative characterization of gene regulation for two distinct classes of sRNA regulators, we have shown that sRNA-mediated regulation possesses many functional properties that are fundamentally different from the classical, protein-mediated mode of gene regulations. Analysis of our model suggests that sRNAs may offer tight regulation below the threshold (repressing the average expression and reducing fluctuations) accompanied by derepression away from the threshold. Taken together, this suggests that sRNAs working in the threshold-linear mode may be particularly suitable for a “stress-relief” mechanism, where no action is elicited until a “tolerance threshold” is exceeded. Knowledge of these properties is essential to an integrated understanding of

gene regulatory systems, and may inspire the design and synthesis of novel genetic circuits⁶⁵ with properties difficult to obtain by using regulatory proteins alone.

4.11 Materials and Methods: Strains & plasmids

All experiments were performed with BW-RI cells derived from *E. coli* K-12 BW25113⁶⁶, with the transfer of the *spr-lacI-tetR* cassette from DH5 α -ZI cells⁴⁵ by phage P1 transduction. This cassette provides the constitutive expression of the *lacI^q* and the *tetR* genes⁴⁵. For some experiments, *ryhB* and/or *sodB* were deleted from BW-RI⁶⁶. These strains were then transformed by the following “target” and “source” plasmids. All strains and plasmids used are summarized in Table 4.4.

pZE12-luc⁴⁵ was used to make the target plasmid pZE12S. Using site-directed mutagenesis, an *EcoRI* site was created by adding GAAT immediately downstream of +1 of the P_{Lac-O1} promoter. The region between the newly created *EcoRI* site and the *EcoRI* site 6 bp upstream of RBS was then deleted by *EcoRI* digestion and subsequent religation, yielding pZE12-lucM. The *KpnI-XbaI* flanking *luc* gene in pZE12-lucM was replaced by the *gfpmut3b* structure gene⁶⁷. This yields pZE12G, which harbors the P_{Lac-O1}:*gfpmut3b* construct with a 5'-UTR defined by an *EcoRI* site immediately downstream of +1 and *KpnI* site immediately upstream of the translation start of *gfpmut3b*. The 15-base sequence sandwiched by the *EcoRI* and *KpnI* sites, ATTAAGAGGAGAAA, contains an RBS indicated by the underlined bases. The 5'-untranslated region from the control region of *sodB* (*crsodB*, from -1 to +88 relative to the transcriptional start site of *sodB* and including the first 11 codons) was cloned into the *EcoRI* and *KpnI* sites of

pZE12G, yielding pZE12S. pZE12S therefore contains the ColE1 *ori*, the P_{Llac-O1} promoter⁴⁵, and *crsodB* fused to the coding sequence of the *gfpmut3b* gene. Similarly, the control region of IS10-Right (from +1 to +36) was substituted for *crsodB* in pZE12S, yielding pZE12IN. To improve the expression level, the RBS in *is10in* control region was modified by changing TC (+16 to +17) to GG.

Three sRNA-source plasmids (pZA30R, pZA31R and pZA31O), were derived from the pZA31-luc plasmid⁴⁵. Each contains the p15A replication *ori* and is marked by chloramphenicol resistance. First, a *NdeI* site was added immediately downstream the +1 of the *luc* gene by inserting ATG between +2 and +3 and a *BamHI* site was added immediately downstream of *luc* by inserting ATC between the 1772th and 1773th nucleotides, yielding pZA31-lucNB. For pZA31R, the *ryhB* gene (from +1 to +96 cloned directly from *E. coli* K-12) was ligated into the *NdeI/BamHI* sites of pZA31-lucNB, replacing the *luc* gene. For pZA30R, the P_{Ltet-O1} promoter and the *luc* gene of pZA31-lucNB were replaced by P_{ryhB} *ryhB* (from -62 to +96 cloned directly from *E. coli* K-12 MG1655). Finally, for pZA31O, the *is10out* gene (from +1 to +103) was substituted for the *luc* gene in pZA31-lucNB.

In addition, we transferred the target *crsodB:gfp* to the *attP* site of strain BW25113 *ryhB* chromosome using the method of Diederich *et al*⁶⁸. Briefly a *SalI/BamHI*-flanked P_{Llac-O1}:*gfpmut3b* containing the downstream terminator was cloned into the same sites of pLDR10 containing the attachment site *attP* and encoding the Cm resistance. The recombinant plasmid was digested with *NotI* and the larger portions of the plasmids containing the fragment of interest but not the *ori* were religated. The circular DNA molecules were transformed into BW25113 *ryhB* cells expressing the *int* gene

contained in pLDR8, a helper plasmid bearing a temperature-sensitive *ori* and encoding the km resistance. The transformations were applied on LB +Ap plates that were incubated at 42 C. The transformants were tested for sensitivity to Cm and Km. The Ap-resistant but Cm- and km-sensitive transformants were identified as the clones that carry the DNA fragment of interest at the *attP* site of *E. coli* chromosome.

Table 4.4: Bacterial strains and plasmids used in this study. All coordinates are relative to +1 position of the given gene

Strain	Genotype	Derived From	Comments
BW-RI	<i>spr-lacI-tetR</i>	BW25113	<i>spr-lacI-tetR</i> cassette derived from DH5 α -ZI ⁴⁵
BW-NULL	pNULL	BW-RI	Negative control
ZZS00	Δ <i>ryhB</i>	BW-RI	<i>ryhB</i> deletion from -54 to +94
ZZS02s	Δ <i>sodB</i>	BW-RI	<i>sodB</i> deletion from +1 to +91
ZZS02q	Δ <i>hfq</i>	BW-RI	<i>hfq</i> deletion from +8 to +279
ZZS22	pZE12S	BW-RI	
ZZS20	Δ <i>ryhB</i> pZE12S	ZZS00	
ZZS23	Δ <i>ryhB</i> pZE12S pZA31R	ZZS00	
ZZS13	Δ <i>ryhB</i> pZE12G pZA31R	ZZS00	
ZZS21	Δ <i>ryhB</i> pZE12S pZA31- <i>luc</i>	ZZS00	
ZZS11	Δ <i>ryhB</i> pZE12G pZA31- <i>luc</i>	ZZS00	
ZZS24	Δ <i>ryhB</i> pZE12S pZA30R	ZZS00	
ZZS22s	Δ <i>sodB</i> pZE12S	ZZS02s	
ZZS23q	Δ <i>hfq</i> Δ <i>ryhB</i> pZE12S pZA31R	ZZS00, ZZS02q	
ZZS21q	Δ <i>hfq</i> Δ <i>ryhB</i> pZE12S pZA31- <i>luc</i>	ZZS00, ZZS02q	
ZZS35	pZE12IS pZA31O	BW-RI	
ZZS31	pZE12IS pZA31- <i>luc</i>	BW-RI	
ZZS01	Δ <i>ryhB</i> Δ <i>lacI</i> Δ <i>lacZYB</i> pZA31- <i>luc</i> pZS-In-RI	MG 1655	
ZZS41	Δ <i>ryhB</i> Δ <i>lacI</i> Δ <i>lacZYB</i> Φ (<i>crsodB</i> - <i>gfp</i>) pZA31- <i>luc</i> pZS-In-RI	MG 1655	
ZZS43	Δ <i>ryhB</i> Δ <i>lacI</i> Δ <i>lacZYB</i> Φ (<i>crsodB</i> - <i>gfp</i>) pZA31- <i>ryhB</i> pZS-In-RI	MG 1655	
Plasmid			
pZE12			ColE1 <i>ori</i> , Amp marker ⁴⁵
pZA31- <i>luc</i>	<i>P_{LTet-OI}:luc</i>		p15A <i>ori</i> , Cm marker ⁴⁵
pNULL	<i>PNULL:gfpmut3b</i>	pZE12	Promoterless <i>gfpmut3b</i> - negative control;
pZE12G	<i>P_{Llac-OI}:gfpmut3b</i>	pZE12	
pZE12S	<i>P_{Llac-OI}:crsodB-gfpmut3b</i>	pZE12	<i>crsodB</i> amplified from +1 to +88
pZE12IS	<i>P_{Llac-OI}:cris10-gfpmut3b</i>	pZE12	<i>cris10</i> amplified from +1 to +36; TC at +17 changed to GG.
pZA31R	<i>P_{LTet-OI}:ryhB</i>	pZA31- <i>luc</i>	<i>ryhB</i> amplified from +1 to +96
pZA31O	<i>P_{LTet-OI}:is10out</i>	pZA31- <i>luc</i>	<i>is10out</i> amplified from +1 to +103
pZA30R	<i>P_{ryhB}:ryhB</i>	pZA31- <i>luc</i>	<i>PryhB:ryhB</i> amplified from -62 to +96
pZSInRI			

4.12 Medium, growth, measurements

BW-RI strains each containing the target and/or source plasmids were grown in M63 minimal media with standard concentrations of the appropriate antibiotics. The overnight cultures were diluted into fresh M63 media ($OD_{600} \approx 0.002$) containing the appropriate antibiotics as well as varying amounts of the inducers (aTc, IPTG, FeSO₄) in the wells of 48-well plates. The plates were incubated with shaking at 37°C and taken for OD_{600} and fluorescence measurements every hour for up to 15 hours (until a final OD_{600} of 0.2 - 0.3) using a TECAN Genios-Pro plate reader. Each measurement was repeated 3 times and the data was analyzed as discussed below.

For RT-PCR measurements, overnight cultures were used to inoculate M63 medium with 0.5% glucose, standard concentrations of the appropriate antibiotics, and various concentrations of inducers to an initial OD_{600} of 0.001 and grown in 48 well plates (Costar) in a 37°C incubator while monitoring OD_{600} and GFP fluorescence (if applicable). When OD_{600} of these cultures reached 0.3-0.5, ~10⁹ cells of each culture were harvested in a microcentrifuge at 4°C, treated with 10 mg/ml lysozyme in TE buffer (pH = 8.0) and RNA was collected using an Absolutely RNA miniprep kit (Stratagene). The prepared samples were then treated with Turbo DNA-free DNase (Ambion), and PCR controls were performed on each sample to verify the absence of contaminating DNA. cDNA was prepared with 1 µg of RNA from each sample using Superscript III First Strand Synthesis system (Invitrogen). Dilutions of the resulting samples were then used as the template in PCR reactions using iQ SYBR Green Supermix (Bio-Rad) in a Smart Cycler thermal cycler (Cepheid).

To measure expression from a chromosomal target, cells were grown overnight in

minimal media with antibiotics. Cultures were then diluted to $OD_{600}=0.01$, and grown in a 12-well plate, 3 ml of culture in each well, with antibiotics and inducers. To establish growth rate, OD_{600} was measured every 60 min. Cultures were grown at 37C with constant shaking until they reach $OD_{600}=0.3$, at which time 1.7 ml of each culture was spun down and resuspended in 1 ml PBS. Expression data were collected using a Becton-Dickinson FACSCalibur flow cytometer with a 488-nm argon excitation laser and a 515- to 545-nm emission filter (FL1) at a low flow rate. Photomultiplier tube (PMT) voltage was set to 950V, and a linear amplifier was set at 9.5X. Forward scatter and fluorescence values were collected for 50,000 cells.

4.13 Data analysis

In order to obtain expression patterns for the different strains, we averaged at each time point the data obtained from the different repeats for each combination of strain and inducers. Fitting $\log_2(OD_{600})$ vs. time to a linear function yielded straight lines (data not shown) whose slope gave the doubling rate (μ) for each strain and condition (yielding doubling time ~ 2 hours for most strains.) Next, for all of the time points concerning each strain and condition, we plotted the average fluorescence vs. average OD_{600} on linear-linear plot and extracted the slope (f). In Fig. 4.2a we show, for example, GFP fluorescence against OD_{600} for *ryhB*⁻ strain (ZZ21), together with the fitted slopes. Each slope gave the average fluorescence per growing cell [in unit of RFU/OD] for that strain and the corresponding growth condition. The raw fluorescence production rate per cell

was computed as $f\mu(1 + \mu\tau)$ ⁶⁹, upon taking into account of the maturation kinetics of GFPmut3 (maturation half-life of ~30 min) ⁶⁷. We then subtracted away from this raw rate the background fluorescence production rate, obtained in the same way from data collected from our negative control (BW-NULL). This yielded the rate of GFP production synthesis from P_{Lac-O1} , and is referred to as the GFP expression. The results are plotted Fig. 4.2b at each IPTG level for each RyhB source.

To fit the experimental data with the steady-state solution Eq. [4.2.2], we define

$$\text{GFP expression} = f(a, a_s, a_\lambda) = \frac{1}{2} \left[(a - a_s - a_\lambda) + \sqrt{(a - a_s - a_\lambda)^2 + 4aa_\lambda} \right].$$

Here $a = b\alpha_m / \beta_m$ is the promoter activity (set by the IPTG concentration); $a_s = b\alpha_s / \beta_m$ is related with the activity of the small RNA (and therefore takes different values for different experiments); and $a_\lambda = b\beta_s / k$ is related to the ‘leakage parameter’ (defined in the text). The latter is independent of the sRNA activity, and should therefore take a single value for all experiments. The coefficient b corresponds to the rate of GFP synthesis and its fluorescent activity. We fit the data to $f(a, a_s, a_\lambda)$ using a standard Levenberg-Marquardt algorithm implemented in MATLAB (MathWorks), with the least-square error defined as

$$\sum_{\text{experiment}(e)} \sum_{[IPTG]} [\text{GFP activity} - f(a([IPTG]), a_s(e), a_\lambda)]^2.$$

The estimated parameters are given in Table 4.2 in terms of 0.5-confidence intervals.

The values of the model parameters can be estimated from various experiments. Consider first RyhB and its targets^{34,35}. The Hfq-bound small RNA RyhB is very stable, with half-life of ~ 30 min^{35,40}, yielding $\beta_s \sim 1/50 \text{ min}^{-1}$. Similarly, from the half-life of ~ 6 min for *sodB* mRNA³⁵, we have $\beta_m \sim 1/10 \text{ min}^{-1}$. Moreover, DNA microarray experiments^{50,70} indicated approximately 10-20 copies/cell for the *sdhCDAB* and *sodB* mRNA in rich medium (where Fe is abundant and RyhB is expected to be repressed). This suggests a target transcription rate (α_m) of ~ 1 nM/min in the state where mRNA is expressed. In general, α_m is controlled by various cellular signals (*e.g.* *sdhCDAB* by Crp-cAMP) and can typically vary ~ 10 fold. (The DNA microarray study of Zhang *et al*⁵⁰ showed approximately 5-fold change in *sdhCDAB* and *sodB* mRNA levels under various physiological conditions.) On the other hand, the activity of the RyhB promoter has a broad range since it is strongly regulated by Fur-Fe²⁺ whose concentration can vary over 1000-fold⁴⁴. We model the latter by allowing α_s to take on the range from 0.1/min to 10/min. Finally the coupled degradation rate k can also be deduced from the results of Masse *et al*³⁵: Since RyhB is shown to disappear in the presence of its targets within 3 min, then using an estimated target mRNA concentration of 20 nM, we find $1/50 \text{ (nM min)}^{-1}$, which is close to the diffusion-limited association rate for typical small proteins^{71,72}.

Next, consider RNA-OUT and its target *is10in*. RNA-OUT itself is extremely stable, with half-life dictated by dilution due to growth $\beta_s \sim 0.02 \text{ min}^{-1}$ ⁷³, while half-life of *is10in* mRNA is typical to bacterial mRNA (2-3 min, $\beta_m \sim 0.3$ ⁴⁹). Binding of RNA-OUT to its target mRNA is characterized by a second-order binding constant in the range of $k \sim 1/50 - 1/20 \text{ (nM min)}^{-1}$. The *pOUT* promoter is a typical promoter, and we assume

that α_s is not very different from that of RyhB⁷⁴. The *pIN* promoter, on the other hand, is atypically weak, and is only enhanced 10-fold upon methylation⁷⁴⁻⁷⁶.

Values of the parameters are summarized in Table 4.3.

4.14 Acknowledgment

We are grateful for helpful discussions with S. Gottesman, W.L. Loomis, R.W. Simons and N. Kleckner during the course of this work. This work was supported by the NSF PFC-sponsored Center for Theoretical Biological Physics (Grants No. PHY-0216576 and PHY-0225630), by the DOE through grant DE-FG02-03-ER63691, and by the NIH through grant GM070915.

Chapter 4 contains material in Levine E, Kuhlman T, Zhang Z, Hwa T. Quantitative features of gene regulation by small RNA. *PLoS Biology*, submitted, used with permission. The dissertation author was the primary experimental investigator of this work.

4.15 References

1. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*. 2004;116:281-97.
2. Gottesman S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet*. 2005;21:399-404.
3. Storz G, Altuvia S, Wassarman KM. An abundance of RNA regulators. *Annu Rev Biochem*. 2005;74:199-217.
4. Zamore PD, Haley B. Ribo-gnome: The Big World of Small RNAs. *Science*. 2005;309:1519-24.
5. Repoila F, Majdalani N, Gottesman S. Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. *Mol Microbiol*. 2003;48:855-61.
6. Gottesman S. The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu Rev Microbiol*. 2004;58:303-28.
7. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell*. 2004;118:69-82.
8. Majdalani N, Vanderpool C, Gottesman S. Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol*. 2005;40:93-113.
9. Ambros V. The functions of animal microRNAs. *Nature*. 2004;431:350-5.
10. Zhang B, Pan X, Cobb GP, Anderson TA. Plant microRNA: a small regulatory molecule with big impact. *Dev Biol*. 2006 Jan 1;289(1):3-16.
11. Kleckner N. Transposon Tn10. In: Berg DE, How NM (eds). *Mobile DNA*. Washington, D.C: American Society of Microbiology, 1989:227-68.
12. Kleckner N. Regulation of transposition in bacteria. *Annu Rev Cell Biol*. 1990;6:297-327.
13. Itoh T, Tomizawa J. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proceedings of the National Academy of Sciences of the United States of America*. 1980 May;77(5):2450-4.
14. Lacatena RM, Cesareni G. Base pairing of RNA I with its complementary sequence in the primer precursor inhibits ColE1 replication. *Nature*. 1981 Dec 17;294(5842):623-6.

15. Light J, Molin S. Post-transcriptional control of expression of the repA gene of plasmid R1 mediated by a small RNA molecule. *Embo J.* 1983;2(1):93-8.
16. Wagner EG, Simons RW. Antisense RNA control in bacteria, phages, and plasmids. *Annu Rev Microbiol.* 1994;48:713-42.
17. Gerdes K, Gulyaev AP, Franch T, Pedersen K, Mikkelsen ND. Antisense RNA-regulated programmed cell death. *Annu Rev Genet.* 1997;31:1-31.
18. Wagner EG, Altuvia S, Romby P. Antisense RNAs in bacteria and their genetic elements. *Adv Genet.* 2002;46:361-98.
19. Hjalte T, Wagner EG. The effect of loop size in antisense and target RNAs on the efficiency of antisense RNA control. *Nucleic Acids Res.* 1992 Dec 25;20(24):6723-32.
20. Nordgren S, Slagter-Jager JG, Wagner GH. Real time kinetic studies of the interaction between folded antisense and target RNAs using surface plasmon resonance. *J Mol Biol.* 2001 Jul 27;310(5):1125-34.
21. Sahagan BG, Dahlberg JE. A small, unstable RNA molecule of *Escherichia coli*: spot 42 RNA. I. Nucleotide sequence analysis. *J Mol Biol.* 1979 Jul 5;131(3):573-92.
22. Aiba H, Matsuyama S, Mizuno T, Mizushima S. Function of micF as an antisense RNA in osmoregulatory expression of the ompF gene in *Escherichia coli*. *J Bacteriol.* 1987 Jul;169(7):3007-12.
23. Kawano M, Reynolds AA, Miranda-Rios J, Storz G. Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.* 2005;33(3):1040-50.
24. Vogel J, Bartels V, Tang TH, et al. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* 2003 Nov 15;31(22):6435-43.
25. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* 2001 Jul 1;15(13):1637-51.
26. Rivas E, Klein RJ, Jones TA, Eddy SR. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol.* 2001 Sep 4;11(17):1369-73.

27. Argaman L, Hershberg R, Vogel J, et al. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol*. 2001 Jun 26;11(12):941-50.
28. Tjaden B, Saxena RM, Stolyar S, Haynor DR, Kolker E, Rosenow C. Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res*. 2002 Sep 1;30(17):3732-8.
29. Storz G, Gottesman S. Versatile roles of small RNA regulators in bacteria. In: Gesteland RF, Cech TR, Atkins JF (eds). *The RNA World*. 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2006:567-94.
30. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell*. 2002 Jan;9(1):11-22.
31. Chen S, Zhang A, Blyn LB, Storz G. MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J Bacteriol*. 2004 Oct;186(20):6689-97.
32. Lease RA, Smith D, McDonough K, Belfort M. The small noncoding DsrA RNA is an acid resistance regulator in *Escherichia coli*. *J Bacteriol*. 2004 Sep;186(18):6179-85.
33. Opdyke JA, Kang JG, Storz G. GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol*. 2004 Oct;186(20):6698-705.
34. Masse E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA*. 2002;99:4620-5.
35. Masse E, Escorcia FE, Gottesman S. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev*. 2003;17:2374-83.
36. Geissmann TA, Touati D. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J*. 2004;23:396-405.
37. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol*. 2003;50:1111-24.
38. Storz G, Opdyke JA, Zhang A. Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol*. 2004;7:140-4.
39. Moller T, Franch T, Hojrup P, et al. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell*. 2002 Jan;9(1):23-30.

40. Moll I, Afonyushkin T, Vytvytska O, Kaberdin VR, Blasi U. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA*. 2003;9:1308-14.
41. Sonnleitner E, Napetschnig J, Afonyushkin T, et al. Functional effects of variants of the RNA chaperone Hfq. *Biochem Biophys Res Commun*. 2004;323:1017-23.
42. Elf J, Paulsson J, Berg OG, Ehrenberg M. Near-critical phenomena in intracellular metabolite pools. *Biophys J*. 2003 Jan;84(1):154-70.
43. Masse E, Vanderpool CK, Gottesman S. Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol*. 2005;187:6962-71.
44. Andrews SC, Robinson AK, Rodriguez-Quinones F. Bacterial iron homeostasis. *FEMS Microbiol Rev*. 2003;27:215-37.
45. Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*. 1997 Mar 15;25(6):1203-10.
46. Ma C, Simons RW. The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site. *Embo J*. 1990 Apr;9(4):1267-74.
47. Simons RW, Kleckner N. Translational control of IS10 transposition. *Cell*. 1983 Sep;34(2):683-91.
48. Case CC, Simons EL, Simons RW. The IS10 transposase mRNA is destabilized during antisense RNA control. *Embo J*. 1990 Apr;9(4):1259-66.
49. Jain C, Kleckner N. IS10 mRNA stability and steady state levels in *Escherichia coli*: indirect effects of translation and role of rne function. *Mol Microbiol*. 1993 Jul;9(2):233-47.
50. Zhang Z, Gosset G, Barabote R, Gonzalez CS, Cuevas WA, Saier MH. Functional interactions between the carbon and iron utilization regulators, Crp and Fur, in *Escherichia coli*. *J Bacteriol*. 2005;187:980-90.
51. Ptashne M, Gann A. *Genes and Signals*. New York, NY: Cold Spring Harbor Press; 2002.
52. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev*. 2005 Apr;15(2):125-35.
53. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*. 2005 Apr;15(2):116-24.

54. Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet.* 2004 May;5(5):396-400.
55. Hornstein E, Shomron N. Canalization of development by microRNAs. *Nat Genet.* 2006 Jun;38 Suppl:S20-4.
56. Swain PS, Elowitz MB, & Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA.* 2002;99:12795-800.
57. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet.* 2005;6:451-64.
58. Swain PS. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J Mol Biol.* 2004 Dec 3;344(4):965-76.
59. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nat Genet.* 2002 May;31(1):69-73.
60. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America.* 2001 Jul 17;98(15):8614-9.
61. Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. *Science.* 2006 Mar 17;311(5767):1600-3.
62. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature.* 2006 Mar 16;440(7082):358-62.
63. Goldbeter A, and Koshland DE, J. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci USA.* 1981;78:6840-4.
64. Baker CS, Morozov I, Suzuki K, Romeo T, Babitzke P. CsrA regulates glycogen biosynthesis by preventing translation of glgC in *Escherichia coli*. *Molecular microbiology.* 2002 Jun;44(6):1599-610.
65. Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ. Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol.* 2004;22:841-7.
66. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America.* 2000 Jun 6;97(12):6640-5.

67. Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*. 1996;173(1 Spec No):33-8.
68. Diederich L, Rasmussen LJ, Messer W. New cloning vectors for integration in the lambda attachment site attB of the *Escherichia coli* chromosome. *Plasmid*. 1992 Jul;28(1):14-24.
69. Leveau JH, Lindow SE. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J Bacteriol*. 2001;183:6752-62.
70. Selinger DW, Cheung KJ, Mei R, et al. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat Biotechnol*. 2000 Dec;18(12):1262-8.
71. Koren R, Hammes GG. A kinetic study of protein-protein interactions. *Biochemistry*. 1976 Mar 9;15(5):1165-71.
72. Northrup SH, Erickson HP. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proceedings of the National Academy of Sciences of the United States of America*. 1992 Apr 15;89(8):3338-42.
73. Case CC, Roels SM, Jensen PD, Lee J, Kleckner N, Simons RW. The unusual stability of the IS10 anti-sense RNA is critical for its function and is determined by the structure of its stem-domain. *Embo J*. 1989 Dec 20;8(13):4297-305.
74. Case CC, Roels SM, Gonzalez JE, Simons EL, Simons RW. Analysis of the promoters and transcripts involved in IS10 anti-sense RNA control. *Gene*. 1988 Dec 10;72(1-2):219-36.
75. Raleigh EA, Kleckner N. Quantitation of insertion sequence IS10 transposase gene expression by a method generally applicable to any rarely expressed gene. *Proc Natl Acad Sci U S A*. 1986 Mar;83(6):1787-91.
76. Roberts D, Hoopes BC, McClure WR, Kleckner N. IS10 transposition is regulated by DNA adenine methylation. *Cell*. 1985 Nov;43(1):117-30.

CHAPTER 5

Noise and Silence in Small RNA Regulation

5.1 Introduction

Living cells often retain key molecules, such as proteins and mRNA, at very low copy number ¹. Fluctuations can therefore be significant, and may have implications on cellular regulation ^{2,3} and lead to phenotypic diversity ⁴⁻⁶. Sources of genetic noise are typically divided into two components, intrinsic and extrinsic ^{2,7,8}. The first is attributed to the stochastic nature of mRNA and protein synthesis and degradation ⁹⁻¹¹, while the latter stems from global factors within the cell, such as ribosome concentration or degree of DNA supercoiling ^{7,8}. Advances in single-cell technologies of recent years have allowed the degree of intrinsic vs. extrinsic noise in single-cell organisms to be probed ^{7,12}, as well as to characterize sources of intrinsic noise ¹³⁻¹⁸. These studies were accompanied by theoretical modeling of genetic noise at the level of a single gene ^{9,10} or small modules of gene networks ^{8,11,19-21}.

Most of the studies mentioned above were done in the context of transcription regulation (except for Ref. 19, which studied theoretically protein mediated post-transcription regulation). In recent years the central role of small regulatory RNA (sRNA) molecules in gene regulation has been unveiled. Organisms that employ small RNAs for gene silencing range from bacteria to vertebrates ²²⁻²⁵. While the pathways and details are different, in all cases small RNAs target specifically messenger RNAs of a target gene, based on some base-pair complementarity and secondary structure. Upon binding – and with the aid of auxiliary proteins – target mRNA is either directed for cleavage or its translation is inhibited. Interestingly, some cases are known where

bacteria use small RNAs as activators of gene expression^{26,27}. However, a eukaryotic counterpart is yet to be discovered.

Are the noise characteristics of genetic regulation by small RNAs significantly different than regulation by protein transcription factors? Here we address this question in the context of RyhB, one of the best-characterized small RNAs in *E. coli*^{28,29}. This small RNA, involved in the control of iron homeostasis and oxidative stress^{28,30}, represses the expression of several target genes by targeting the 5'-UTR of their mRNAs. Binding of the sRNA to its target promotes degradation of both molecules^{29,31}. Previously, we described the threshold-linear response that characterizes target genes under this mode of regulation (see Chapter 4, *Quantitative Characteristics of Gene Regulation by small RNA*). Here we extend our model to incorporate stochastic fluctuations.

When the synthesis rate of the small RNA significantly exceeds that of its target mRNA, the average copy number of the latter is kept extremely low. Under these circumstances, one might expect that mRNA level should fluctuate considerably. However, our model predicts that sRNA regulation keeps the noise level extremely low by reducing the burstiness of translation, consistent with the notion that bursty translation is a major source of intrinsic noise^{10,13,16,17}.

Average mRNA level is expected to remain low even when synthesis of sRNA and mRNA occur at comparable rates. However, in this case our model predicts overwhelming fluctuations, which may increase mRNA level, even to its unrepressed level. Interestingly, the nature of these fluctuations allows for a population diversity, in which some cells may possess high mRNA copy number, while others possess almost

none. We tested this prediction experimentally, by constructing synthetic sRNA-target pairs on plasmids, and transforming them into *E. coli* cells. Synthetic sRNAs were based on RyhB, while targets consisted of fusion of a 5'-control-region of a RyhB target to a GFP reporter. Measuring single-cell fluorescence using flow cytometry allowed us to identify population diversity at the threshold of gene expression, as predicted by the model.

5.2 Model for sRNA regulation

Consider a simplified model for a small regulatory RNA which represses expression of a target gene by targeting its 5'-UTR. The two RNA species are coded by independent genes on the bacterial chromosome, and their synthesis can be independently regulated, e.g. by transcription factors, at the two promoters. Small RNAs and target mRNAs are therefore synthesized at two independent rates, denoted α_s and α_m respectively. Typical mRNAs in bacteria are subject to active degradation, which we assume to be linear and characterized by a rate β_m . Small RNAs may also be subject to active degradation or may be diluted due to cell growth. Both processes are coded as a linear depletion with rate β_s . Finally, binding of the small RNA to its target occludes ribosome binding and prevents translation initiation, and promotes degradation of both molecules. Within our model, this is summarized as an irreversible second-order binding process with rate k , after which the mRNA and sRNA molecules no longer participate in any gene expression process. The values of these parameters can be estimated for various sRNAs ³² (see also Table 4.1).

The master equation for the distribution of mRNA, sRNA and protein copy number (m , s and g , resp.) in this model is given by

$$\begin{aligned} \frac{d}{dt}P(m, s, g) = & \left[\left\{ \alpha_m (a_m - 1) + \beta_m (a_m^+ - 1) m \right\} \right. \\ & + \left\{ \alpha_s (a_s - 1) + \beta_s (a_s^+ - 1) s \right\} + k (a_m^+ a_s^+ - 1) ms \\ & \left. + \left\{ \gamma m (a_g - 1) + \delta (a_g^+ - 1) g \right\} \right] P(m, s, g), \end{aligned} \quad [5.2.1]$$

where the lowering and raising operators are defined through $a_x h(x) = h(x-1)$, $a_x h(0) = 0$, and $a_x^+ h(x) = h(x+1)$ for any function $h(x)$. In this paper we focus on the steady-state properties of gene expression, and therefore all time derivatives are hereafter set to zero.

Multiplying the master equation by either m or by s , and summing over m , s and g yields

$$\alpha_m - \beta_m \langle m \rangle - k \langle ms \rangle = 0, \quad \alpha_s - \beta_s \langle s \rangle - k \langle ms \rangle = 0 \quad [5.2.2]$$

Here $\langle \cdot \rangle$ denotes a steady-state ensemble average over different realizations of the stochastic process. As usual, one assumes ergodicity and identifies this average with the population average, namely the quantity measured in culture.

5.3 Mass-action model

As a first approximation, we ignore correlations between m and s , and replace the term $\langle ms \rangle$ on the right hand side of Eq. [5.2.2] by $\langle m \rangle \langle s \rangle$. The equations then have the same structure as the mass-action model presented in Chapter 4, *Quantitative Characteristics of Gene Regulation by Small RNA*. It is now straightforward to obtain the

steady-state concentrations of all molecule species, and investigate their dependence on the different parameters. Let us briefly review these results.

The expression of a target gene depends strongly on the ratio between the transcription rate of the target, α_m , and that of the small RNA, α_s . When the sRNA is produced at much higher rate, all mRNAs are rapidly targeted for degradation, and protein synthesis cannot take place. Gene expression is then repressed ($w\alpha_s$)-fold, where $w \equiv k/(\beta_m\beta_s)$ measures the efficiency of the coupled degradation channel. Based on available data, we estimate that for RyhB in *E. coli* at log phase, $w\alpha_s$ can be as large as $10^2 - 10^3$ (see Chapter 4, *Quantitative Characteristics of Gene Regulation by Small RNA*). In contrast, if the transcription rate of the target gene significantly exceeds that of the small RNA, repression is effectively relieved. The concentration of mRNA is then given by $(\alpha_m - \alpha_s)/\beta_m + O(w^{-1})$. In summary, for large w gene expression is effectively silenced below the threshold $\alpha_m \approx \alpha_s$, and increases linearly with the promoter activity above it. The sharpness (‘sensitivity’) of the transition at the threshold increases with w . This *threshold linear* behavior is depicted in the inset of Fig. 5.1. This behavior was tested experimentally for RyhB, an endogenous small RNA in *E. coli* (see Chapter 4, *Quantitative Characteristics of Gene Regulation by Small RNA*), as described below (see Fig. 5.6).

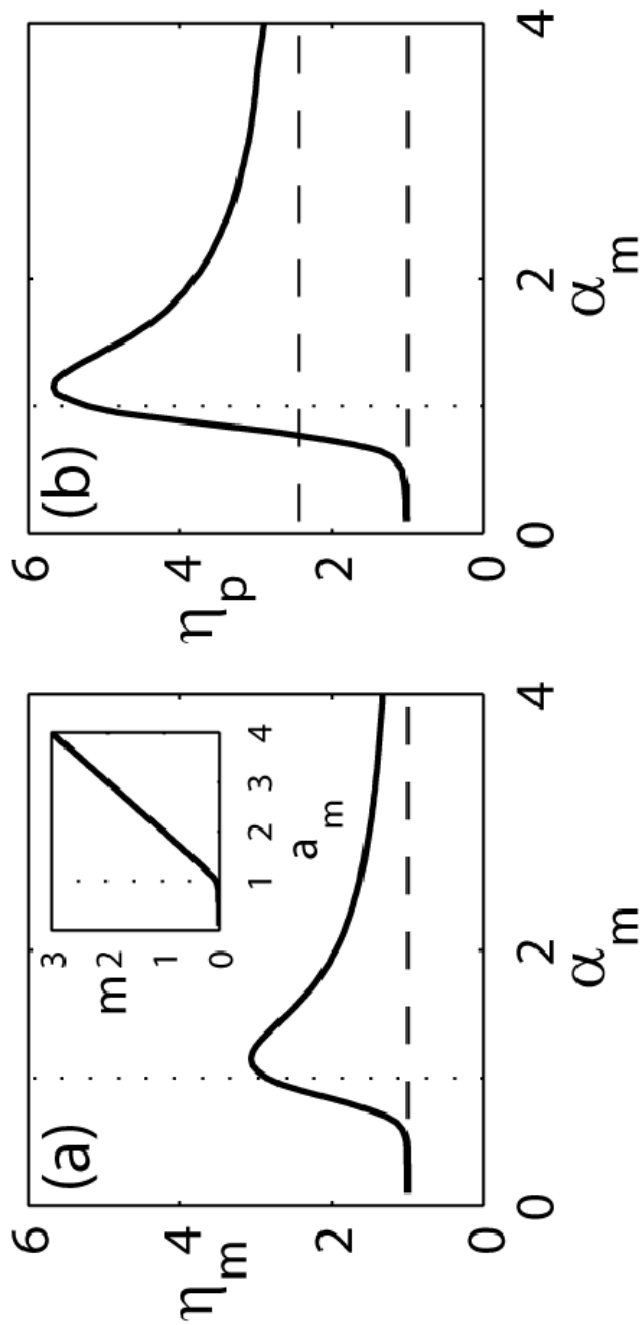


Figure 5.1: Noise indices for target mRNA and protein. (a) For mRNA, noise approaches the ‘bare’ Poissonian limit $\eta_m = 1/\text{min}$ (dashed line) for high and low transcription rates, but shows enhanced fluctuations near the threshold (vertical dotted line, $\alpha_m \approx \alpha_s = 1/\text{min}$). (b) For proteins, the ‘bare’ noise level depends on the burst size (upper horizontal dashed line, see text). One observes in addition that noise is attenuated at low expression rates. Here $\beta_m = \beta_s = 0.02/\text{min}$ and $k = 0.1$. (inset) Threshold-linear behavior of the mass-action model.

5.4 Noise in target gene expression

We now turn to describe internal noise in the expression of a gene regulated by a small RNA. We focus on a noise index $\eta_x \equiv C_{x,x} / \langle x \rangle$, where $C_{x,y} = \langle xy \rangle - \langle x \rangle \langle y \rangle$ is the steady-state correlation between the copy numbers of molecules x and y (so $C_{x,x}$ is the variance of x). Note that η_x has the same dimensions as a molecule number.

Multiplying the master equation by m^2 or s^2 , summing over all variables, and using Eq. [5.2.2], one finds that at steady state

$$\alpha_m - \beta_m C_{m,m} - k C_{m,ms} = 0, \quad \alpha_s - \beta_s C_{s,s} - k C_{s,ms} = 0. \quad [5.4.1]$$

Again, one can make the approximation where m and s are uncorrelated, in which case the variance of each species follows the same steady-state equation as its mean (cf. Eq. [5.2.2]). For the mRNA, these equations correspond to a birth-death process, with birth rate α_m and death rate $\beta_m^* = \beta_m + k \langle s \rangle$. Under this approximation $\eta_m = \eta_s = 1$, just as for a Poisson process. However, it is natural to suspect that at least for some choices of parameters m and s should be highly correlated and this approximation should not hold. In Fig. 5.1a we plot the noise index η_m as obtained from exact numerical integration of the master equation, Eq. [5.2.1]. When the target transcription rate is very high or very low compared with that of the small RNA, η_m indeed approaches one. However, when the two rates are comparable, correlations between the two species cannot be neglected. In fact, strong anti-correlation between the two results in a large noise index. This phenomenon will be addressed theoretically and experimentally below.

In most cases, the essential gene product to the cellular program is the proteins, rather than the transcribed RNA. Indeed, the known natural targets of RyhB are all

structural proteins^{28,30}, and accordingly the synthetic reporter we will use below to test our model is the green fluorescent protein (GFP). We therefore direct our attention to η_p , plotted in Fig. 5.1*b*. In Fig. 5.1*b* we also plot the noise index for a protein that is not regulated by a small RNA, but otherwise have the same characteristics (upper horizontal dashed line). For such a protein it is straightforward to show that $\eta_p = 1 + b / (1 + \delta / \beta_m)$, where $b \equiv \gamma / \beta_m$ is the size the burst of proteins translated from a single mRNA¹⁰.

Two features of the mRNA noise index propagate to the protein level. First, when the target transcription rate is much larger than that of the sRNA, noise is not affected by the sRNA regulation. Also, as the transcription levels become comparable, the strong noise in the mRNA level is also observed for the protein. One new feature is found as the mRNA transcription level becomes even lower. Here, the noise in the protein level is significantly reduced as compared with an sRNA-free gene. This will be the focus of the next section.

5.5 Population diversity at the threshold – Model

According to the threshold-linear picture presented above (inset of Fig. 5.1), whenever transcription of the two genes – the small RNA and its target – occurs at similar rates, the steady-state average of both species is considerably suppressed. In this case, where all participants in the process are at low numbers, fluctuations can dominate. This situation has been studied in another context in Ref. 35 within a linear noise approximation. This approximation proved useful in quantifying the strong anti-

correlation between species, but misses the main feature discussed below, namely the population diversity.

To gain insight into the nature of these fluctuation, Fig. 5.2a shows typical temporal evolution of the sRNA and target copy number in our model, as obtained from Monte-Carlo simulations³³. While the average mRNA number $\langle m \rangle$ (as obtained by numerical solution of the master equation and compared with the simulation) is not very different from the one predicted by the mass-action model, these averages do not capture the dynamical behavior of the mRNA copy number. In fact, simulations suggest that the cell goes through alternating periods of complete mRNA suppression (with high level of sRNA) and mRNA expression (with the sRNA suppressed). Starting from a state where both RNA species are suppressed, a small fluctuation makes the mRNA the majority species. In this state, most mRNAs decay through the sRNA-independent degradation pathway. At the same time, the developing mRNA pool keeps the sRNA level suppressed. After a while, fluctuations drive the mRNA pool back to a level comparable to that of the sRNA. Now, similar fluctuations can drive the cell to another round of abundant mRNA, or – as it happens in this particular realization – to a round of abundant sRNA, accompanied by mRNA suppression.

These simulations suggest an appealing scenario, where the transition from the silenced regime ($\alpha_m \ll \alpha_s$) to the active regime ($\alpha_m \gg \alpha_s$) occurs via separation of the probability $P(m)$ into two modes: One mode is a reminiscent of the silenced state, while the other is the bud of the active state. This scenario is depicted in Fig. 5.2b, where $P(m)$ - as obtained from numerical solution of the master equation, Eq. [5.2.1] - is followed through the transition. Just above the transition one indeed finds a regime where

the population of cells is divided into two sub-populations, which do or do not express the target gene.

To gain insight into the mechanism behind this behavior, consider the following picture. The steady-state dynamics of the mRNA molecule number, $m(t)$, can be described as a random walker on the positive axis. This number is increased by one due to a transcription event, with rate α_m . It is decreased by one whenever an mRNA is degraded by a ribonuclease, independent of the small RNA interaction, a process that occurs with rate $\beta_m m(t)$. Finally, m also decreases by one whenever an sRNA binds to an mRNA, a process that occurs with rate $km(t)s(t)$. At steady state, it is tempting to use the mass-action equations and let $km(t)s(t) = \alpha_s [m / (m + \beta_s / k)] < \alpha_s$. Under this description the random walker is biased to the right, and is only kept bound by the independent degradation. However, at small and intermediate values of m , correlations between the small RNA and its target cannot be neglected. A better estimate for the left-hopping rate due to coupled degradation would therefore be $k \langle ms | m \rangle$. This function is plotted in Fig. 5.2c, and compared with the mass-action result. The two estimates approach α_s at high values of m . On the other hand, taking correlations into account significantly increases the coupled degradation rate at low – but finite – values of m . Thus, (anti-) correlations between the sRNA and its target form a potential barrier, which makes it hard for $m(t)$ to escape the origin. This potential barrier shapes the bimodal distribution, as seen in Fig. 5.3.

To study the dependence of this bimodal behavior on the model parameter, we draw in Fig. 5.4 the ‘phase diagram’ of the model. Bimodality takes place just above the threshold $\alpha_m \geq \alpha_s$. Strong interaction between the sRNA and its target (large k) increases

the anti-correlations between that two species, and extends the range of α_m at which bimodality occurs. In addition, slower degradation rate of the target makes this range of α_m values larger, and also allows for bimodality at smaller k .

Our estimation of the relevant parameters for RyhB and its targets in *E. coli* seem, in fact, not to allow for bimodality. Nevertheless, as the phenomenon described here is induced by noise, other sources of noise may in fact contribute to expand the region of bimodality. For example, a more detailed account for the intermediate ribonucleoprotein complex formed between the small RNA, its target and Hfq allows for bimodality even outside the domain defined in Fig. 5.4 (data not shown). In addition, higher values of the synthesis rates of the two RNA genes, which favors bimodality, may be relevant for fast growing cells, or in the case where these genes are coded on multicopy plasmids.

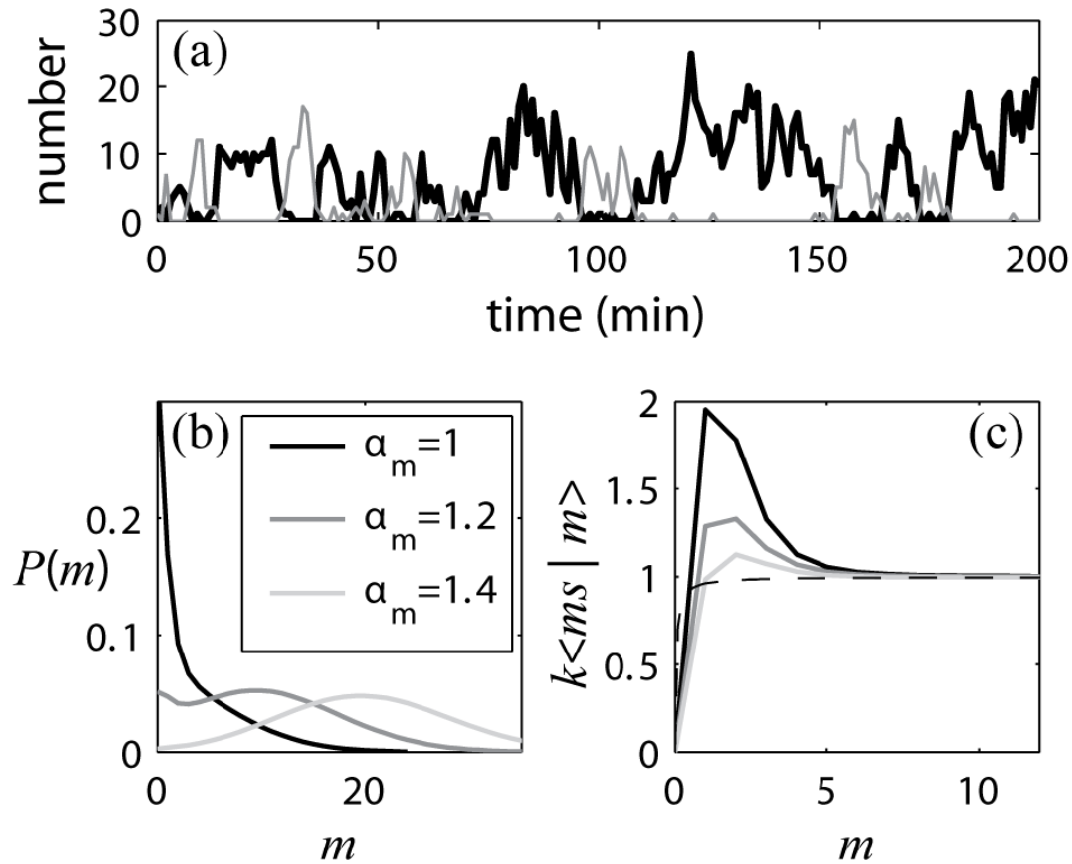


Figure 5.2: Diversity in gene expression near the transition. (a) Temporal evolution of mRNA (black) and sRNA (gray) expression levels in a Monte-Carlo simulation. Here $\alpha_m = 1.2$, $\alpha_s = 1$, $\beta_m = \beta_s = 0.02$, $k = 0.5$. (b) Probability distribution for the mRNA number at steady-state, calculated by direct numerical integration of the master equation Eq. [5.2.1]. The coexistence of two populations at the onset of gene expression appears here as two modes in the distribution function. (c) Correlations between the numbers of mRNA and sRNA molecules amplify the coupled degradation at low but finite m . For comparison, dashed line is the coupled degradation strength as predicted by the mass-action model.

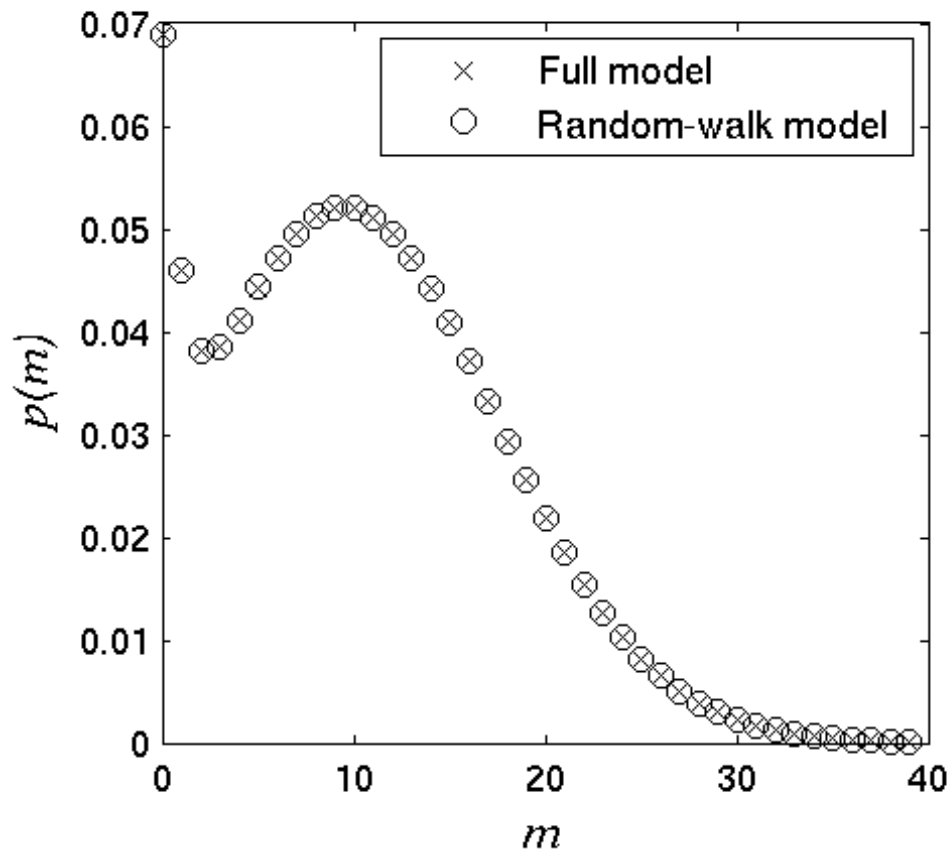


Figure 5.3: Random-walk picture. The random-walk picture maps the dynamics of the mRNA number, m , onto a random walker on the positive axis, which takes steps to the right with rate α_m and steps to the left with rate $\beta_m + k\phi(m)$, where $\phi(m) = \langle ms|m \rangle$. The steady state distribution of such a random walker is given by

$$P(m) = P(0) \prod_{n=1}^m \frac{\alpha_n}{\beta_n + \phi(n)}.$$

In this picture the shape of $\phi(m)$ determines the number of maxima – namely, the number of modes – in the distribution function. To demonstrate the applicability of this picture, we compare in this figure the steady-state distribution function, as obtained from an exact numerical solution to the master equation, with the steady-state distribution function of the random walker. The function $\phi(m)$ is again obtained from the master equation (in this sense, the random-walker picture allows for an intuitive description of the problem, but does not help in solving it). Parameters are the same as in Fig. 5.2.

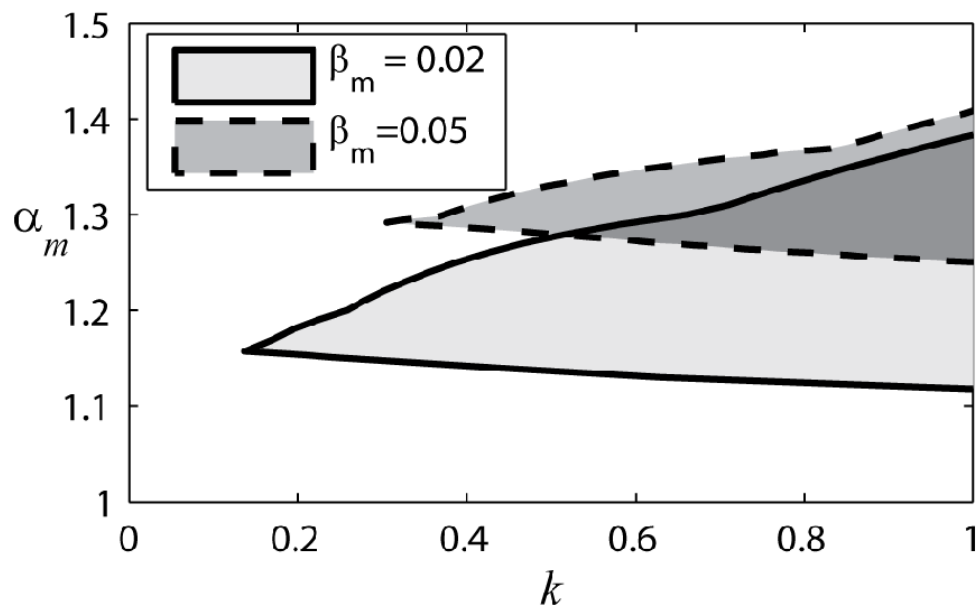


Figure 5.4: Phase Diagram. The domains of bimodality (filled areas) were determined through exact numerical integration of the master equation, with $\alpha_s = 1/\text{min}$ and $\beta_s = 0.02/\text{min}$.

5.6 Population diversity at the threshold – Experiment

The model presented above is consistent with available data about RyhB, an endogenous small RNA in *E. coli*, involved in regulation of iron homeostasis and oxidative stress^{24,29,31}. Expression of the *ryhB* gene is regulated by Fur, the ferric uptake regulator, which inhibits gene expression in response to high cellular levels of free Fe^{2+} ions. At least 18 genes are directly repressed by RyhB²⁸⁻³⁰, resulting in reduced accumulation at low concentration of intra-cellular iron³⁶. Among these genes are iron storage genes, building blocks for iron-sulfur clusters, and oxidative stress response genes. Most of these genes are subject to complex transcription regulation at their promoters, and thus their mRNA expression can be considered independent of that of RyhB. RyhB and at least some of its targets bind the RNA chaperone Hfq^{31,37-39}. Binding to Hfq may stabilize the mRNA and promotes the interaction between the small RNA and its targets^{31,40}. Upon binding, the two molecules are degraded through an RNase E dependent process^{29,31,38}.

To test the results of our model, we constructed a synthetic pair of small RNA and its target. The *ryhB* gene was cloned from the *E. coli* chromosome onto a medium-copy plasmid. Expression of *ryhB* was driven by the Ptet promoter⁴¹. This promoter is inducible by anhydrotetracycline (aTc), which allowed us to control the synthesis rate of the small RNA (α_s). As a target, we fused the control region of *sodB*, one of the strongest known targets of RyhB, to a *gfp* gene. The target was driven by the *lac* promoter, inducible by Isopropyl beta-D-1-thiogalactopyranoside (IPTG)⁴¹.

For each composition of inducers, we estimated the activity of the target promoter by measuring fluorescence from cells with no RyhB source. We then plotted the

fluorescence from cells with plasmid borne RyhB against the promoter activity (Fig. 5.5). The quality of a single-parameter fit of these data to the mass-action model suggested that this approximation captures the main features of bulk activity (see Chapter 4, *Quantitative Characteristics of Gene Regulation by Small RNA*).

Addressing fluctuations refined our predictions for the *threshold-linear* response of a target gene to activity at its promoter. For $\alpha_m \ll \alpha_s$, gene expression is suppressed in all cells. At the onset of gene expression, $\alpha_m \geq \alpha_s$, a snapshot of a cell population at a given point in time is predicted to capture a sub-population of cells expressing the target gene, and another sub-population which does not. Finally, above the threshold, $\alpha_m \gg \alpha_s$, almost full expression is expected in all cells.

To test this prediction, we go back to Fig. 5.5, and observe that with no IPTG added to the media, α_m is very small, much smaller than α_s . Also, IPTG concentration above 0.5 mM corresponds to $\alpha_m \gg \alpha_s$. According to the fit to our model, the threshold is expected around $[\text{IPTG}] \approx 0.3 \text{ mM}$. It is around this concentration of IPTG that we expect to find two sub-populations of cells.

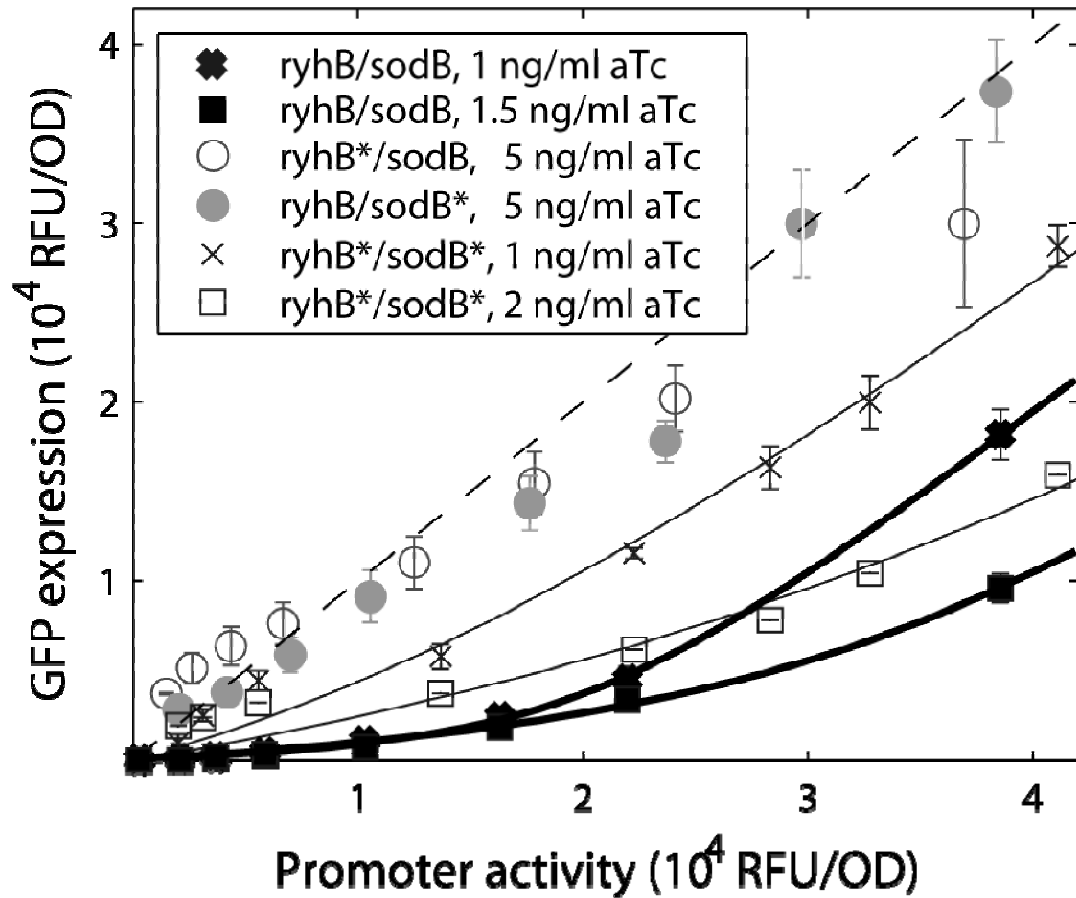


Figure 5.5: Bulk response of a target gene to small RNA regulation. Promoter activity is defined as the GFP fluorescence level in a strain with no *ryhB* gene. This activity is controlled by changing the concentration of IPTG in the media (see Section 5.9, *Medium, growth, measurements*). In cells bearing *ryhB* on a plasmid, aTc concentration controls the expression level of this gene. Data was fitted to the solution of the mass-action model using one global parameter for each pair, and an additional parameter for each aTc level; see Chapter 4, *Quantitative Characterization of Gene Regulation by Small RNA*

We collected single-cell fluorescence data using flow-cytometry. Typical histograms are plotted in Fig. 5.6*a*. At low IPTG concentrations, all cells show only background level of fluorescence, while at high level of IPTG virtually all cells show high fluorescence level. However, at intermediate levels of IPTG, the histogram is bi-modal. One mode corresponds to cells whose fluorescence is at or near the background level; the other mode corresponds to cells expressing some amount of GFP.

In order to show that this effect is the result of an interaction between the reporter target and RyhB, we perform a single mutation in the 5'-control region of our target. This mutation, C $\rightarrow$ G, is located one nucleotide away from the start codon in the core recognition sequence of RyhB and the SodB mRNA³¹. Plasmids harboring this mutated target, *crsodB**-*gfp*, were transformed into *ryhB*⁻ cells with either a *ryhB*-harboring plasmid or a control plasmid. Bulk fluorescence from the two strains was indistinguishable (Fig. 5.5), suggesting that interaction between RyhB and the target has been completely removed. Single cell *crsodB**-*gfp* fluorescence data is shown in Fig. 5.6*b*. These histograms suggest a homogeneous population of cells at all IPTG levels. As promoter activity is turned on, the population mean and variance of protein concentrations is smoothly increased. Similar results were obtained by removing the control-region altogether (Fig. 5.7). Thus, interaction between the target and RyhB is required for the bi-modality at the onset of gene expression. For a comparison of our full data sets of the two targets, see Fig. 5.8.

It is also possible that the observed RyhB-dependent bimodality is an indirect effect of RyhB interactions with many endogenous targets. To address this possibility, we mutated a single nucleotide in *ryhB* at the position complementary to the mutation in

*crsodB**. Plasmids harboring this small RNA gene, *ryhB**, were transformed into *ryhB** cells along with plasmids coding for either the target with the wild-type control region, *crsodB-gfp*, or the one with the mutated control-region, *crsodB*-gfp*. Measurements of bulk fluorescence are given again in Fig. 5.5. These data suggest that RyhB* interacts with *crsodB*-gfp*, in the same threshold-linear manner that characterizes the interaction between RyhB and *crsodB-gfp*, while RyhB* does not affect the fluorescence of *crsodB-GFP*. Using RT-PCR we verified that RyhB* expression has also no effect on the mRNA levels of chromosomal *sodB* and *fumA*, two endogenous targets of RyhB²⁸ (data not shown). Taken together, these results suggest that RyhB* does not interact with a wild-type control regions targeted by RyhB. Single-cell data for the RyhB* – *crsodB*-gfp* pair are presented in Fig. 5.6c. As for the wild-type small RNA and control region, we find again homogeneous populations of cells in the low and high ends of gene expression, and inhomogeneous populations at intermediate levels.

To test our predictions further, we make use of the fact that RyhB* is expressed from the Ptet promoter, inducible by aTc. Thus, keeping the IPTG level constant at some intermediate level (i.e., constant α_m), we can tune α_s by changing the concentration of aTc in the media. Typical histograms are shown in Fig. 5.9. As the small RNA transcription level is increased, the sub-population of cells which show background fluorescence increases on the expense of the other sub-population, in accordance with our prediction.

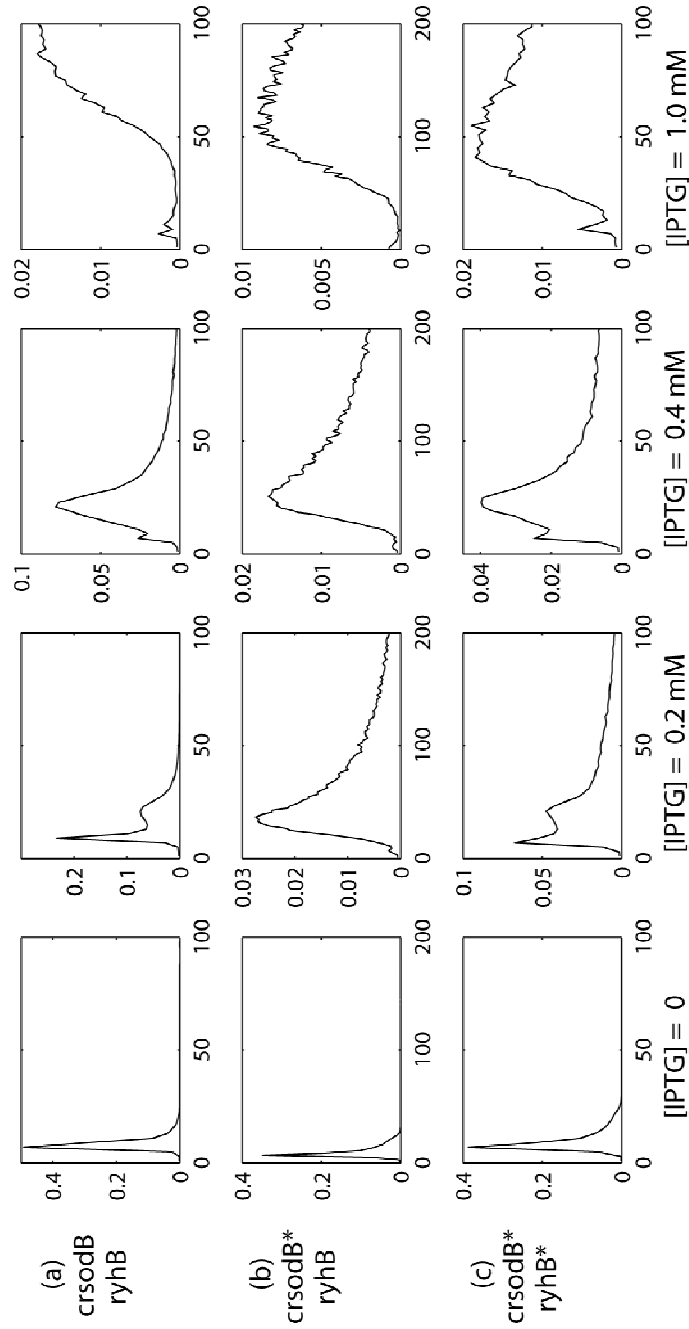


Figure 5.6: Histograms of single-cell fluorescence in response to small RNA regulation, obtained by flow-cytometry. The level transcription of target, expressed from the *lac* promoter⁴¹ is induced by IPTG, while the level of RyhB expression from the *tet* promoter remains constant in these experiments. Concentrations of aTc were 1.5 ng/ml (ryhB experiments) and 2 ng/ml (ryhB* experiments). Bimodality is observed near the threshold in the cases of matching sRNA and target.

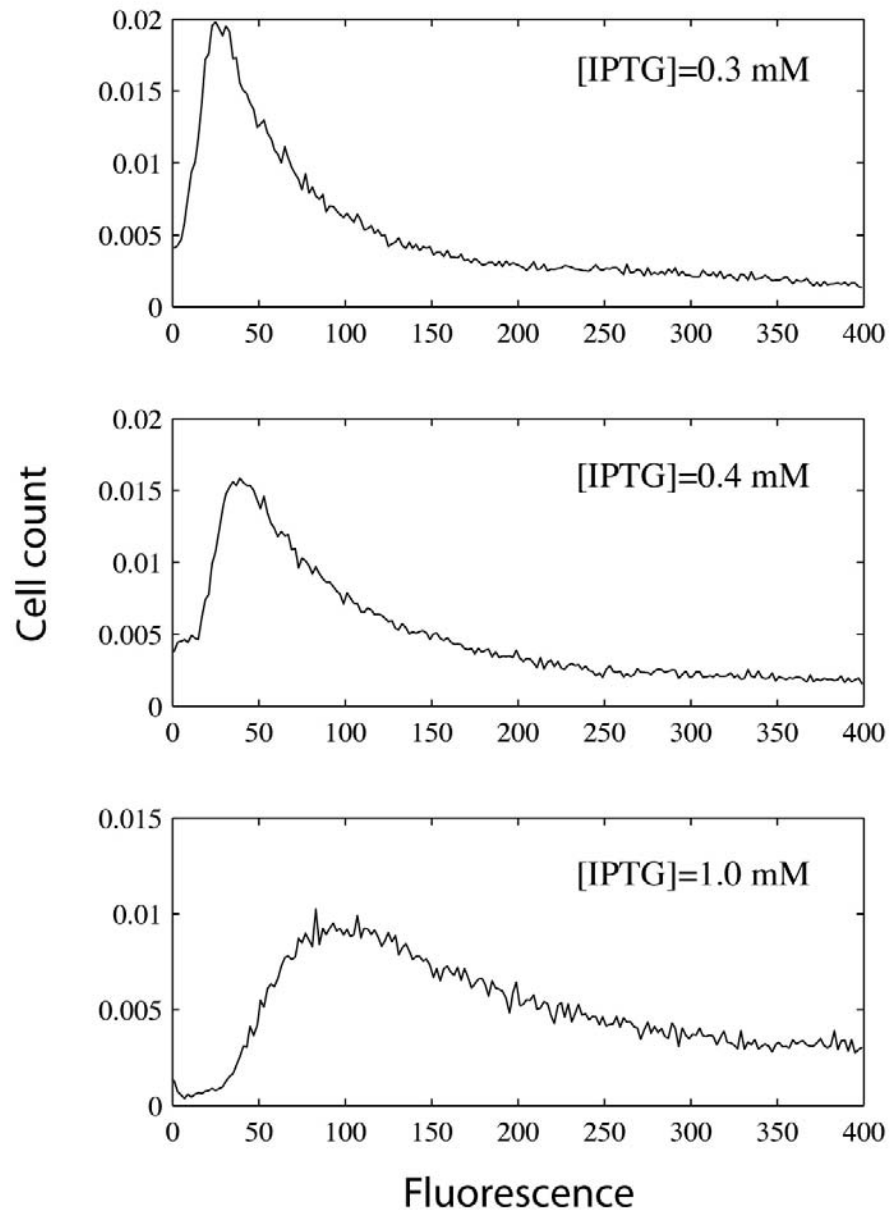


Figure 5.7: Single cell expression level of P_{L-lac} -*gfp*. In the absence of any structured 5'-UTR the mRNA of the *gfp* gene does not interact with RyhB, and one observes no bimodality. Here aTc concentration is 5 ng/ml.

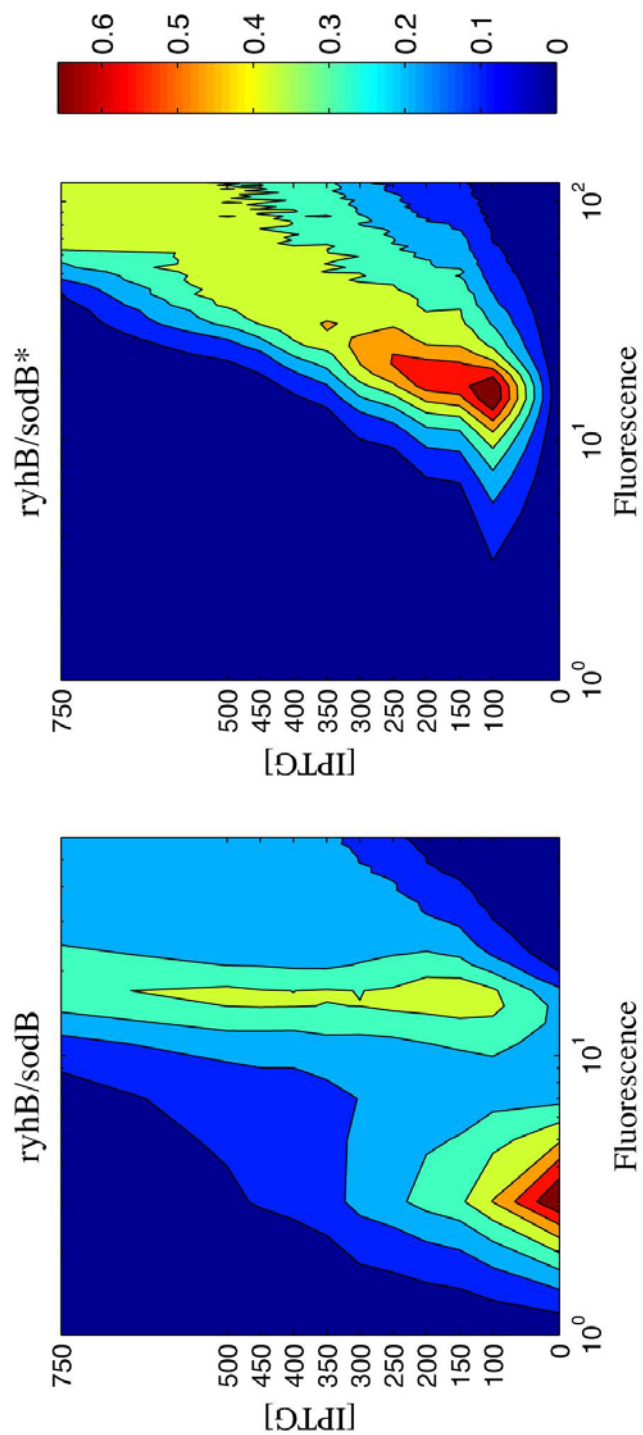


Figure 5.8: Flow cytometry data for RyhB/crsodB-gfp and RyhB/crsodB*-gfp. Contours are based on $g \cdot C(g)$, where $C(g)$ is the count of cell of fluorescence g . Concentration of aTc in both experiments is 2 ng/ml.

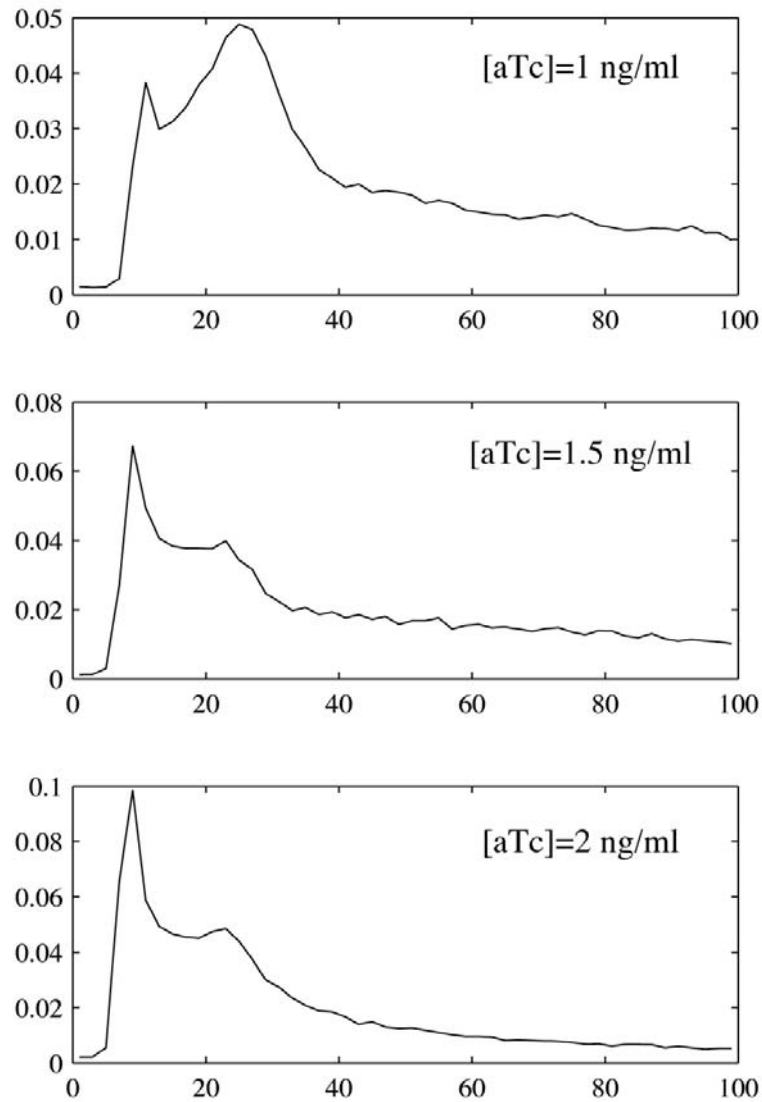


Figure 5.9: Single cell expression levels of $P_{L-lac}-crsodb-gfp$ at different levels of *ryhB* expression. Histograms of single-cell fluorescence in response to small RNA regulation, obtained by flow-cytometry. The level transcription of the small RNA, expressed from the *tet* promoter⁴¹ is induced by IPTG, while the level of target expression from the *lac* promoter remains constant in these experiments. Concentration of IPTG was 0.2 mM.

5.7 Noise characteristics of sRNA vs. protein regulation

When gene expression is silenced by the small RNA, the noise index for mRNA approaches the limit of ‘bare’ Poisson noise. For proteins noise is considerably reduced, an exciting feature given the fact that in this region the average number of molecules is kept very low.

To gain insight into the mechanism of noise attenuation, let us compare the nature of this basal expression – namely gene expression in the repressed state – for two genes. Expression of the first, which we call gene g_R , is repressed at the transcription level by a protein transcription repressor. The second, named gene g_S , is repressed post-transcriptionally by a small RNA. An efficient transcription factor can keep the mRNA synthesis rate (which, at steady state, is the same as the rate of transcription initiation) very low; the small RNA makes the life-time of its target mRNA very low. We can therefore conduct the numerical experiment where the two have the same effect, yielding equal mean mRNA copy numbers for the two genes.

In Fig. 5.10a we present typical results of Monte-Carlo simulations³³ for the two genes. Expression of gene g_R is characterized by rare ‘bursts’ of protein, embedded in long periods with not a single protein in the cell. Each burst of proteins corresponds to a single transcription event. The stray mRNA is then amplified by a series of translation events, until the mRNA molecule is degraded. The burst size is therefore given by $b = \gamma / \beta_m$. This type of kinetics has been recently observed experimentally, by tracking expression of Venus-Tsr fusion protein from the *lac* promoter, controlled by the *lac* repressor¹⁷.

Let us now follow the kinetics of gene g_s . Here the expression level is much smoother, in the sense that no two distinct behaviors are observed. Rather, gene expression seems to fluctuate around some mean. This is understood by the fact that mRNA transcripts are being constantly synthesized. In the silenced regime, the noise index $\eta_m \simeq 1$, which means that m and s are weakly correlated^{§§}. In this case, as discussed following Eq. [5.4.1], the moments of m are described by the same equations as for a birth-death process, with an effective death rate $\beta_m^* \equiv \beta_m + k\langle s \rangle$. The mRNA transcripts are short-lived, resulting in a decreased burst size $b^* = \gamma / \beta_m^*$. This manifests itself in a reduced noise index, $\eta_p = 1 + b^* / (1 + \delta / \beta_m^*) \simeq 1 + b^*$. Note that the burst size b – and with it the noise index η_p – is now increasing with α_m : an increased target transcription rate results in a smaller sRNA pool $\langle s \rangle$, and therefore in larger bursts.

Another possible manifestation of the different kinetics in the two cases is the shape of the probability distribution, demonstrated in Fig. 5.1*b*. For a translationally repressed gene, the maximum probability occurs at zero protein expression. For an sRNA repressed gene with the same average, the probability may be maximized at low but finite protein number, as in the example given here.

Can one probe these noise properties experimentally? Naturally, with multi-copy plasmid, noise is expected to become smoother even in the transcription-regulated case, as the burst frequency is approximately linear in the plasmid copy number. A single chromosomal reporter target is required, accompanied by a single-molecule resolution, as

^{§§} Using a linear noise approximation³⁴ one can show that for $\alpha_s \ll \alpha_m$ the correlation coefficient $C_{m,s} / (\sigma_m \sigma_s) \simeq \sqrt{w^{-1} \alpha_m / \alpha_s^2}$

in Ref. 17. However, the temporal resolution of a system based on fluorescent reporter is estimated at about 3 minutes¹⁷. While this resolution is good enough to track rare bursts, it is not expected to suffice for the smooth kinetics of an sRNA target.

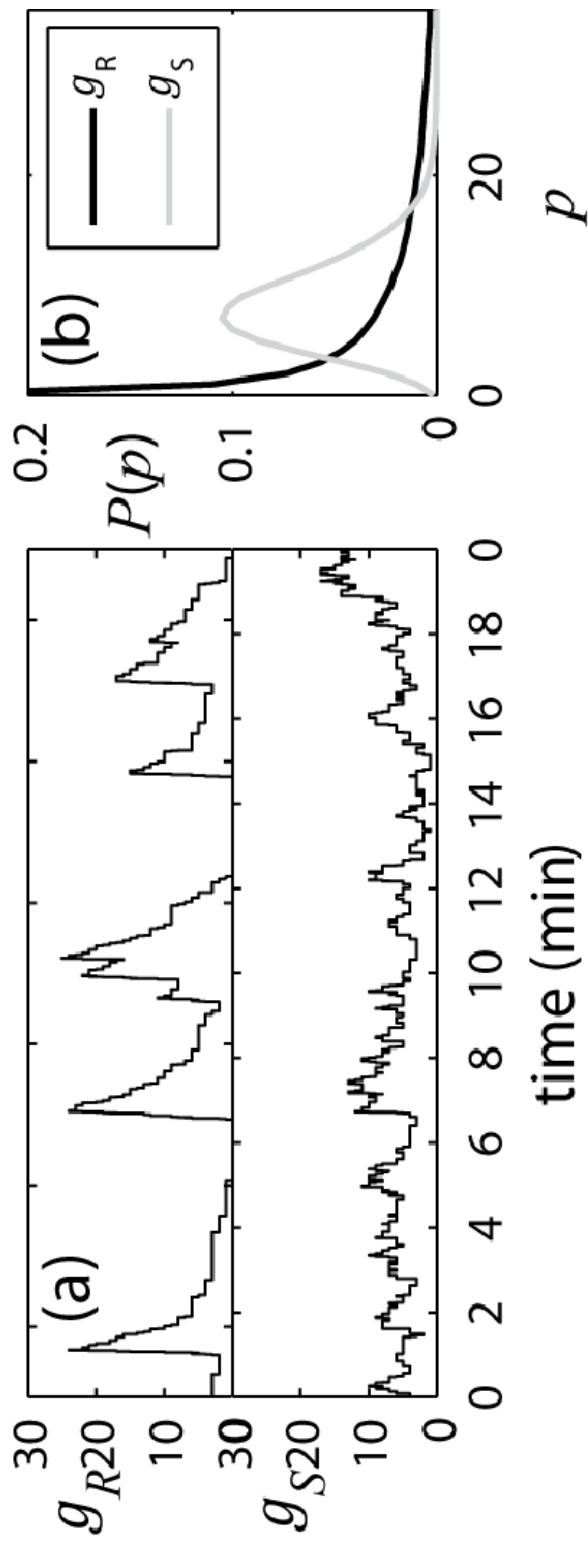


Figure 5.10: Noise properties in the silenced state. (a) Typical temporal evolution for a transcriptionally repressed gene (g_R) and an sRNA target (g_S), as obtained from Monte-Carlo simulations. Here, the transcription rates g_R and g_S are adjusted such that the mean mRNA copy numbers are equal; $\alpha_m = 0.084$ (g_R) and 0.3 (g_S), $\alpha_s = 1$, $\beta_m = 0.1$, $\beta_s = 0.02$, $k = 0.1$, $\gamma = 2$ and $\delta = 0.02$, all in min^{-1} . (b) Steady-state distributions for the two genes, obtained from Monte-Carlo simulations.

5.8 Materials and Methods: Strains & plasmids.

All experiments were performed with BW-RI cells derived from *E. coli* K-12 BW25113⁴², with the transfer of the *spr-lacI-tetR* cassette from DH5 α -ZI cells⁴¹ by phage P1 transduction. This cassette provides the constitutive expression of the *lacIq* and the *tetR* genes⁴¹. Chromosomal *ryhB* was deleted from BW-RI⁴². These strains were then transformed by the following "target" and "source" plasmids. All strains and plasmids used are summarized in Table 5.1.

5.9 Medium, growth, measurements

BW-RI strains each containing the target and/or source plasmids were grown in M63 minimal media with standard concentrations of the appropriate antibiotics. The overnight cultures were diluted into fresh M63 media (OD600 $\approx$ 0.002) containing the appropriate antibiotics as well as varying amounts of the inducers (aTc, IPTG) in the wells of 12-well plates. The plates were incubated with shaking at 37° C and taken for OD600 and fluorescence measurements every hour for up to 8 hours (until a final OD600 of 0.2 - 0.3) using a TECAN Genios-Pro plate reader. Each measurement was repeated 3 times and the data was analyzed as described in Chapter 4, *Quantitative Characteristics of Gene Regulation by Small RNA*.

To obtain single-cell data, 1.7 ml of each culture was spun down and resuspended in 1 ml PBS. Expression data were collected using a Becton-Dickinson FACS Calibur flow cytometer with a 488-nm argon excitation laser and a 515- to 545-nm emission filter (FL1) at a low flow rate. Photomultiplier tube (PMT) voltage was set to 950V, and a

linear amplifier was set at 9.5X. Forward scatter and fluorescence values were collected for 50,000 cells.

Table 5.1. Strains & plasmids. All coordinates are relative to +1 position of the given gene.

Strain	Genotype	Derived From	Comments
BW-RI	<i>spr-lacI-tetR</i>	BW25113	<i>spr-lacI-tetR</i> cassette derived from DH5 α -ZI (Ref. 41)
BW-NULL	pNULL	BW-RI	Negative control
ZZS00	Δ <i>ryhB</i>	BW-RI	<i>ryhB</i> deletion from -54 to +94
ZZS23	Δ <i>ryhB</i> pZE12S pZA31R	ZZS00	
ZZS23*	Δ <i>ryhB</i> pZE12S pZA31R*	ZZS00	
ZZS2*3	Δ <i>ryhB</i> pZE12S* pZA31R	ZZS00	
ZZS13	Δ <i>ryhB</i> pZE12G pZA31R	ZZS00	
ZZS21	Δ <i>ryhB</i> pZE12S pZA31- <i>luc</i>	ZZS00	
Plasmid	Genotype	Source	Comments
pZE12		Ref. 41	ColE1 <i>ori</i> , Amp marker
pZA31- <i>luc</i>	<i>P_{LTet-O1}:luc</i>	Ref. 41	p15A <i>ori</i> , Cm marker
pNULL	<i>PNULL:gfpmut3b</i>	Chapter 4	Promoterless <i>gfpmut3b</i> - negative control;
pZE12G	<i>P_{Llac-O1}:gfpmut3b</i>	Chapter 4	
pZE12S	<i>P_{Llac-O1}:crsodB-gfpmut3b</i>	Chapter 4	<i>crsodB</i> amplified from +1 to +88
pZE12S	<i>P_{Llac-O1}:crsodB-gfpmut3b</i>	This study	C→G mutation at +54 of <i>crsodB</i>
pZA31R	<i>P_{LTet-O1}:ryhB</i>	Chapter 4	<i>ryhB</i> amplified from +1 to +96
pZA31R	<i>P_{LTet-O1}:ryhB</i>	This study	G→C mutation at +44 of <i>ryhB</i>

5.10 Acknowledgment

We are grateful to Natalie Ostroff and Jeff Hasty for advice and hospitality. This work was supported by NSF through the PFC-sponsored Center for Theoretical Biological Physics (Grants No. PHY-0216576 and PHY-0225630). TH acknowledges additional support by NSF Grant No. DMR-0211308.

Chapter 5 contains material in Levine E, Kuhlman T, Zhang Z, Huang M, Huang Y, and Hwa T. Noise and silence in small RNA regulation. *Proc Natl Acad Sci*, in preparation, used with permission. The dissertation author was the primary experimental investigator of this work.

5.11 References

1. Guptasarma P. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays*. 1995 Nov;17(11):987-97.
2. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*. 2005;6:451-64.
3. Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. *Science (New York, NY)*. 2005 Sep 23;309(5743):2010-3.
4. Kussell E, Leibler S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science (New York, NY)*. 2005 Sep 23;309(5743):2075-8.
5. Kearns DB, Losick R. Cell population heterogeneity during growth of *Bacillus subtilis*. *Genes & development*. 2005 Dec 15;19(24):3083-94.
6. Suel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*. 2006 Mar 23;440(7083):545-50.
7. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science (New York, NY)*. 2002 Aug 16;297(5584):1183-6.
8. Swain PS, Elowitz MB, & Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA*. 2002;99:12795-800.
9. McAdams HH, Arkin A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*. 1997 Feb 4;94(3):814-9.
10. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2001 Jul 17;98(15):8614-9.
11. Paulsson J. Summing up the noise in gene networks. *Nature*. 2004 Jan 29;427(6973):415-8.
12. Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS, Hasty J. Origins of extrinsic variability in eukaryotic gene expression. *Nature*. 2006 Feb 16;439(7078):861-4.

13. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nature genetics*. 2002 May;31(1):69-73.
14. Becskei A, Kaufmann BB, van Oudenaarden A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature genetics*. 2005 Sep;37(9):937-44.
15. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. *Cell*. 2005 Dec 16;123(6):1025-36.
16. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature*. 2006 Mar 16;440(7082):358-62.
17. Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. *Science (New York, NY)*. 2006 Mar 17;311(5767):1600-3.
18. Bar-Even A, Paulsson J, Maheshri N, et al. Noise in protein expression scales with natural protein abundance. *Nature genetics*. 2006 Jun;38(6):636-43.
19. Swain PS. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J Mol Biol*. 2004 Dec 3;344(4):965-76.
20. Walczak AM, Onuchic JN, Wolynes PG. Absolute rate theories of epigenetic stability. *Proceedings of the National Academy of Sciences of the United States of America*. 2005 Dec 27;102(52):18926-31.
21. Pedraza JM, van Oudenaarden A. Noise propagation in gene networks. *Science (New York, NY)*. 2005 Mar 25;307(5717):1965-9.
22. Ambros V. The functions of animal microRNAs. *Nature*. 2004;431:350-5.
23. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*. 2004;116:281-97.
24. Gottesman S. The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu Rev Microbiol*. 2004;58:303-28.
25. Gottesman S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet*. 2005;21:399-404.
26. Lease RA, Smith D, McDonough K, Belfort M. The small noncoding DsrA RNA is an acid resistance regulator in *Escherichia coli*. *J Bacteriol*. 2004 Sep;186(18):6179-85.

27. Opdyke JA, Kang JG, Storz G. GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol.* 2004 Oct;186(20):6698-705.
28. Masse E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA.* 2002;99:4620-5.
29. Masse E, Escorcia FE, Gottesman S. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev.* 2003;17:2374-83.
30. Masse E, Vanderpool CK, Gottesman S. Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol.* 2005;187:6962-71.
31. Geissmann TA, Touati D. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J.* 2004;23:396-405.
32. Wagner EG, Altuvia S, Romby P. Antisense RNAs in bacteria and their genetic elements. *Adv Genet.* 2002;46:361-98.
33. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem.* 1977;81:2340-61.
34. Kampen NV. *Stochastic Processes in Physics and Chemistry.* 2 ed. Amsterdam: North-Holland; 2001.
35. Elf J, Paulsson J, Berg OG, Ehrenberg M. Near-critical phenomena in intracellular metabolite pools. *Biophys J.* 2003 Jan;84(1):154-70.
36. Jacques JF, Jang S, Prevost K, et al. RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in *Escherichia coli*. *Molecular microbiology.* 2006 Nov;62(4):1181-90.
37. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol.* 2003;50:1111-24.
38. Moll I, Afonyushkin T, Vytvytska O, Kaberdin VR, Blasi U. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA.* 2003;9:1308-14.
39. Sonnleitner E, Napetschnig J, Afonyushkin T, et al. Functional effects of variants of the RNA chaperone Hfq. *Biochem Biophys Res Commun.* 2004;323:1017-23.
40. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell.* 2002 Jan;9(1):11-22.

41. Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* 1997 Mar 15;25(6):1203-10.
42. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 use in PCR products. *Proc Natl Acad Sci USA.* 2000;97:6640-5.