View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DSpace@MIT

Noise and Multistability in Gene Regulatory Networks

by

Ertugrul M. Ozbudak

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2004

© Massachusetts Institute of Technology 2004. All rights reserved.

Author . . . .
Department of Physics
March 29, 2004

Certified by . . . .
Alexander van Oudenaarden
Assistant Professor
Thesis Supervisor

Accepted by . . . .
Thomas J. Greytak
Chairman, Department Committee on Graduate Students

Noise and Multistability in Gene Regulatory Networks

by

Ertugrul M. Ozbudak

Submitted to the Department of Physics
on March 29, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Proteins are the functional machinery in living cells. Proteins interact with each other and bind to DNA to form so-called gene regulatory networks and in this way regulate the level, location and timing of expression of other proteins. Cells implement feedback loops to create a memory of their gene expression states. In this way, every differentiated cell in a multicellular organism remembers its expression profile throughout its life. On the other hand, biochemical reactions that take place during gene expression involve small numbers of molecules, and are therefore dominated by large concentration fluctuations. This intrinsic noise has the potential to corrupt memory storage and might result in random transitions between different gene expression states. In the first part of my thesis, I will discuss how the fluctuations in gene expression levels are regulated. The results provided the first experimental evidence that cells can regulate noise in their gene expression by tuning their genetic parameters. In the second half of my thesis, I will discuss how cells create memory by experimentally studying a gene regulatory network that implements a positive feedback loop. A positive feedback loop with nonlinear interactions creates two distinct stable gene expression states. A phase diagram, coupled with a mathematical model of the network, was used to quantitatively investigate the biochemical processes in this network. The response of the network depends on its previous history (hysteresis). Despite the fluctuations in the gene expression, the memory of the gene expression state is preserved for a long time for a broad range of system parameters. On the other hand, for some of the parameters, noise causes random transitions of the cells between different gene expression states and results in a bimodal response. Finally, the hysteretic response of the natural system is experimentally converted to an ultrasensitive graded response as predicted by our model.

Thesis Supervisor: Alexander van Oudenaarden
Title: Assistant Professor

Acknowledgments

I would like to thank Juan Pedraza and Jeff Chabot for giving me inputs at every stage of my studies, to Mukund Thattai working together with me for four years, to Han Lim for working together with me on the lactose multistability project and his high quality of humor, to Attila Becskei for training me in yeast molecular biology techniques, supporting me psychologically during my hard times and showing me first hand the meaning of the sufferings of a scientist, to Fatih M. Yanik for being an example of an impressive hard-worker and his funny jokes, to Cagri A. Savran for having together many social activities together and helping me to use LateX in my thesis, to my advisor Alexander van Oudenaarden for his deep support and patience throughout my graduate study and to David Litster and Mehran Kardar for accepting to be members of my thesis committee and for their helpful suggestions. Also to my parents for encouraging me throughout my studies, and especially to my wife for all her support and making life joyful for me. This research was supported in part by Merck/MIT computational biology fellowship, a grant from NIH on “Multistability And Noise In Gene Regulation” and a grant from DARPA on “Stochastic Fluctuations In Gene Regulation”.

Contents

1 Introduction **25**

1.1 From genes to proteins: Information passaging in living matter 25

1.1.1 Central Dogma of Molecular Biology 25

1.1.2 Transcription 26

1.1.3 Translation 28

1.1.4 Protein bursts 29

1.2 Regulation of gene expression can be accomplished at different steps . 31

1.3 Genetic switches 32

1.4 Gene regulatory networks 34

1.5 Operating principles in gene regulatory networks 35

1.6 Noise in gene expression 38

1.7 Positive feedback mediated memory creation in cells 38

2 Regulation of Noise in the Expression of a Single Gene **41**

2.1 Variations in protein synthesis levels 41

2.2 Exploiting noise in genetic switches 41

2.3 How do cells achieve predictable outcomes? 43

2.4 Noise in the expression of a single gene 45

2.5 Modeling noise in the expression levels of a single gene 52

2.6 Gene intrinsic and extrinsic sources of the noise 57

2.7 Noise in the gene expression of eukaryotic cells 58

2.8 The balance between noise and cost reduction in gene expression . . . 59

3 Multistability in the Lactose Utilization Network of *Escherichia coli* 61
3.1 Multistability . . . 61
3.2 The lac operon . . . 62
3.3 Bistable response of the lac operon . . . 63
3.4 Global analysis of the lactose transport network . . . 67
3.5 Theoretical phase diagram and calculation of parameters . . . 70
3.6 Transitions in the phase diagram . . . 73
3.7 Correlation between green and red fluorescence values . . . 78
3.8 Population averaged measurements and mean fluorescences . . . 80
3.9 Conclusion . . . 82

4 Conclusion and Future Directions 83

A General Cloning Tools 85
A.1 PCR . . . 85
A.2 Insertion of DNA segments into plasmid vectors . . . 87
A.3 Transformation of plasmids into bacteria . . . 88
A.4 Gene insertion into the chromosome of Bacillus subtilis . . . 89
A.5 Gene insertion into the chromosome of E. coli . . . 89

B Methods 1 93
B.1 Strains, growth conditions and media . . . 93
B.2 Data acquisition and analysis . . . 94
B.3 Determination of transcriptional and translational efficiencies . . . 94
B.4 Monte Carlo simulations . . . 95
B.5 Software . . . 95

C Methods 2 97
C.1 Bacterial strains and plasmids . . . 97
C.2 Growth conditions and media . . . 98
C.3 Data acquisition . . . 98
C.4 Data analysis . . . 98

C.5 Calculation of the repression factor 99
C.6 Growth in IPTG and lactose 100

List of Figures

1-1 "**Start and stop signals for RNA synthesis by a bacterial RNA polymerase**. Here, the lower strand of DNA is the template strand. (A) Start signal. The polymerase begins transcribing at the start site. Two short sequences (shaded red), about -35 and -10 nucleotides upstream the start site, determine where the polymerase binds; (B) A stop (termination) signal. The *E. coli* RNA polymerase stops when it synthesizes a run of U residues (shaded blue) from a complementary run of A residues on the template strand, provided that it has just synthesized a self-complementary RNA nucleotide sequence (shaded green), which rapidly forms a hairpin helix that is crucial for stopping transcription. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc." 27

1-2 "**Information flow in protein synthesis.** (A) The nucleotides in an mRNA molecule are joined together to form a complementary copy of a segment of one strand of DNA. (B) They are then matched three at a time to complementary sets of three nucleotides in the anticodon regions of tRNA molecules. At the other end of each type of tRNA molecule, a specific amino acid is held in a high-energy linkage, and when matching occurs, this amino acid is added to the end of the growing polypeptide chain. Thus translation of the mRNA nucleotide sequence into an amino acid sequence depends on complementary base-pairing between codons in the mRNA and corresponding tRNA anticodons. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc." 28

1-3 "**A polyribosome.** Schematic drawing showing how a series of ribosomes can simultaneously translate the same mRNA molecule. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc." . . 30

1-4 "**Summary of the mechanisms by which specific gene regulatory proteins control gene transcription in procaryotes.** (A) Negative regulation; (B) positive regulation. Note that the addition of an inducing ligand can turn on a gene either by removing a gene repressor protein from the DNA (upper left panel) or by causing a gene activator protein to bind (lower right panel). Likewise, the addition of an inhibitory ligand can turn off a gene either by removing a gene activator protein from the DNA (upper right panel) or by causing a gene repressor protein to bind (lower left panel). Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc." 33

1-5 "Large-scale protein interaction networks. Each dot corresponds to a protein and each arrow points to an interaction. (Image courtesy of D. Figeys [1])" 37

2-1 ”Histogram showing the result of a typical experiment in which the expression level of a fluorescent reporter protein is measured in a population of isogenic bacterial cells. Traditional population-averaged measurements would summarize the entire histogram by its mean value $\overline{p}$; however, our single-cell measurements show that the expression level varies from cell to cell, with a standard deviation σ_P. The phenotypic noise strength, defined as the quantity $\sigma_P^2/\overline{p}$, is a measure of the spread of expression levels in a population. The relative standard deviation $\sigma_P/\overline{p}$, although a more common measure of phenotypic noise, obscures its essential behavior. For instance, the relative standard deviation for a Poisson distribution is $\sigma_P/\overline{p} = 1/\overline{p}^{1/2}$, which decreases as the mean increases; but the noise strength for this distribution, $\sigma_P^2/\overline{p} = 1$, is independent of the mean. In general, the noise strength circumvents the trivial effect of decreased noise with increased mean, and measures deviations from Poisson behavior.” 47

2-2 ”Phenotypic noise strength for the four different translational mutants at fixed inducer concentration. Noise strength is clearly dependent on translational efficiency.” 48

2-3 ”Phenotypic noise strength for one strain (ERT3) as inducer concentration is varied. The transcriptional efficiency does not significantly affect noise strength.” 49

2-4 ”Complete experimental data. Each data point is the summarized result of an entire histogram corresponding to a flow cytometer run of a population of typically $10^4 - 10^5$ cells. The phenotypic noise strength of the population (z, in arbitrary fluorescence units) is plotted as a function of transcriptional efficiency (x, depending on the IPTG concentration) and translational efficiency (y, depending on the translational mutant used). Transcriptional and translational efficiencies are normalized to those of the wildtype ERT25 strain, allowing these parameters to be directly compared. These data are fitted to a plane of the form $z = a_0 + a_x x + a_y y$ using a least-square routine, giving $a_0 = 7.10.9$, $a_x = 6.50.4$, $a_y = 21.80.9$. The ratio $a_y/a_x = 3.4$ gives the relative effect of translational versus transcriptional efficiency on phenotypic noise strength.” 50

2-5 ”For clarity, the three-dimensional data are projected parallel to the fit plane onto the boundary planes $x = 1$, noise strength as a function of translation. The intersection of the fit plane with each boundary plane is shown as a solid line; dotted lines indicate an interval of 1 s.d. Data are summarized separately for each translational mutant (dark circles with error bars that represent 95% c.i.).” 51

2-6 ”The three-dimensional data are projected parallel to the fit plane onto the boundary planes $y = 1$, noise strength as a function of transcription. The intersection of the fit plane with each boundary plane is shown as a solid line; dotted lines indicate an interval of 1 s.d.” . . . 52

2-7 'The noise strength as a function of transcription. The results of the control experiments conducted on transcriptional mutants at full induction. Three strains (ERT51, ERT53 and ERT55) are very similar, both in transcriptional efficiency and in noise strength, suggesting that biochemical noise is determined by the actual transcription rate rather than by the specific method used to achieve it. The strain ERT57 shows a highly amplified transcriptional efficiency, allowing reliable estimation of correlations. Data are summarized separately for each transcriptional mutant. A linear fit through these points gives a slope $a_x = 7.30 \pm 0.3$, which is consistent with the slope $a_x = 6.5 \pm 0.4$ obtained from Figure 2-4" 53

2-8 "Modeling single-gene expression. mRNA molecules are transcribed at rate k_R from the template DNA strand. Proteins are translated at a rate k_P from each mRNA molecule. Proteins and mRNA degrade at rates γ_P and γ_R, respectively. Degradation into constituents is denoted by a slashed circle." 53

2-9 ”Upper panel:Typically, mRNA is unstable when compared with the protein product of a gene. During its brief lifetime, however, an mRNA molecule can inject a large burst of proteins into the cytoplasm. A Monte Carlo timecourse over a 30 min time interval shows bursts of protein creation of average size $b = k_P/\gamma_R$ occurring at average rate k_R. The magnitudes of these parameters are indicated on the figure by bars. The timecourse in upper panel is a magnified section of the middle panel. Middle and lower panels: Monte Carlo simulations of typical timecourses for protein number. Deterministic timecourses are indicated as solid lines; the corresponding population histogram is shown to the right of each timecourse. The following examples both achieve the same mean protein concentration, but with different noise characteristics. In both cases, $\gamma_R = 0.1s^{-1}$ and $\gamma_P = 0.002s^{-1}$; the burst size b is varied to obtain different noise strengths, whereas the transcript initiation rate k_R is chosen to fix the mean protein number at 50. A gene with low transcription but high translation rates (middle panel; $k_R = 0.01s^{-1}$, $b = 10$) produces bursts that are large, variable and infrequent, resulting in strong fluctuations. Conversely, a gene with high transcription and low translation rates (lower panel, $k_R = 0.1s^{-1}$, b=1) produces bursts that are small and frequent, causing only weak fluctuations in protein concentration and producing a smaller phenotypic variation in the population. Regulation of a two-step process, that of transcription followed by translation, can therefore be used to independently adjust the mean protein concentration and the level of phenotypic noise in a bacterial population.” 56

2-10 ”Total noise is plotted. The arrow shows the basal level of the total noise, which is defined as the extrinsic noise. Experiments are carried with strain ERT3 as [IPTG] varied.” 58

3-1 ”The lactose utilization network. Red lines represent regulatory interactions, with pointed ends for activation and blunt ends for inhibition; black arrows represent protein creation through transcription and translation, and dotted arrows represent uptake across the cell membrane. In our experiments we vary two external inputs, the extracellular concentrations of glucose and TMG, and measure the resulting levels of two fluorescent reporter proteins: GFP, expressed at the lac promoter, and HcRed, expressed at the gat promoter. LacY catalyses the uptake of TMG, which induces further expression of LacY, resulting in a positive feedback loop.” 64

3-2 ”Overlayed green fluorescence and inverted phase-contrast images of cells that are initially uninduced for lac expression, then grown for 20 h in 18 μM TMG. The cell population shows a bimodal distribution of lac expression levels, with induced cells having over one hundred times the green fluorescence of uninduced cells. Scale bar, 2 μm.” 65

3-3 ”Behavior of a series of cell populations, each initially uninduced (lower panel) or fully induced (upper panel) for lac expression, then grown in media containing various amounts of TMG. Scatter plots show log(green fluorescence) versus log(red fluorescence) for about 1,000 cells in each population. Each scatter plot is centered at a position that indicates the underlying TMG concentration. The scale bar represents variation in red fluorescence by a factor of 10. White arrows indicate the initial states of the cell populations in each panel. The TMG concentration must increase above 30 μM to turn on initially uninduced cells (up arrow), whereas it must decrease below 3 μM to turn off initially induced cells (down arrow). The grey region shows the range of TMG concentrations over which the system is hysteretic.” 66

3-4 ”The mean red fluorescence level of each cell population is independent of its history but decreases with increasing glucose concentrations.” . 67

3-5 ”The phase diagram of the wild-type lactose utilization network. When glucose is added to the medium, the hysteretic region moves to higher levels of TMG. At each glucose level, the lower (down arrow) and upper (up arrow) switching thresholds show those concentrations of TMG at which less than 5% of the cells are in their initial states.” 68

3-6 ”Model parameters are extracted by fitting to measured fluorescence values at the switching thresholds. Upper panel: The maximal promoter activity, α, increases linearly with red fluorescence. Lower panel: The operon repression factor, ρ, is constant.” 74

3-7 ”The TMG uptake rate per active permease, β_T, increases with extracellular TMG levels. The dashed line shows a power-law fit with an exponent of 0.6. The data can also be fitted using a hyperbola, giving a half-saturation concentration of 680 μM.” 75

3-8 ”The elements of catabolite repression. At each glucose concentration, we show the transport activity of LacY molecules (β_G, measuring inducer exclusion) versus the transcriptional activity of the lac operon (α, measuring CRP activation). We see that permease activity drops rapidly as glucose is added, falling to 65% of its maximal value. Operon activity drops more gradually, but falls to 20% of its maximal value. Error bars were determined by propagating the experimental error in measured fluorescence values.” 76

3-9 ”Histograms of log(fluorescence) for cells that are initially uninduced, then grown in media containing 1 mM glucose and various levels of TMG (indicated in μM on each panel). a, Response of the wild-type network. b, c, Response of the network with extraneous LacI-binding sites on a 4-copy plasmid (b) and a 25-copy plasmid (c). d, Theoretical phase diagram of the lactose utilization network.” 77

3-10 ”The mean green fluorescence level of each cell population in Figure 3-9c is shown as a function of the TMG concentration. The response is highly sigmoidal (Hill coefficient ≈ 6) owing to positive feedback.” 78

3-11 ”Fluorescence levels of single cells in a bimodal population. Cells in the induced state show strongly correlated green and red fluorescence levels.” 79

3-12 ”Population averaged lac expression levels as a function of glucose and TMG concentrations. These results are obtained by incorporating our fitted parameters into a stochastic version of the positive feedback model. Axes are oriented so that cells are uninduced at the bottom left corner.” 81

A-1 ”PCR amplification. PCR produces an amount of DNA that doubles in each cycle of DNA synthesis and includes a uniquely sized DNA species. Three steps constitute each cycle, as described in the text. After many cycles of reaction, the population of DNA molecules becomes dominated by a single DNA fragment, X nucleotides long, provided that the original DNA sample contains the DNA sequence that was anticipated when the two oligonucleotides were designed. In the example illustrated, three cycles of reaction produce 16 DNA chains, 8 of which have this unique length (yellow); but after three more cycles, 240 of the 256 DNA chains would be X nucleotides long. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc.” 86

A-2 "The formation of a recombinant DNA molecule. The cohesive ends produced by many kinds of restriction nucleases allow two DNA fragments to join by complementary base-pairing. DNA fragments joined in this way can be covalently linked in a highly efficient reaction catalyzed by the enzyme DNA ligase. In this example a recombinant plasmid DNA molecule containing a chromosomal DNA insert is formed. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc." 88

A-3 "Purification and amplification of a specific DNA sequence by DNA cloning in a bacterium. Each bacterial cell carrying a recombinant plasmid develops into a colony of identical cells, visible as a spot on the nutrient agar. By inoculating a single colony of interest into a liquid culture, one can obtain a large number of identical plasmid DNA molecules, each containing the same DNA insert. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc." 90

List of Tables

2.1 Noise induced switches 43
2.2 Point mutations in RBS and initiation codon 46
2.3 Point mutations in the P_{spac} promoter 46

Chapter 1

Introduction

1.1 From genes to proteins: Information passaging in living matter

Over ten million of species are living on this planet. They are either single-celled organisms: like bacteria and yeast, or multicellular organisms: such as nematodes, insects or mammals, some containing more than 10^{13} cells. Each single cell represents a chemical factory, receiving raw materials from outside and using these materials to grow and multiply. All the information for the functioning of these little factories are encoded in the DNA of the cells. Thus, single cells can be perceived as the vehicles of hereditary information for all living matter. Hereditary information of all genes is stored as a linear chemical code in the DNA molecule of a single cell. Each gene corresponds to the genetic information of one protein. During cell replication, DNA copies itself to pass this information accurately to the next offspring of the mother cell.

1.1.1 Central Dogma of Molecular Biology

During the life cycle of a cell, this conserved genetic information has to be converted into the functional machinery of each cell. Proteins are the main functional molecules in the cells and make up most of the cell mass. Therefore protein synthesis should be very well organized. Protein synthesis consists of two sequential templated polymerization processes, which are called transcription and translation. This two step process of information transfer constitutes the so-called “central dogma of molecular biology”.

1.1.2 Transcription

There are different types of RNA molecules in each cell: messenger RNAs (mRNA), transfer RNAs (tRNA), ribosomal RNAs (rRNA) and some other type of RNA molecules which have catalytic functions. All of these RNA molecules are synthesized by an enzyme called “RNA polymerase” by using one of the strands of DNA as a template. This step is called transcription. During the transcription process, an RNA transcript is created as a chain of single-stranded nucleic acids, whose length ranges from 50 to 10000 nucleotides [2]. Free RNA polymerase molecules diffuse in the cell and attach randomly to the DNA. They slide over the DNA but weakly sticking to most of it. However, when they reach a DNA segment called “promoter”, they bind tightly. The promoter contains a region called ”start site”, where RNA polymerase starts synthesizing a complementary RNA molecule from the DNA template. The RNA polymerase moves stepwise on the DNA and the elongation of the RNA ends when the polymerase reaches a termination signal on the DNA (Figure 1-1).

Each gene or gene cluster (operon) has its own upstream promoter sequences. In bacteria, when all these promoter sequences are compared, a ”consensus sequence” is obtained. Strong promoters (where RNA polymerase produces large amounts of mRNAs from genes downstream of these promoters) have sequences that very closely match to the consensus sequence, whereas weak promoters (where RNA polymerase produces small amounts of mRNAs) have sequences that deviate significantly from the consensus sequence.

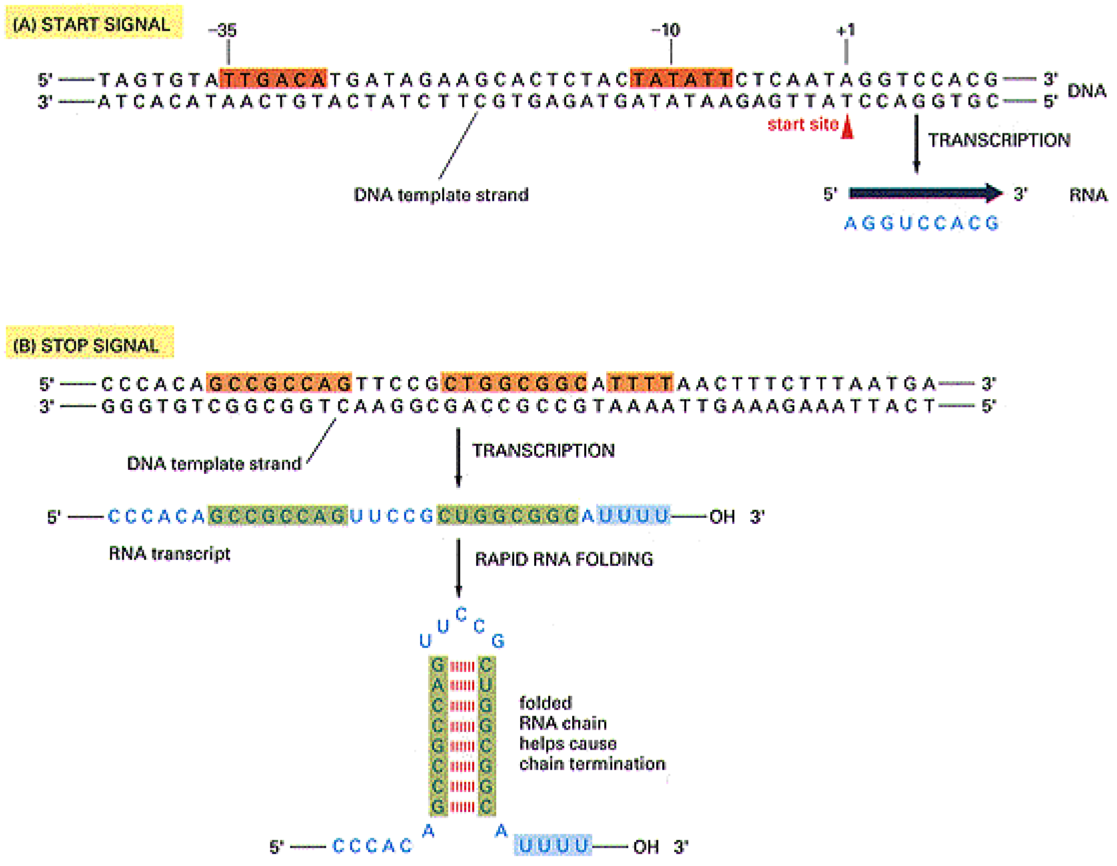
(A) START SIGNAL
−35
−10
+1
5' — T A G T G T A T T G A C A T G A T A G A A G C A C T C T A C T A T A T T C T C A A T A G G T C C A C G — 3'
3' — A T C A C A T A A C T G T A C T A T C T T C G T G A G A T G A T A T A A G A G T T A T C C A G G T G C — 5'
DNA
start site
TRANSCRIPTION
DNA template strand
5'
3'
RNA
A G G U C C A C G
(B) STOP SIGNAL
5' — C C C A C A G C C G C C A G T T C C G C T G G C G G C A T T T T A A C T T T C T T T A A T G A — 3'
3' — G G G T G T C G G C G G T C A A G G C G A C C G C C G T A A A A T T G A A A G A A A T T A C T — 5'
DNA template strand
TRANSCRIPTION
5' — C C C A C A G C C G C C A G U U C C G C U G G C G G C A U U U U — OH 3'
RNA transcript
RAPID RNA FOLDING
U U C C G
G A C C G C C G
C U G G C G G C
folded
RNA chain
helps cause
chain termination
5' — C C C A C
A
A
U U U U — OH 3'

Figure 1-1: ”**Start and stop signals for RNA synthesis by a bacterial RNA polymerase**. Here, the lower strand of DNA is the template strand. (A) Start signal. The polymerase begins transcribing at the start site. Two short sequences (shaded red), about -35 and -10 nucleotides upstream the start site, determine where the polymerase binds; (B) A stop (termination) signal. The *E. coli* RNA polymerase stops when it synthesizes a run of U residues (shaded blue) from a complementary run of A residues on the template strand, provided that it has just synthesized a self-complementary RNA nucleotide sequence (shaded green), which rapidly forms a hairpin helix that is crucial for stopping transcription. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc.”

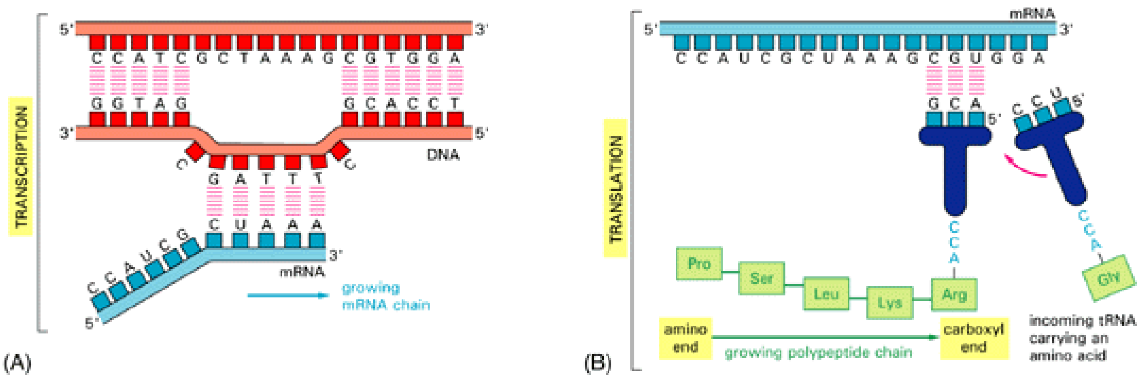
TRANSCRIPTION
5'
3'
C C A T C G C T A A A G C G T G G A
G G T A G
G C A C C T
3'
5'
DNA
G A T T T
C U A A A
C C A U C G
3'
mRNA
5'
growing mRNA chain
(A)
TRANSLATION
mRNA
5'
3'
C C A U C G C U A A A G C G U G G A
G C A
5'
C C U
5'
C C A
C C A
Pro
Ser
Leu
Lys
Arg
Gly
amino end
growing polypeptide chain
carboxyl end
incoming tRNA carrying an amino acid
(B)

Figure 1-2: ”**Information flow in protein synthesis.** (A) The nucleotides in an mRNA molecule are joined together to form a complementary copy of a segment of one strand of DNA. (B) They are then matched three at a time to complementary sets of three nucleotides in the anticodon regions of tRNA molecules. At the other end of each type of tRNA molecule, a specific amino acid is held in a high-energy linkage, and when matching occurs, this amino acid is added to the end of the growing polypeptide chain. Thus translation of the mRNA nucleotide sequence into an amino acid sequence depends on complementary base-pairing between codons in the mRNA and corresponding tRNA anticodons. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc.”

1.1.3 Translation

At the second step of the central dogma, RNA molecules are used as templates to synthesize protein molecules. This step is called “translation”. Proteins are polymers of amino acids linked together by peptide bonds. During this process another type of RNA, which is called tRNA, functions as an adaptor. tRNA translates nucleotide sequence information into amino acid sequence information (Figure 1-2). tRNAs are short RNA molecules, which bind at one end to a specific codon (composed of three nucleotides) on the mRNA and at their other end to the specific amino acid dictated by that codon. The region on the tRNA that binds to a codon on the mRNA is called an anticodon. Codon-anticodon pairing is required to attach each specific amino acid to a growing protein chain.

Proteins are translated from mRNA templates by a complex machinery called the “ribosome”. Ribosomes are composed of rRNAs and proteins. Ribosomes have two major subunits. The smaller subunit binds to mRNA and tRNAs. It helps the codon-anticodon base pairing and prevents any slipping between mRNA and tRNA. The larger subunit catalyzes peptide bond formation between successive amino acids. Ribosome moves stepwise along the mRNA. As it passes over a codon, a new amino acid is added to the growing chain of amino acids. The complete synthesis of an average sized protein takes about 20 seconds in bacteria[2].

In bacteria, there are specific sequences on mRNA molecules that are recognized and bound by ribosomes. These are called ”ribosome binding sequences” (RBS). On mRNAs, the exact sequences that defines for the start of translation is called the ”start codon”. Each RBS is usually located at a close distance upstream of a start codon. Ribosomes that bind to the RBS recognize this start codon on mRNAs and start synthesizing a new protein. RBS sites might occur in more than one place of a bacterial mRNA. This leads to the synthesis of more than one species of protein from one mRNA transcript. These kinds of mRNAs are called polycistronic transcripts. Protein synthesis ends when a ribosome reaches either one of the three specific codons, which are called stop codons, on a mRNA. At this point, the mature protein dissociates from the ribosome.

1.1.4 Protein bursts

On average the synthesis of a protein takes about half a minute in bacteria. During this time period, many translation initiations might take place on the same mRNA transcript. Usually, a new ribosome jumps onto the start codon of mRNA, immediately after the preceding ribosome clears the RBS as it moves along the mRNA. Therefore a series of ribosomes can simultaneously translate the same mRNA molecule, before that mRNA is degraded, giving rise to bursts of newly produced proteins in the cell. This binding of multiple ribosomes on an mRNA molecule generates a structure called polyribosomes (Figure 1-3).

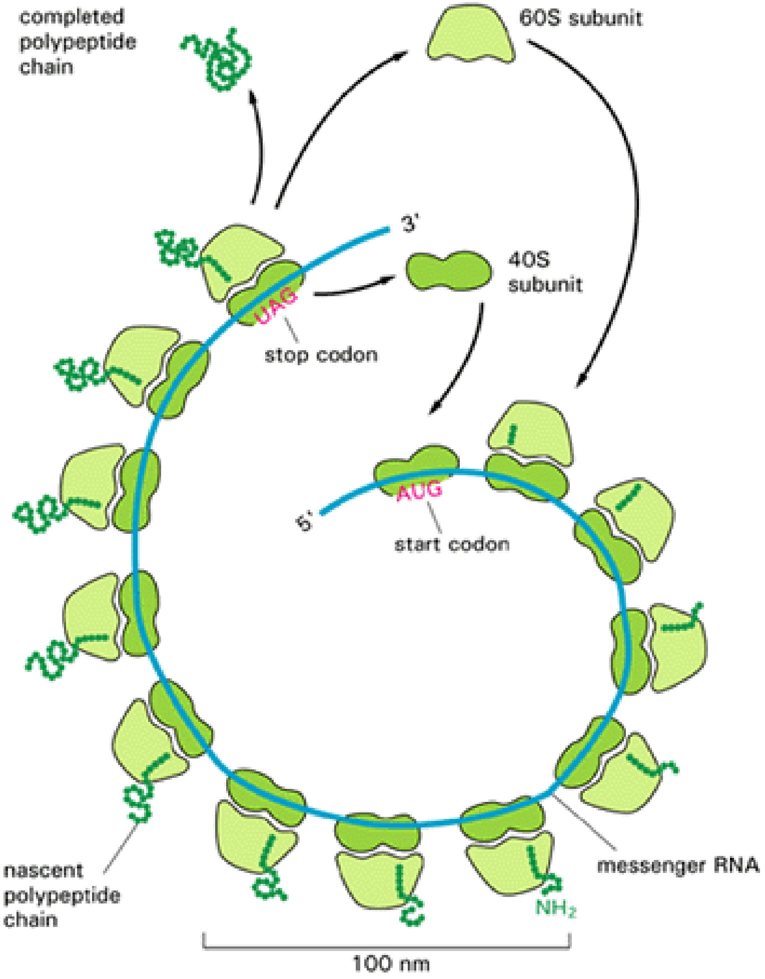
completed polypeptide chain
60S subunit
3'
40S subunit
UAG
stop codon
5'
AUG
start codon
nascent polypeptide chain
messenger RNA
NH_2
100 nm

Figure 1-3: "**A polyribosome.** Schematic drawing showing how a series of ribosomes can simultaneously translate the same mRNA molecule. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc."

1.2 Regulation of gene expression can be accomplished at different steps

Proteins are the functional machinery in the cell factories. They are used as catalysts in most of the reactions happening in any cell. These reactions include replication of DNA molecules and the passage of information from DNA to RNA and then to the proteins themselves, which closes the feedback loop between DNA and proteins. This feedback loop is the core of the self-reproducing capacity of living cells.

While DNA is a single stable molecule in every cell, mRNAs and proteins are unstable molecules. Cells do not synthesize all of the proteins that their genome encodes continuously at high levels. They adjust the rates of transcription of each gene and translation of each mRNA molecule separately to regulate the levels of each of their protein species separately. There are regulatory sequences on the DNA that are called non-coding sequences. They do not code for any protein, but instead define where a gene starts and ends or determine the efficiency of RNA and protein production.

The complexity of the regulatory and non-coding regions of DNA changes from organism to organism. Thus, the genome of an organism not only defines the functions of its proteins but also when and how much of them will be synthesized.

A bacterium can control the amounts of the proteins it makes at different stages: by controlling when and how often a given gene is transcribed (transcriptional control), by selecting which mRNAs in the cytoplasm are translated by ribosomes (translational control), by selectively degradating certain mRNA molecules (degradation control), or by selectively activating or inactivating specific protein molecules after they have been made (activity control).

In majority of the cases, transcriptional controls are the most widely used way of regulating gene expression. Since transcription is the first stage in gene expression, this is the most economic way of achieving regulation. It prevents wasting energy for the production of unused superfluous intermediate molecules. Most of the mRNAs in bacteria have very short lifetime. They are usually degraded within five minutes.

Because of this rapid synthesis and degradation of mRNAs, a bacterium can quickly adapt to any changes in the environment.

Sometimes RNA molecules regulate the translation of other RNA molecules by binding to them and targeting them to degradation. This type of control is called antisense RNA strategy. This strategy is implemented in regulation of the copy number of one of the well-known plasmid family. This strategy creates a feedback control on the initiation of DNA replication for a large family of bacterial DNA plasmids. The control system limits the number of copies of the plasmid made in the cell, thereby preventing the plasmid from killing its host cell by over replicating [2].

1.3 Genetic switches

Transcriptional control of gene expression is mostly done by DNA binding proteins. They bind to specific recognition sequences on the DNA to turn transcription of a gene (or a set of genes) on or off. Some of these regulatory proteins, which are called "repressors", bind to a region close to the promoter of a gene that they regulate. In this case, they inhibit the binding of RNA polymerase to the promoter region of that gene. Some of the other regulatory proteins use another strategy. They bind to the DNA and induce a striking bend in the DNA. This bending sometimes blocks the access of RNA polymerase to the promoter region [3], whereas in other cases it helps RNA polymerase for binding the promoter [4].

Some of the bacterial promoters are only weakly functional on their own, either because they are recognized poorly by RNA polymerase or because the polymerase has difficulty opening the DNA helix when it tries to start transcription. In some cases, these poorly functioning promoters can be activated by gene regulatory proteins that bind to a nearby site to the promoter, contacting the RNA polymerase in a way that dramatically increases the probability of transcription. In many cases, the binding of the repressor or activator proteins are also regulated by secondary small molecules (Figure 1-4).

Transcriptional switches are widely used in bacteria to adapt to changes in their

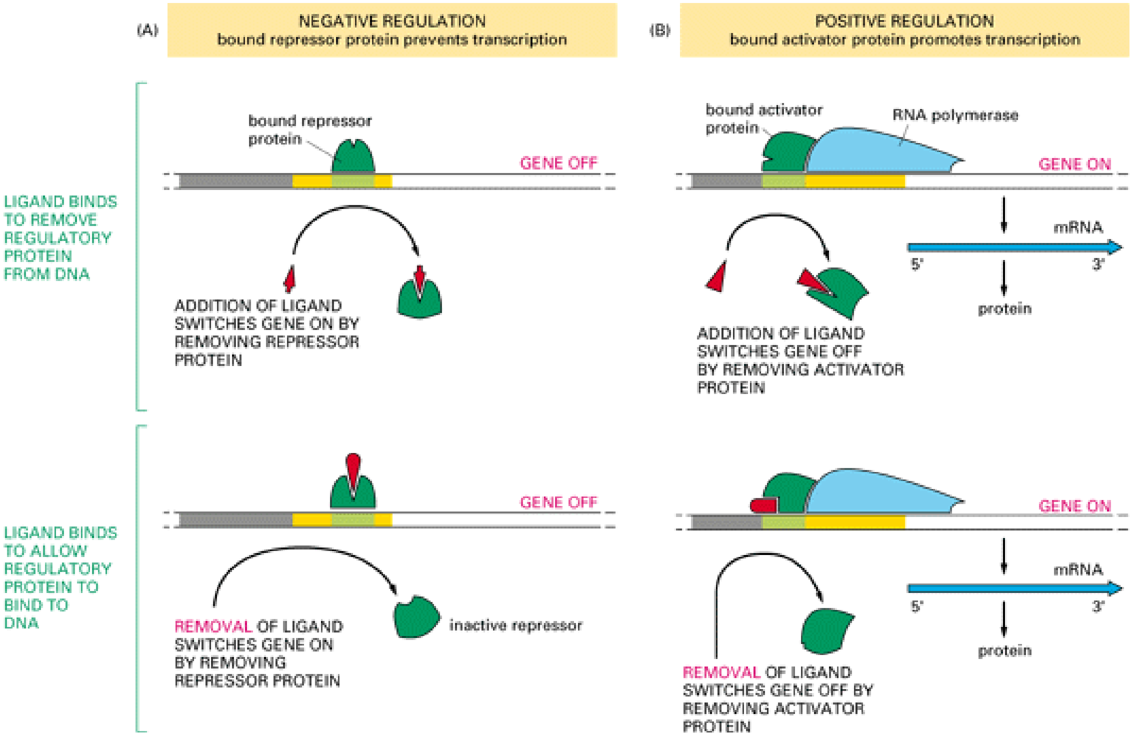
(A)
NEGATIVE REGULATION
bound repressor protein prevents transcription
(B)
POSITIVE REGULATION
bound activator protein promotes transcription
LIGAND BINDS TO REMOVE REGULATORY PROTEIN FROM DNA
bound repressor protein
GENE OFF
ADDITION OF LIGAND SWITCHES GENE ON BY REMOVING REPRESSOR PROTEIN
bound activator protein
RNA polymerase
GENE ON
mRNA
5'
3'
protein
ADDITION OF LIGAND SWITCHES GENE OFF BY REMOVING ACTIVATOR PROTEIN
LIGAND BINDS TO ALLOW REGULATORY PROTEIN TO BIND TO DNA
GENE OFF
REMOVAL OF LIGAND SWITCHES GENE ON BY REMOVING REPRESSOR PROTEIN
inactive repressor
GENE ON
mRNA
5'
3'
protein
REMOVAL OF LIGAND SWITCHES GENE OFF BY REMOVING ACTIVATOR PROTEIN

Figure 1-4: "**Summary of the mechanisms by which specific gene regulatory proteins control gene transcription in procaryotes.** (A) Negative regulation; (B) positive regulation. Note that the addition of an inducing ligand can turn on a gene either by removing a gene repressor protein from the DNA (upper left panel) or by causing a gene activator protein to bind (lower right panel). Likewise, the addition of an inhibitory ligand can turn off a gene either by removing a gene activator protein from the DNA (upper right panel) or by causing a gene repressor protein to bind (lower left panel). Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc."

environment. The bacterial CAP (catabolite activator protein), for example, activates genes that enable *E. coli* to use other carbon sources when glucose, its preferred carbon source, is limited or unavailable. Decreasing levels of glucose induce an increase in the intracellular signaling molecule cyclic AMP (cAMP), which binds to the CAP protein. The cAMP-CAP complex binds to specific recognition sequences on the DNA near the target promoters and turns on the appropriate genes. In this way the expression of a target gene is switched on or off, depending on whether cyclic AMP levels in the cell are high or low, respectively [2].

Some bacterial proteins (including CAP) can act as either activators or repressors, depending on the exact placement of the DNA sequences they recognize with respect to the promoter: if the binding site for the protein overlaps the promoter, the polymerase cannot bind and the protein acts as a repressor; whereas if the binding site does not block the access of the polymerase, the protein acts as an activator.

The lac operon in *E. coli* has a switch, which is regulated by two different DNA binding proteins: a lac promoter specific transcriptional repressor and a globally acting protein CAP (CRP). This operon encodes for three proteins involved in the uptake and metabolism of the lactose. This operon's expression is induced depending on the availability of inducer galactosides (lactose is one of them) outside the cell. The lac operon integrates different signals and switches on and off depending on the outcome of these signals. This operon's response under different conditions is explained in the third chapter of my thesis in more detail.

1.4 Gene regulatory networks

Complex developmental switches are built from smaller ones. In a well-studied model organism, *Drosophila melanogaster*, it was found that, during early development, the expression of segmentation genes is regulated by a hierarchy of positional signals. The products of the egg-polarity genes provide global positional signals that cause downstream, so-called, gap genes to be expressed in special regions of the embryo, and the products of the gap genes then provide a second level of positional signals that act more locally to regulate finer details of patterning by influencing the expression of the pair-rule genes. In this way, the global gradients organize the creation of a fine-grained pattern through a process of sequential positional controls [2].

In many cases, cells control their gene expression in combinatorial ways. In a recent study, it was shown that, opposing gradients of two *D. melanogaster* transcriptional repressors dictates the positions of several segments by differentially repressing two distinct regulatory regions (enhancers) of one of the pair-rule genes [5]. As mentioned before, expression of critical gene regulatory proteins lead to dramatic changes in a whole set of downstream genes.

1.5 Operating principles in gene regulatory networks

Cells have complex gene regulatory networks. The interaction diagrams resembles electrical circuits in a computer. However, there are important differences between our understanding of how a computer and a cell functions. First, computers are built by humans and we know how each component of a computer connects to other components. This enables us to predict the outcome of any change in the network circuitry in a quantitative way. Whereas, we have just started to explore the genomic interaction networks in living cells in the last decades. The genomes of many model organisms, including man, have been sequenced. Biologists have been working hard to decipher which genes code for which proteins. Recently, there has been a great interest to discover the protein interaction circuits in cells [6, 7, 8]. What is still missing, is an understanding of how the global protein interaction network defines specific functions in a cell. From a given interaction network, we should be able to reach a state of knowledge, where we will be able to predict the outcome of any perturbation in this network. Before understanding how the global protein interaction network of a cell works robustly, first we have to learn how the modular parts of this global network function.

A second significant difference between gene regulatory networks and electrical circuits is the signal to noise ratio during information passage in each network. Electronics are designed to have large signal to noise ratios. Whereas, in cells, there are considerable fluctuations (noise) in gene expression levels. Then, one might ask whether or not this variation has any significant impact on the cells. Is it advantages for the cells? Or is it just an unavoidable statistical fact that cells have to cope with and try to filter out? In the literature, one finds examples of cells exploiting the noise in gene expression to introduce heterogeneity in its population to adapt better and faster to changing environmental conditions [9]. Although noise might be advantageous for cells in some cases, one would expect to see that cells have adopted mechanisms to cope with and survive in the presence of these fluctuations. For example, during development of complex organisms, noise has to be filtered out to achieve precise regulation of the differentiation of an embryo [10]. Is there any method that cells use for controlling the noise levels in their gene expression?

Thirdly, computers are designed to have high memory storage capacity. Whereas, as we discuss above, cells live in the presence of large internal fluctuations, which have the potential to corrupt memory storage. However, during embryonic development, all the cells of an adult organism are derived from the same fertilized egg. They differentiate into different cell-fates that have different gene expression patterns at early stages of development. Cells remember their initial commitments and hold on to their distinct fates throughout the life of the adult organism. How is this memory created in the presence of large fluctuations? Are there simple mechanisms that would allow cells to create memory of gene expression pattern and life-long fates?

To answer above-posted questions, we tried to understand how two of the essential operating principles (noise regulation and memory creation) are implemented in gene regulatory networks.

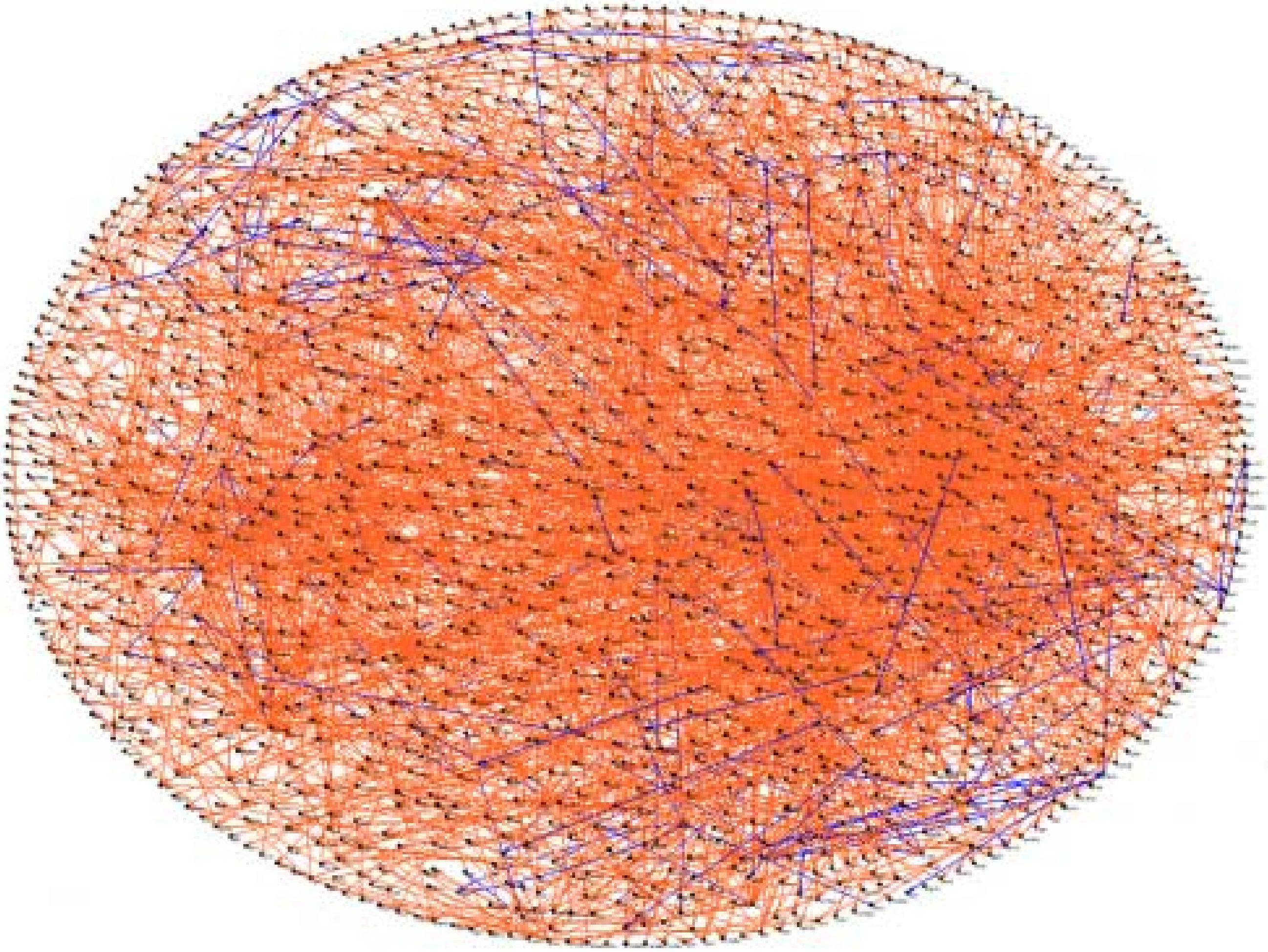

Figure 1-5: "Large-scale protein interaction networks. Each dot corresponds to a protein and each arrow points to an interaction. (Image courtesy of D. Figeys [1])"

1.6 Noise in gene expression

Cells respond to environmental or internal signals by changing the repertoire or the amount of the proteins that they produce. Protein synthesis consumes more energy than any other biosynthesis process in the cells. That's why, every cell has to regulate the changes in its gene expression levels in the most economical way. I made an analogy between a cell and a factory in the beginning of the introduction chapter. But at this point, that analogy fails. In a factory, every machine produces precise quantities of materials within a given time interval, however in a living cell, mRNA and protein concentrations fluctuate significantly. In a population of cells, these fluctuations result in significant differences in the amounts of proteins in each cell at a given time.

To understand the regulation of noise, we took a reductionist approach. The fluctuation in the gene expression is studied at the single gene level. The details of this study are described in the second chapter of my thesis. By changing the sequences of the regulatory regions of a single gene, we showed that noise in the expression levels of any gene is determined by its genetic parameters, such as transcriptional and translational efficiencies. Cells can tune the mean and the noise of each gene's expression independently, which will allow them to achieve any signal to noise ratio. The experimental results are in close agreement with a theoretical prediction published from our group [11].

1.7 Positive feedback mediated memory creation in cells

In computers, memory is stored in a binary state. Cells also achieve storing memory by creating a bistable state. The easiest way to achieve a bistable state is to use feedback loops. Positive feedback loops are one of the frequently recurring network motifs in gene regulatory networks. Positive feedback loops with nonlinear cooperative interactions create two distinct stable states (say, low and high) for the expression of the output protein of a gene network. Bistable networks usually have a history-dependent response (hysteresis). We wanted to understand how cells can store memory by using a single positive feedback loop. As our case-study, we focused on the "lactose transport network" in *E. coli.* This network is composed of an operon that synthesizes essential proteins for the uptake and metabolism of the lactose sugar into bacterial cells. Cells that are expressing the genes from this operon at high levels (high state) take up high levels of lactose from the extracellular environment. Uptake of lactose stimulates the expression of these genes even more. Thus, cells will stay in the high expression state persistently. Whereas cells that are expressing the genes from this operon at low levels (low state) are not able to uptake lactose molecules. Cells that are initially in the low expression state, stay in that state unless they are pushed up by very high levels of the stimulus (extracellular lactose). On the other hand, cells in the high state will keep high gene expression levels unless they are put into an environment which has trace amounts of the lactose molecules. When cells are given intermediate level of stimulus, they stay in their initial states. Thus, the positive feedback network creates a hysteretic bistable response, which allows cells to remember their history (have a memory) until their history is erased by drastic changes in the environment. Hysteretic response is a similar phenomenon that is observed in ferromagnetism and thermodynamics. I discuss this project in more detail in the third chapter of my thesis.

Chapter 2

Regulation of Noise in the Expression of a Single Gene

2.1 Variations in protein synthesis levels

Even in a population of genetically identical cells, grown in uniform conditions, there are significant differences in the levels of gene expression from individual to individual. One of the obvious reasons for this variation comes from the statistical variation in random partitioning of small number of regulatory molecules between daughter cells, when the mother cells divide. Another important part of this variation, the intrinsic genetic noise, is caused by random fluctuations in the underlying biochemical reactions [9, 12, 13]. The concentrations of key molecules that regulate gene expression are in the nanomolar range (1nM = 1 molecule/cell in *E.coli*). This leads to considerable fluctuations in the concentration of each protein species over time in one cell and causes a cell-to-cell variation within a population of clonal cells.

2.2 Exploiting noise in genetic switches

One of the interesting examples for exploiting noise is a gene regulatory network that consists of two proteins, where each protein represses synthesis of the other protein. In this specific case, there would be two stable states in each cell: only one of these two proteins will be synthesized exclusively. Now, if each one of these two proteins regulates the expression of a different set of genes, then in each cell only one distinct set of genes will be expressed. Cells can take advantage of this stochasticity in gene expression and partition their population into two subpopulations with distinct fates. This gene network is the core of the phage-lambda lysis-lysogeny decision circuit [9, 14]. Phage λ exploits this mechanism to achieve diversity in its population and therefore increases the likelihood of the survival of its species under different environmental conditions. Similar behavior is observed in the commitment decision of *Bacillus subtilis* cells with respect to competence and sporulation [15].

Bistable regulatory mechanisms are used in gene networks to produce stochastic phenotypic outcomes. There are many examples of these random bistable switching mechanisms that are used in the control of virulence of pathogenic organisms. Random alterations of surface proteins or random inversions of DNA segments aid pathogens to avoid the host's immune response [12, 16] (Table 2.1). Another example of a stochastic bistable switch is the induction of lac operon in *E.coli* cells. This network implements a cooperative positive feedback to create two separate stable states of gene expression. Stochastic fluctuations in gene expression in each cell drives clonal population of cells into these two separate states. This will be the focus of the third chapter of my thesis.

In a recent study [25], it was shown *Drosophila melanogaster* gene *Dscam*, which is essential for axon guidance, has 38,016 possible alternative protein forms. Different forms of this protein are synthesized from the same initial mRNA script. The difference among them depends on how that mRNA is processed to become a protein. It was found that, the Dscam protein repertoire of each cell is different from those of its neighbors, providing a potential mechanism for generating diversity in the nervous system.

Some other examples of population heterogeneity include: individual swimming behavior and chemotaxis in bacteria [26, 27], differentiation of progenitor haematopoietic stem cells [28], random activation of genes in different T cells [29] and variation of dendrite formation due to haploinsufficiency-mediated increased noise in the ex-

Organism and Reference	Mechanism and Function
Escherichia coli Pap system, [17]	Differential methylation of different Lrp binding sites. Phase variation in pili expression, affecting virulence.
Escherichia coli Fim system, [18]	Invertible DNA segments. Phase variation in type 1 pili, affecting virulence.
Haemophilus influenzae, [19]	lipopolysaccharide epitope expression
Moraxella bovis, [20]	Invertible DNA segments. Phase variation in pilin alters antigen response.
Neisseria gonorrheae, [21]	Pili expression
Phage Mu, [22]	Invertible DNA segments. Phase variation in type 1 pili, affecting virulence.
Salmonella typhimurium Hin system, [22]	Invertible DNA segments. Phase variation in flagellin alters antigen response.
Staphylococcus epidermidis, [23]	polysaccharide intercellular adhesin synthesis
Vibrio vulnificus, [24]	Capsular polysaccharide expression

Table 2.1: Noise induced switches

pression of a tumor-suppressor gene[30].

2.3 How do cells achieve predictable outcomes?

Despite the above mentioned unavoidable stochasticity, in majority of the cases, cells display predictable outcomes. This becomes even more striking in the development of complex multicellular organisms. How are these deterministic outcomes achieved? Up to now, redundancy in genes or in regulatory networks, feedback loops and checkpoint mechanisms are proposed as an answer to this question. Checkpoint mechanisms pause cells to make sure that certain tasks during the cell cycle are accomplished properly. Redundancy in genes and regulatory pathways provides back up against mutational and environmental perturbations.

In nematodes (worms) anchor cells (AC) and ventral uterus cells (VU) differentiate from two precursor cells: Z1 and Z4. In some of the worm species, differentiation of AC/VU cell lineages has a fixed pathway from their precursor cells, however in *C. elegans*, different cell lineage (AC or VU) decisions are made randomly from Z1 and Z4 precursor cells. In about half of the animals, AC is derived from the Z1 precursor cell, whereas in the rest of the embryos it is derived from the Z4 precursor cell. Because of an intercellular feedback loop, the other precursor cell always differentiate into VU fate. Thus, stochastic decision in the beginning of lineage selection is compensated by an intercellular feedback loop to give rise to precise ratio of final differentiated cell types [31].

In a recent study, using a synthetic gene circuit, it was shown that negative feedback reduces the noise in gene expression levels substantially [32]. Negative feedback works like a low-pass filter. On the other hand, integral feedback results in band-pass filter that amplifies intermediate frequencies while attenuating low and high frequencies. This type of feedback is used in bacterial chemotaxis [33]. This feedback enables cells to have a robust adaptation in response to chemotactic stimuli. Circadian rhythms are observed in nearly all living organisms with a characteristic period close to 24h. In a theoretical study [34], robustness of circadian rhythms with respect to molecular noise was investigated. It was found that, robustness increases with the degree of cooperativity of the autorepression circuit that composes the essential core of the circadian network. The entrainment by light/dark cycles was found to stabilize the phase of the oscillations with respect to molecular noise. Whereas, a different theoretical study claims that noise may even be beneficial to the somitogenesis oscillator in vertebrates. It allows the oscillations to continue under conditions, where a purely deterministic reaction model does not allow sustained oscillations [35] and results in spatial patterns of somites in vertebrates.

Another recent study investigated the precision in the establishment of regulatory protein gradients during embryonic development [10]. Although a gradient of an upstream regulatory protein varies strongly from one embryo to another, its positional readout (downstream regulated protein) is still precise and scales with the embryo length. Although the specific mechanism that leads to this noise filtration was not identified, nevertheless the results implied the existence of such mechanisms.

2.4 Noise in the expression of a single gene

As is apparent from the previous discussion, it is important to develop a quantitative understanding of noise in gene expression. Sometimes cells exploit it to achieve diversity in their population response in a widely changing environment. Whereas, in other cases they have to filter it to achieve robust phenotypes. However there has not been any experimental study to identify the sources and the regulation of noise at the expression level of a single gene. Therefore, we investigated whether or not cells can differentially regulate the noise levels in the production of each one of their different protein species. If this is true, then, each gene should have some distinct property to regulate the level of fluctuations in its expression.

To get an experimental answer to this question, we examined how the cell-to-cell variation in the expression level of a single gene depends on its genetic parameters such as transcription and translation rates. There are regulatory non-coding sequences upstream of each gene. We specifically explored how some of the sequences determine the transcriptional and translational efficiencies (Chapter 1) of any gene and how they also affect the noise in the gene expression levels. We selected as our reporter system a single copy chromosomal gene with an inducible promoter. Since an estimated 50-80% of bacterial genes are transcriptionally regulated [36], this system typifies the majority of naturally occurring genes, allowing our results to be extended to natural systems.

A single copy of our reporter, the green fluorescent protein gene (*gfp*), was incorporated in the chromosome of *B. subtilis* (Appendix B.1). We chose to integrate *gfp* in the chromosome itself, rather than in the form of plasmids, since plasmid copy number variation [37, 38] can act as an additional and unwanted source of noise. We also introduced a repressor protein (LacI) into the chromosome under the control of a constitutively active and strong promoter. *gfp* is put under the control of the P_{spac} promoter. This promoter has specific binding sites for LacI repressors. The LacI repressors will bind to their recognition sequences and repress gene expression from this promoter. An inducer molecule, isopropyl-β-D-thiogalactopyranoside (IPTG),

Strain	Ribosome Binding Site	Initiation Codon	Translational Efficiency
ERT25	GGG AAA AGG AGG TGA ACT ACT	ATG	1.00
ERT27	GGG AAA AGG AGG TGA ACT ACT	TTG	0.87
ERT 3	GGG AAA AGG TGG TGA ACT ACT	ATG	0.84
ERT29	GGG AAA AGG AGG TGA ACT ACT	GTG	0.66

Table 2.2: Point mutations in RBS and initiation codon

Strain	-10 Regulatory Region	Transcriptional Efficiency
ERT 57	CAT AAT GTG TGT AAT	6.63
ERT 25	CAT AAT GTG TGG AAT	1.00
ERT 53	CAT AAT GTG TGC AAT	0.79
ERT 51	CAT AAT GTG TGA AAT	0.76
ERT 55	CAT AAT GTG TAA AAT	0.76

Table 2.3: Point mutations in the P_{spac} promoter

binds to LacI and impairs its binding to the promoter.

Varying the concentration of IPTG in the growth medium was used to regulate the transcriptional efficiency of *gfp* expression. Translational efficiency was regulated by constructing a series of *B. subtilis* strains (Table 2.2) that contained point mutations in the ribosome binding site (RBS) and initiation codon of *gfp* [39]. As it was discussed in the introduction, binding affinity of ribosomes depends on their recognition sequences on the mRNA and how they match to their strongly preferred sequences.

Since we used two different type of strategies to regulate transcriptional (by changing the probability of RNA polymerase binding to the promoter region with the use of an intermediate repressor molecule) and translational (by mutating the RBS site, so directly changing the affinity of ribosome to the RBS) processes, this might introduce a potential bias in the relative contributions of these processes to biochemical noise. As a control, we constructed four additional strains (Table 2.3) whose transcription rates were altered by mutations in the promoter region, which changed the binding probability of RNA polymerase to the promoter region directly. As described below, both strategies of transcriptional regulation produce similar results.

The GFP expression levels for single cells in a bacterial population were measured by flow cytometry. Differences in GFP expression from cell to cell (phenotypic noise)

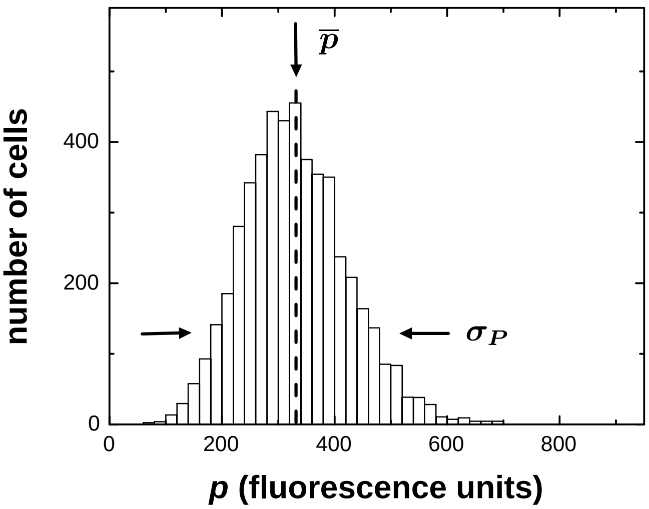

Figure 2-1: "Histogram showing the result of a typical experiment in which the expression level of a fluorescent reporter protein is measured in a population of isogenic bacterial cells. Traditional population-averaged measurements would summarize the entire histogram by its mean value $\overline{p}$; however, our single-cell measurements show that the expression level varies from cell to cell, with a standard deviation σ_P. The phenotypic noise strength, defined as the quantity $\sigma_P^2/\overline{p}$, is a measure of the spread of expression levels in a population. The relative standard deviation $\sigma_P/\overline{p}$, although a more common measure of phenotypic noise, obscures its essential behavior. For instance, the relative standard deviation for a Poisson distribution is $\sigma_P/\overline{p} = 1/\overline{p}^{1/2}$, which decreases as the mean increases; but the noise strength for this distribution, $\sigma_P^2/\overline{p} = 1$, is independent of the mean. In general, the noise strength circumvents the trivial effect of decreased noise with increased mean, and measures deviations from Poisson behavior."

are clearly seen in a histogram showing the protein expression levels (p) measured during a typical experiment Figure 2-1.

The histogram is characterized by a mean value $\overline{p}$ and a standard deviation σ_P. The phenotypic noise strength, defined as the quantity $\sigma_P^2/\overline{p}$ (variance/mean), is sensitive to the microscopic sources of stochasticity (which depends on the sequence properties of each gene) that we wish to study, and is the unit in which we report our results. We measured the phenotypic noise strength for the four different translational mutants as IPTG concentration was varied between 30 μM (near-basal transcription)

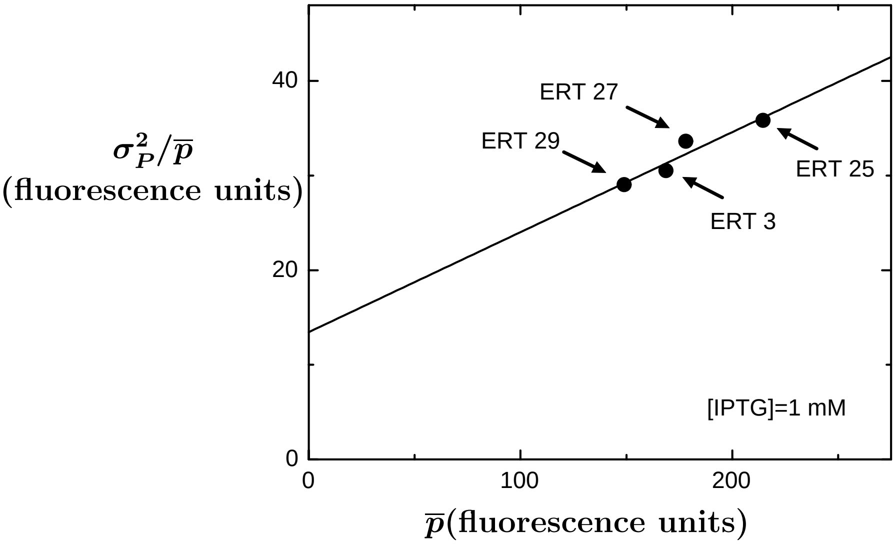

Figure 2-2: "Phenotypic noise strength for the four different translational mutants at fixed inducer concentration. Noise strength is clearly dependent on translational efficiency."

and 1 mM (full operon induction).

For example: Figure 2-2 shows flow cytometer results for these four strains at full induction; Figure 2-3 shows the results from a series of flow cytometer experiments conducted on a single strain (ERT3) as IPTG concentration was varied.

Figure 2-4 summarizes all of our experimental results, showing the measured noise strength as a function of both transcriptional efficiency (varying [IPTG] in the growth medium) and translational efficiency (using different strains with mutations in the RBS and initiation codon). Note that each data point is the result of an entire histogram corresponding to a flow cytometer run of a population of typically 10^4 to 10^5 cells.

Since the addition of IPTG and mutations in the gfp RBS are not expected to affect normal cellular processes, all contributions to phenotypic noise remain unchanged throughout our experiment, other than transcriptional and translational efficiencies. The response of the phenotypic noise strength to a change in either the translational efficiency (the slope of the curve shown in Figure 2-5) or the transcriptional efficiency

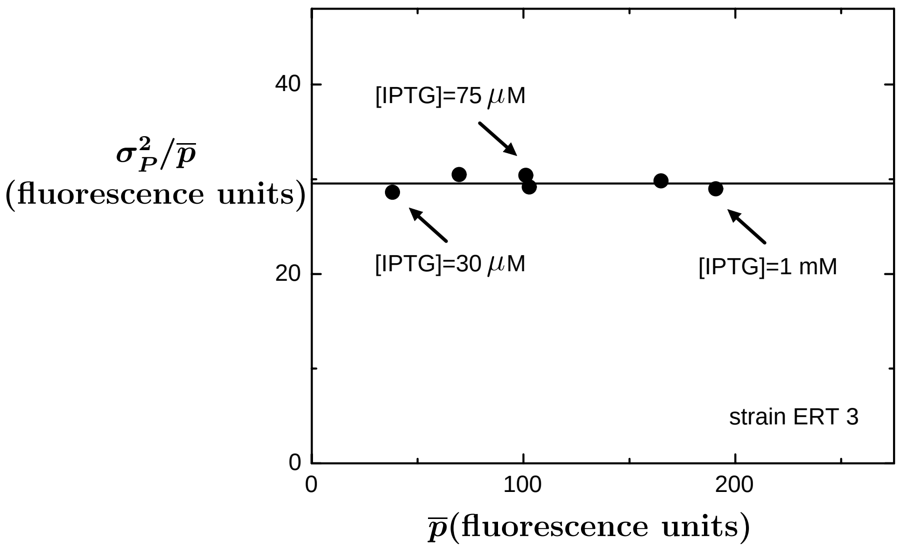

Figure 2-3: "Phenotypic noise strength for one strain (ERT3) as inducer concentration is varied. The transcriptional efficiency does not significantly affect noise strength."

(the slope in Figure 2-6) therefore isolates the contribution of that parameter to the phenotypic noise.

We find that the phenotypic noise strength shows a strong positive correlation with translational efficiency (Figure 2-5 , slope=21.8), compared to only a weak positive correlation with transcriptional efficiency (Figure 2-6 , slope=6.5). Switching from the ERT27 strain to the ERT25 strain (an increase in translational efficiency of about 15%, see Table 2.2) increases the noise strength from 32 to 35 units; the same effect is achieved only upon doubling transcriptional efficiency (a 100% increase) from the half-induction to the full-induction level. Experiments conducted on the control strains, in which transcription rates were altered by mutation rather than by operon induction, corroborated the weak correlation between noise strength and transcriptional efficiency (Figure 2-7, slope=7.3). The differential nature of our measurements makes these results independent of the specific properties of the reporter protein, such as gene locus or folding characteristics.

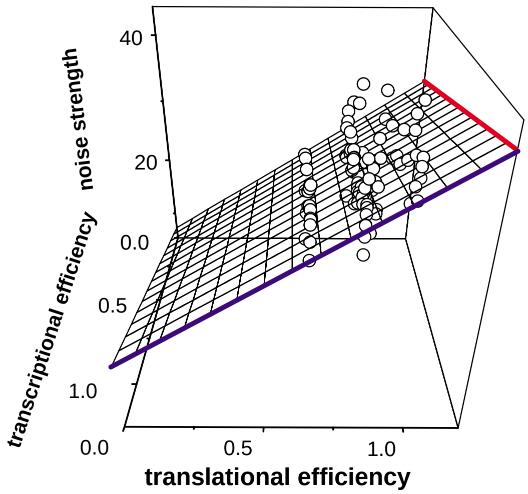

Figure 2-4: ”Complete experimental data. Each data point is the summarized result of an entire histogram corresponding to a flow cytometer run of a population of typically $10^4 - 10^5$ cells. The phenotypic noise strength of the population (z, in arbitrary fluorescence units) is plotted as a function of transcriptional efficiency (x, depending on the IPTG concentration) and translational efficiency (y, depending on the translational mutant used). Transcriptional and translational efficiencies are normalized to those of the wildtype ERT25 strain, allowing these parameters to be directly compared. These data are fitted to a plane of the form $z = a_0 + a_x x + a_y y$ using a least-square routine, giving $a_0 = 7.10.9$, $a_x = 6.50.4$, $a_y = 21.80.9$. The ratio $a_y/a_x = 3.4$ gives the relative effect of translational versus transcriptional efficiency on phenotypic noise strength.”

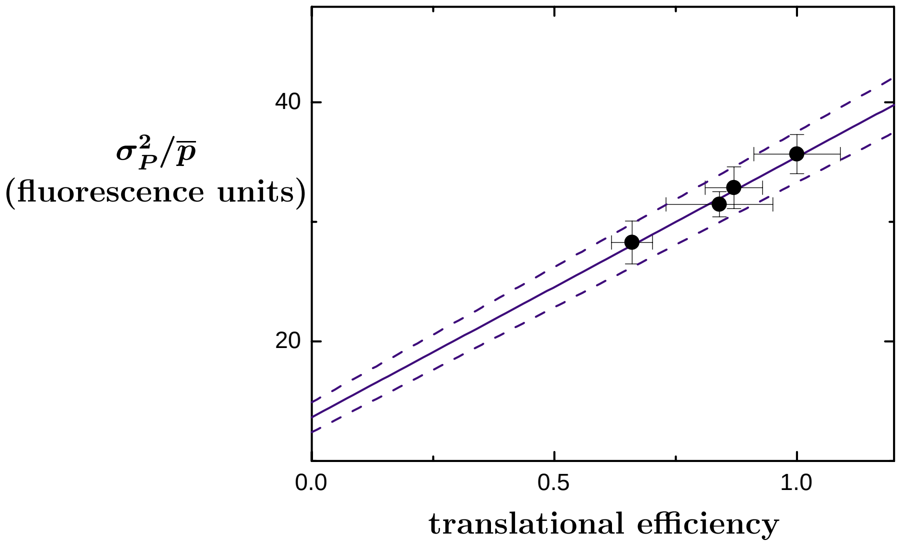

Figure 2-5: "For clarity, the three-dimensional data are projected parallel to the fit plane onto the boundary planes $x = 1$, noise strength as a function of translation. The intersection of the fit plane with each boundary plane is shown as a solid line; dotted lines indicate an interval of 1 s.d. Data are summarized separately for each translational mutant (dark circles with error bars that represent 95% c.i.)."

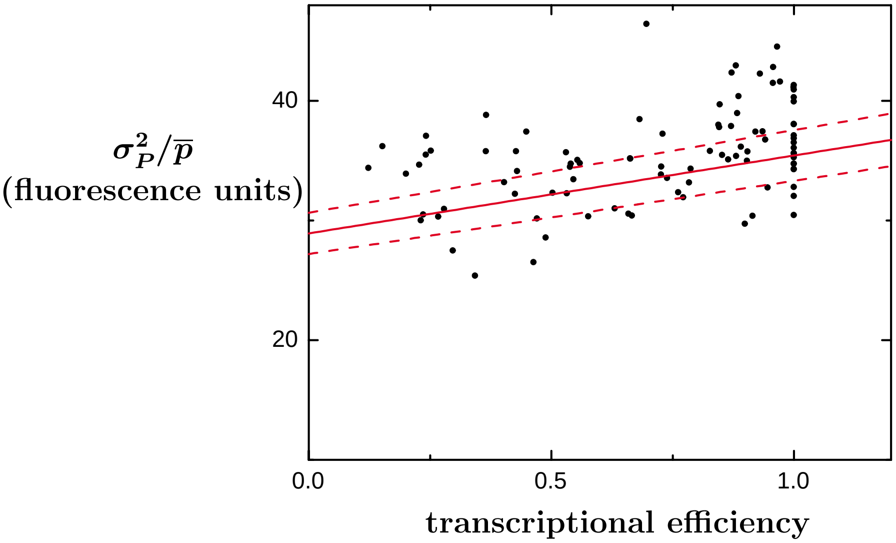

Figure 2-6: "The three-dimensional data are projected parallel to the fit plane onto the boundary planes $y = 1$, noise strength as a function of transcription. The intersection of the fit plane with each boundary plane is shown as a solid line; dotted lines indicate an interval of 1 s.d."

2.5 Modeling noise in the expression levels of a single gene

The noise properties of a single gene can be derived using the Langevin technique. This approach yields statistics equivalent to those generated by large-scale Monte Carlo simulations, but has the added advantage of providing insight into system behavior [40]. We treat the mRNA number r and protein number p as continuous quantities and assume that fluctuations are introduced by gaussian white noise sources:

$$\frac{dr}{dt} + \gamma_R r = k_R + \eta_R \tag{2.1}$$

$$\frac{dp}{dt} + \gamma_P p = k_P r + \eta_P \tag{2.2}$$

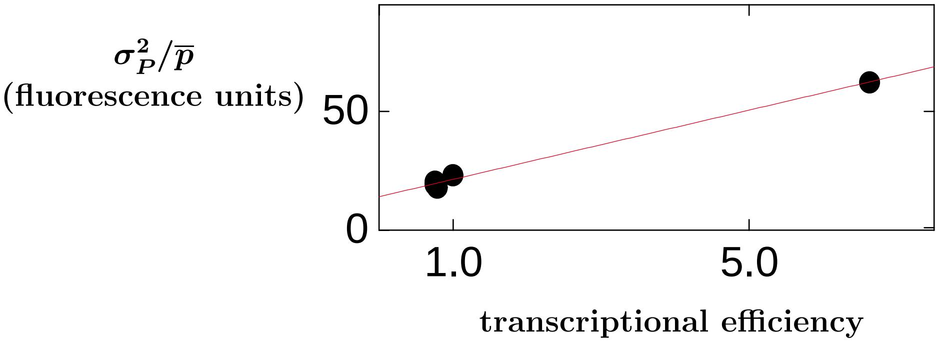

Figure 2-7: 'The noise strength as a function of transcription. The results of the control experiments conducted on transcriptional mutants at full induction. Three strains (ERT51, ERT53 and ERT55) are very similar, both in transcriptional efficiency and in noise strength, suggesting that biochemical noise is determined by the actual transcription rate rather than by the specific method used to achieve it. The strain ERT57 shows a highly amplified transcriptional efficiency, allowing reliable estimation of correlations. Data are summarized separately for each transcriptional mutant. A linear fit through these points gives a slope $a_x = 7.30 \pm 0.3$, which is consistent with the slope $a_x = 6.5 \pm 0.4$ obtained from Figure 2-4"

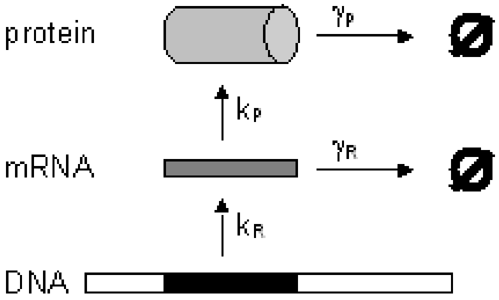

Figure 2-8: "Modeling single-gene expression. mRNA molecules are transcribed at rate k_R from the template DNA strand. Proteins are translated at a rate k_P from each mRNA molecule. Proteins and mRNA degrade at rates γ_P and γ_R, respectively. Degradation into constituents is denoted by a slashed circle."

Here, γ_R and γ_P represent the decay rates of mRNA and protein, respectively; k_R is the transcription rate and k_P is the translation rate, so the rate of protein creation is k_Pr (Figure 2-8). η_R and η_P are white noise sources with the following statistics:

$$\overline{\eta_i(t)} = 0 \tag{2.3}$$

$$\overline{\eta_i(t)\eta_i(t+\tau)} = q_i\delta(\tau) \tag{2.4}$$

where i = R or P, angular braces represent population averages, and δ is the Dirac δ-function. The noise magnitudes q_i are chosen so that they are consistent with the steady-state Poisson statistics of chemical reactions. For example, in steady-state, the mRNA number is given by $\overline{r} = k_R/\gamma_R$. Expanding around this steady-state by setting r = $\overline{r}$ + δr gives:

$$\frac{d\delta r}{dt} + \gamma_R \delta r = \eta_R \tag{2.5}$$

Fourier-transforming these equations by setting $x(t) = \int e^{i\omega t} x(\omega) d\omega/2\pi$ gives:

$$\frac{\delta r(\omega)}{\eta_R(\omega)} = \frac{1}{\gamma_R + i\omega} \tag{2.6}$$

$$\overline{(|\eta_R(\omega)|)^2} = q_R \tag{2.7}$$

so that the steady-state value of the fluctuations is given by:

$$\overline{\delta r^2} = \int \frac{d\omega}{2\pi} \frac{1}{\gamma_R^2 + \omega^2} q_R = \frac{q_R}{2\gamma_R}$$

Now we impose Poisson statistics by setting $\overline{\delta r^2} = \overline{r}$, giving $q_R = 2k_R$, and similarly, $q_p = 2k_P k_R/\gamma_R$. Protein number fluctuations can then be derived as:

$$\overline{\delta p^2} = \int \frac{d\omega}{2\pi} \frac{1}{\gamma_P^2 + \omega^2} (q_P + \frac{q_R}{\gamma_R^2 + \omega^2}) = \overline{p}(1 + \frac{k_P}{\gamma_R + \gamma_P})$$

We define the noise strength to be the quantity $\nu = \overline{\delta p^2}/\overline{p}$, also known as the Fano factor (Figure 2-1). For a Poisson process, $\nu = 1$; for an arbitrary stochastic process, the noise strength reveals deviations from Poissonian behavior. Setting $\phi = \gamma_P/\gamma_R$ and defining the burst size $b = k_P/\gamma_R$ finally gives:

$$\frac{\overline{\delta p^2}}{\overline{p}} \equiv \frac{\sigma_P^2}{\overline{p}} = 1 + \frac{b}{1+\phi} \tag{2.8}$$

Typically, ϕ is a small quantity (mRNA is unstable compared with protein), so that the result above reduces to:

$$\frac{\sigma_P^2}{\overline{p}} \cong 1 + b \tag{2.9}$$

This equation shows that $(\sigma_P^2/\overline{p})$ is greater than Poissonian noise strength ($\sigma_P^2/\overline{p}$ $= 1$) and is simply an increasing function of translational efficiency.

Here, $b = k_P/\gamma_R$ is the average number of proteins synthesized per mRNA transcript. These proteins are injected into the cytoplasm in sharp bursts (upper panel in Figure 2-9). The measured asymmetry between the transcriptional and translational contributions is consistent with this prediction, and is strong evidence for the biochemical origin of phenotypic variability (lower panels in Figure 2-9). Phenotypic noise in a population is therefore indicative of protein concentration fluctuations over time in single cells.

The cell to cell variation in gene expression, and fluctuations over time in single cells, have broad implications. Noise is often thought as being harmful, garbling cell signals, corrupting circadian clocks [41], and disrupting the fine-tuned process of development. Cell signaling pathways [42] and developmental switches [43] have evolved so as to minimize the disruptive effect of such fluctuations in ways which are only now beginning to be understood.

Previously, it was reported that variation in gene expression could be reduced by autoregulation [32]. In our experiments [44], we experimentally demonstrate that phenotypic variation can also be controlled by the genetic parameters of a single gene.

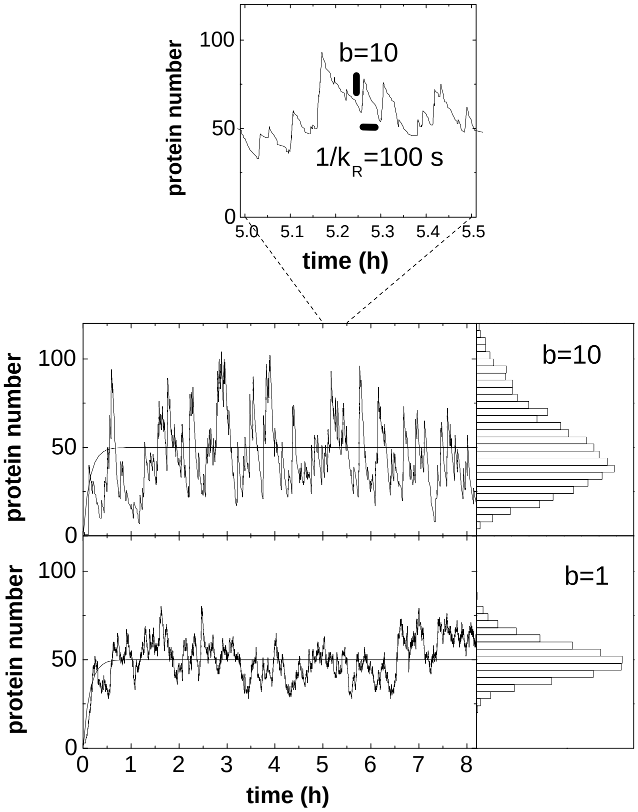

Figure 2-9: "Upper panel:Typically, mRNA is unstable when compared with the protein product of a gene. During its brief lifetime, however, an mRNA molecule can inject a large burst of proteins into the cytoplasm. A Monte Carlo timecourse over a 30 min time interval shows bursts of protein creation of average size $b = k_P/\gamma_R$ occurring at average rate k_R. The magnitudes of these parameters are indicated on the figure by bars. The timecourse in upper panel is a magnified section of the middle panel. Middle and lower panels: Monte Carlo simulations of typical timecourses for protein number. Deterministic timecourses are indicated as solid lines; the corresponding population histogram is shown to the right of each timecourse. The following examples both achieve the same mean protein concentration, but with different noise characteristics. In both cases, $\gamma_R = 0.1s^{-1}$ and $\gamma_P = 0.002s^{-1}$; the burst size b is varied to obtain different noise strengths, whereas the transcript initiation rate k_R is chosen to fix the mean protein number at 50. A gene with low transcription but high translation rates (middle panel; $k_R = 0.01s^{-1}$, $b = 10$) produces bursts that are large, variable and infrequent, resulting in strong fluctuations. Conversely, a gene with high transcription and low translation rates (lower panel, $k_R = 0.1s^{-1}$, b=1) produces bursts that are small and frequent, causing only weak fluctuations in protein concentration and producing a smaller phenotypic variation in the population. Regulation of a two-step process, that of transcription followed by translation, can therefore be used to independently adjust the mean protein concentration and the level of phenotypic noise in a bacterial population."

2.6 Gene intrinsic and extrinsic sources of the noise

Another study differentiated the noise contributions coming from gene intrinsic or extrinsic factors [45]. In this study, inserted cyan and yellow fluorescent protein genes (*cfp* and *yfp*) under the control of the same promoter were inserted into the chromosome of *E.coli* and the expression levels of these two proteins were measured. The degree of correlation in the fluctuations of these two proteins gives information about the extrinsic contribution to the total noise. Extrinsic noise stems from the fluctuations in the common regulatory elements, such as: fluctuations in the amount of repressor proteins, RNA polymerase molecules or ribosomes. Uncorrelated fluctuations in these two fluorescent protein levels measure the gene intrinsic noise. If the intrinsic noise dominates, individual genes could be selected for their noise properties. If the extrinsic part of the noise dominates, the noise levels of all of the genes would be regulated similarly independent of their different transcription and translation rates. The outcome of this study was that the gene extrinsic contribution of the noise is the dominant part of the total noise in gene expression levels. However, a recent analysis [46] showed that, most of the extrinsic noise in the two gene study originates from repressor proteins that were synthesized from plasmids. Variation in plasmid copy number could lead to an increased extrinsic noise.

When we plot our data from one strain at different transcriptional induction levels with different units: $\sigma_P^2/\overline{p}^2$, we found that (Figure 2-10) most of the noise comes from the fluctuations in the intrinsic factors. The extrinsic part only adds up a constant displacement to the total noise. If we divide Equation 2.9 by $\overline{p}$, we obtain:

$$\frac{\sigma_P^2}{\overline{p}^2} \cong \frac{C_1}{\overline{p}} \tag{2.10}$$

where $C_1 = b+1$. The right hand side of this equation corresponds to the intrinsic noise term. Any additional constant terms to the right hand side will correspond to the extrinsic part of the total noise. Therefore the total noise is:

$$\eta_{TOT}^2 = \eta_{int}^2 + \eta_{ext}^2 = \frac{C_1}{\overline{p}} + C_2 \tag{2.11}$$

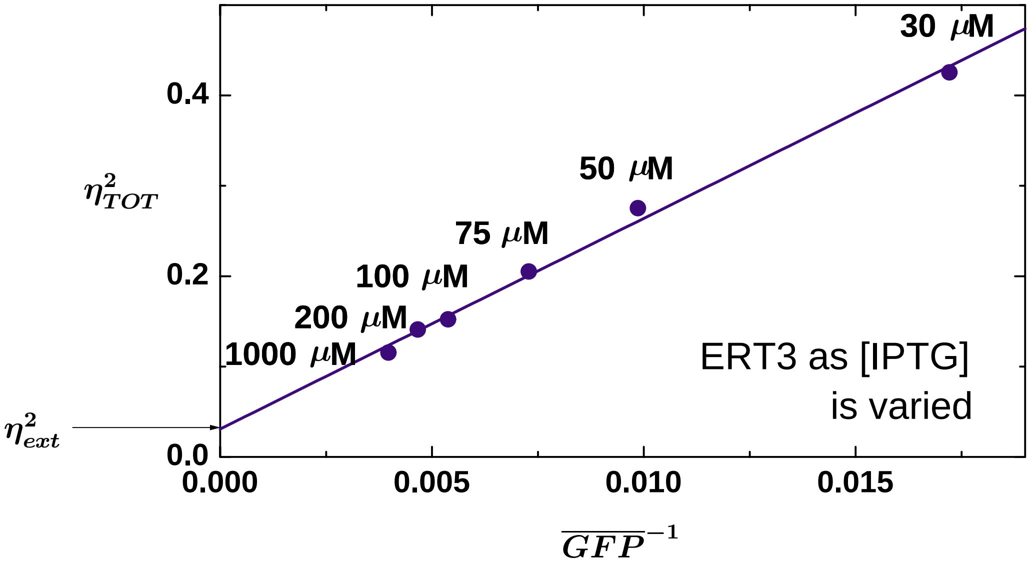

Figure 2-10: "Total noise is plotted. The arrow shows the basal level of the total noise, which is defined as the extrinsic noise. Experiments are carried with strain ERT3 as [IPTG] varied."

where $\eta^2_{ext} = C_2$. Based on this analysis (Figure 2-10) it can be concluded that extrinsic noise is smaller compared to the intrinsic part of the noise.

2.7 Noise in the gene expression of eukaryotic cells

A recent study investigated noise in the gene expression of *Saccharomyces cerevisiae* (budding yeast) cells [47]. Yeast is a single-celled eukaryotic (which has a nucleus) organism. In yeast cells, transcription takes place in the nucleus. Yeast DNA is bound to special proteins that constitute the so-called chromatin. Bacteria do not have a nucleus and both transcription and translation take place in the cytoplasm. Bacteria also do not have a chromatin structure. The motivation of this study was to see, whether or not compartmentalization of transcription and chromatin structure has any effect on the noise levels in eukaryotic cells. They showed a clear difference in the noise levels between bacterial and eukaryotic gene expression. The noise strength has a non-monotonical dependence on the transcriptional efficiency in eukaryotic gene expression. This result is attributed to pulsatile mRNA production due to transcription reinitiation, which is crucial for the dependence of noise on transcriptional efficiency. However, experiments have not been performed yet that directly explore the effects of transcriptional reinitiation on the fluctuations in the protein levels in eukaryotic cells [46].

2.8 The balance between noise and cost reduction in gene expression

It is the random births and deaths of mRNA transcripts that dominantly determine the noise levels in gene expression. The average number of proteins produced per gene is equal to the product of the average number of mRNAs produced per gene and the average number of proteins produced per mRNA. One cell can obtain the same average protein number by reducing the transcriptional efficiency by a certain factor and increasing the translational efficiency by that same factor. Exactly the opposite case is also possible. The same average number of proteins is obtained by increasing the transcriptional efficiency by a factor and decreasing the translational efficiency by the same factor. In the first scenario, protein production would be more economical (Chapter 1.2), since it will require much less transcript production to achieve the same average number of proteins. But this case will result in a noisier protein production within a clonal population of cells. In the second scenario, the protein production is less economical but will also be less noisy. Thus, the reduction in noise will only be achieved by spending more cellular materials to produce the same amount of proteins [13]. This scenario suggests the existence of a balance between noise and energy management in the cell.

The technique of transcriptional and translational noise control can be applied in the fast growing field of artificial genetic networks [48]. The current capabilities of artificially engineered circuits such as genetic toggle switches [49] or ring oscillators

[50] are limited by intrinsic noise. Novel noise reduction methods will allow these circuits to mimic the robust behavior of natural biological systems, and will enable their practical application in areas such as biocomputation, or in the construction of genetic biosensors.

Chapter 3

Multistability in the Lactose Utilization Network of *Escherichia coli*

3.1 Multistability

Multistability, the capacity to achieve multiple internal states in response to a single set of external inputs, is the essence of a biological switch. Biological switches are essential for the determination of cell fate in multicellular organisms [51], the activation of mitogen-activated protein kinase (MAPK) cascades in animal cells [52], the regulation of cell-cycle oscillations during mitosis with mutually exclusive cell cycle phases [53, 54, 55], the threshold response of lateral propagation of EGFR phosphorylation and the maintenance of epigenetic traits in microbes [56].

The multistability of several natural [51, 52, 53, 54, 55, 57, 56, 58, 59] and synthetic [49, 60, 61] systems has been attributed to positive feedback loops or mutually exclusive double negative feedback loops in their regulatory networks [62]. Many years ago [63], it was mathematically proven that the existence of at least one positive feedback loop is a necessary condition for having multiple steady states in the system. However, feedback alone does not guarantee multistability. The phase diagram of a multistable system, a concise description of internal states as key parameters are varied, reveals the conditions required to produce a functional switch [64, 65].

3.2 The lac operon

Complex developmental switches are built from smaller ones. We wanted to understand the basic underlying principles of a simple natural switch. If we can understand how this simple switch works quantitatively, we hope that, in the future, we will be able to understand and design more complex biological switches. With this goal in mind, we picked one of the most intensively studied natural networks; the lactose utilization network of *Escherichia coli* [66]. The bistability of the lactose utilization network has been under investigation since 1957 [67, 68].

The basic components of this network have been well characterized [66], making it an ideal candidate for global analysis. The lac operon comprises three genes required for the uptake and metabolism of lactose and related sugars (Figure 3-1): lacZ, lacY and lacA. lacZ codes for β-galactosidase, an enzyme responsible for the conversion of lactose into allolactose and subsequent metabolic intermediates. lacY codes for the lactose permease (LacY), which facilitates the uptake of lactose and similar molecules, including thio-methylgalactoside (TMG), a non-metabolizable lactose analogue. lacA codes for an acetyltransferase, which is involved in sugar metabolism. The operon has two transcriptional regulators: a repressor (LacI) and an activator (the cyclic AMP receptor protein, CRP). Inducers, among them allolactose and TMG, bind to and inhibit repression by LacI, whereas cAMP binds to and triggers activation by CRP. The concentration of cAMP drops in response to the uptake of various carbon sources, including glucose and lactose [69]; glucose uptake also interferes with LacY activity, leading to exclusion of the inducer [69]. Together these effects mediate catabolite repression; the ability of glucose to inhibit lac expression. Crucially, cAMP levels are not affected by TMG uptake. Therefore, the extracellular concentrations of TMG and glucose can be used to independently regulate the activities of LacI and CRP, the two cis-regulatory inputs of the lac operon [70]. However, the response of the operon must be considered within the broader context of the network. The uptake of TMG induces the synthesis of LacY, which in turn promotes further TMG uptake; the resulting positive feedback loop creates the potential for bistability [63, 71]

We wanted to use a phase diagram, coupled with a mathematical model of the network, to quantitatively investigate processes such as sugar uptake and transcriptional regulation in vivo and then to test whether or not the wild-type natural response of this genetic network could be changed by a perturbation into a completely different kind of response [72, 73].

3.3 Bistable response of the lac operon

In order to probe the network's bistable response, we incorporated a single copy of the green fluorescent protein gene (gfp) under the control of the lac promoter into the chromosome of *E. coli* MG1655 (Figure 3-1) (Appendix C). We placed this reporter in the chromosome rather than on a multicopy plasmid to minimize the titration of LacI molecules by extraneous LacI-binding sites. The cells also contained a plasmid encoding a red fluorescent reporter (HcRed) under the control of the galactitol (gat) promoter. This promoter includes a CRP-binding site, as well as a binding site for the galactitol repressor, GatR. However, GatR is absent in *E. coli* MG1655 [74]. Therefore, transcription at the gat promoter, measured by red fluorescence, is a direct measure of CRP-cAMP levels. In our experiments, we measure the response of single cells, initially in a given state of lac expression, to exposure to various combinations of glucose and TMG levels (Figure 3-2). It is crucial to use cells with well-defined initial states, either uninduced or fully induced, because the response of a bistable system is expected to depend on its history.

We find, in the absence of glucose, that the lac operon is uninduced at low TMG concentrations (< 3 μM) and fully induced at high TMG concentrations (> 30 μM) regardless of the cell's history. Between these switching thresholds, however, system response is hysteretic (history dependent): TMG levels must exceed 30 μM to turn on initially uninduced cells but must drop below 3 μM to turn off initially

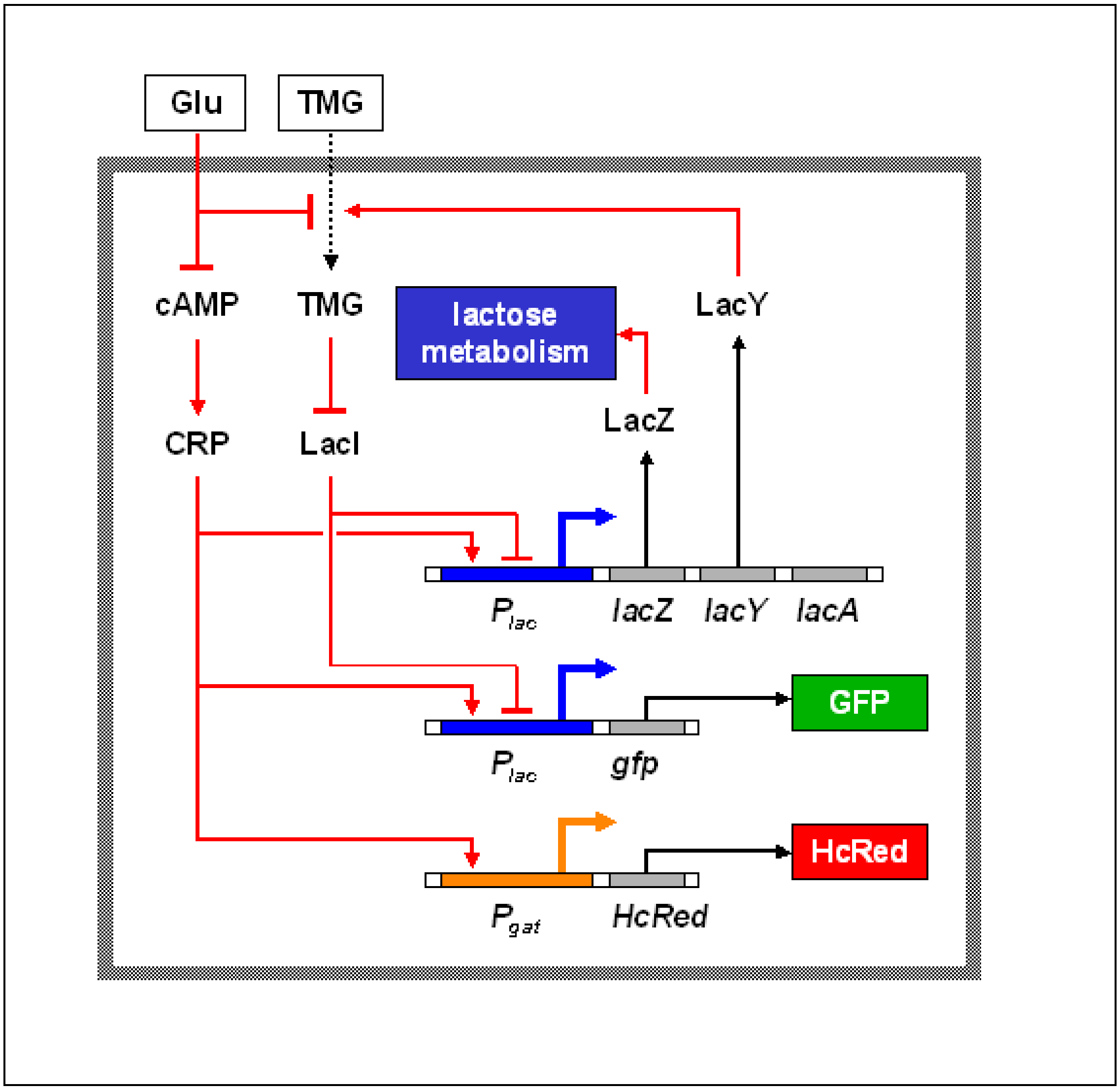
Glu
TMG
cAMP
TMG
lactose metabolism
LacY
CRP
LacI
LacZ
P_{lac}
lacZ
lacY
lacA
P_{lac}
gfp
GFP
P_{gat}
HcRed
HcRed

Figure 3-1: "The lactose utilization network. Red lines represent regulatory interactions, with pointed ends for activation and blunt ends for inhibition; black arrows represent protein creation through transcription and translation, and dotted arrows represent uptake across the cell membrane. In our experiments we vary two external inputs, the extracellular concentrations of glucose and TMG, and measure the resulting levels of two fluorescent reporter proteins: GFP, expressed at the lac promoter, and HcRed, expressed at the gat promoter. LacY catalyses the uptake of TMG, which induces further expression of LacY, resulting in a positive feedback loop."

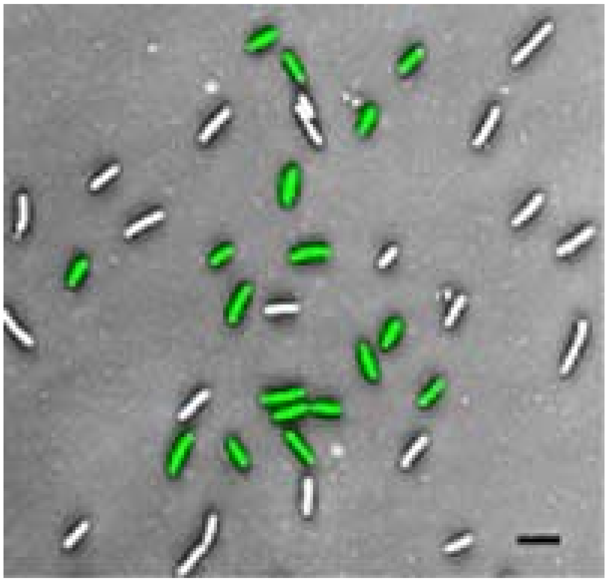

Figure 3-2: "Overlayed green fluorescence and inverted phase-contrast images of cells that are initially uninduced for lac expression, then grown for 20 h in 18 μM TMG. The cell population shows a bimodal distribution of lac expression levels, with induced cells having over one hundred times the green fluorescence of uninduced cells. Scale bar, 2 μm."

induced cells (Figure 3-3). As one approaches the boundaries of this bistable region, stochastic mechanisms cause growing numbers of cells to switch from their initial states, resulting in a bimodal distribution of green fluorescence levels, with induced cells having over one hundred times the fluorescence levels of uninduced cells. This behavior shows the importance of performing single-cell experiments, as a population-averaged measurement (Section 3.8) would have shown the mean fluorescence level to move smoothly between its low and high endpoints [53], obscuring the fact that individual cells never display intermediate fluorescence levels.

We find that the red fluorescence level is independent of cell history and of TMG concentration, showing that the observed history dependence of lac induction is not due to CRP. Red fluorescence levels do decrease in response to an increase in glucose concentration, ultimately dropping fivefold (Figure 3-4). There is a proportional drop in the green fluorescence levels of induced cells, reflecting the reduction in the levels of CRP-cAMP. The network continues to respond hysteretically in the presence of

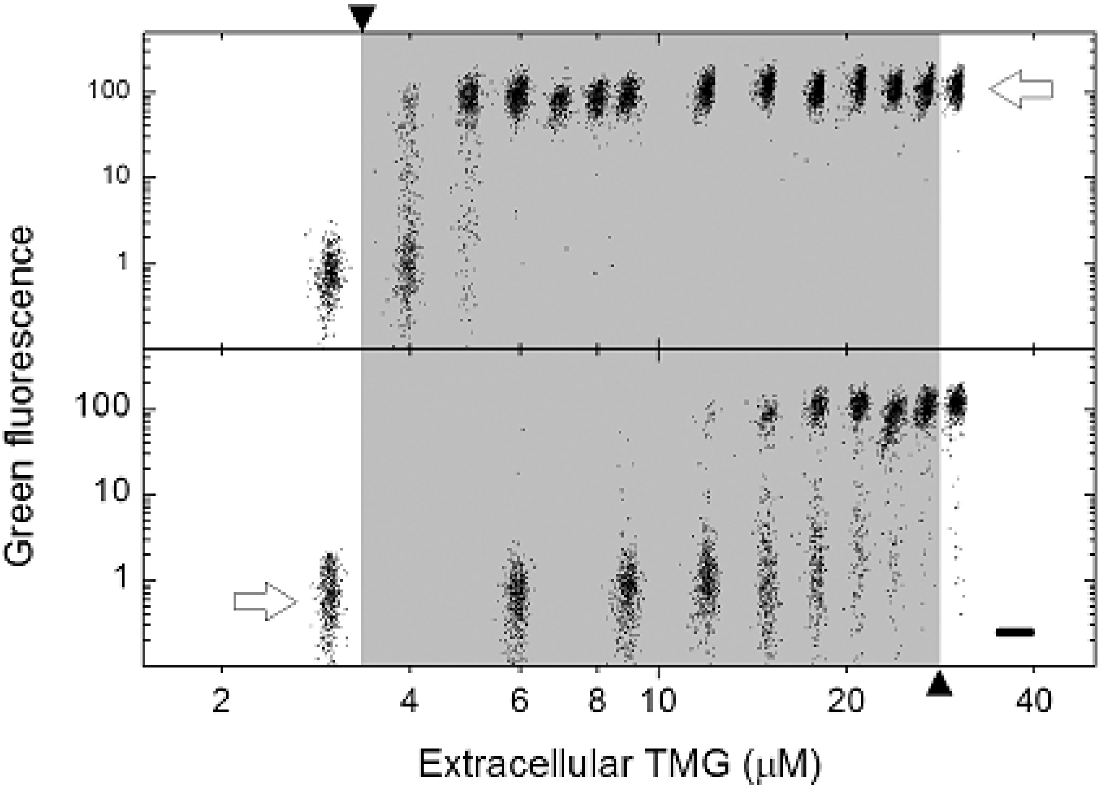

Figure 3-3: "Behavior of a series of cell populations, each initially uninduced (lower panel) or fully induced (upper panel) for lac expression, then grown in media containing various amounts of TMG. Scatter plots show log(green fluorescence) versus log(red fluorescence) for about 1,000 cells in each population. Each scatter plot is centered at a position that indicates the underlying TMG concentration. The scale bar represents variation in red fluorescence by a factor of 10. White arrows indicate the initial states of the cell populations in each panel. The TMG concentration must increase above 30 μM to turn on initially uninduced cells (up arrow), whereas it must decrease below 3 μM to turn off initially induced cells (down arrow). The grey region shows the range of TMG concentrations over which the system is hysteretic."

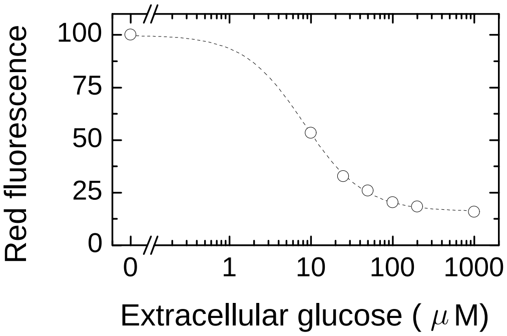

Figure 3-4: "The mean red fluorescence level of each cell population is independent of its history but decreases with increasing glucose concentrations."

glucose, but higher levels of TMG are required to induce switching. By measuring system response at several glucose concentrations ranging from 0 to 1 mM, we are able to map out the complete range of glucose and TMG levels over which the system is bistable (Figure 3-5). This is the phase diagram of the wild-type lactose utilization network.

3.4 Global analysis of the lactose transport network

The switching boundaries of this phase diagram correspond to special conditions of the network dynamics. By imposing these conditions within a mathematical model of the lac system, one is able to use the phase diagram as a quantitative probe of the molecular processes in living single cells. The lac system was modeled by using the following equations:

$$\frac{R}{R_T} = \frac{1}{1 + (x/x_0)^n} \tag{3.1}$$

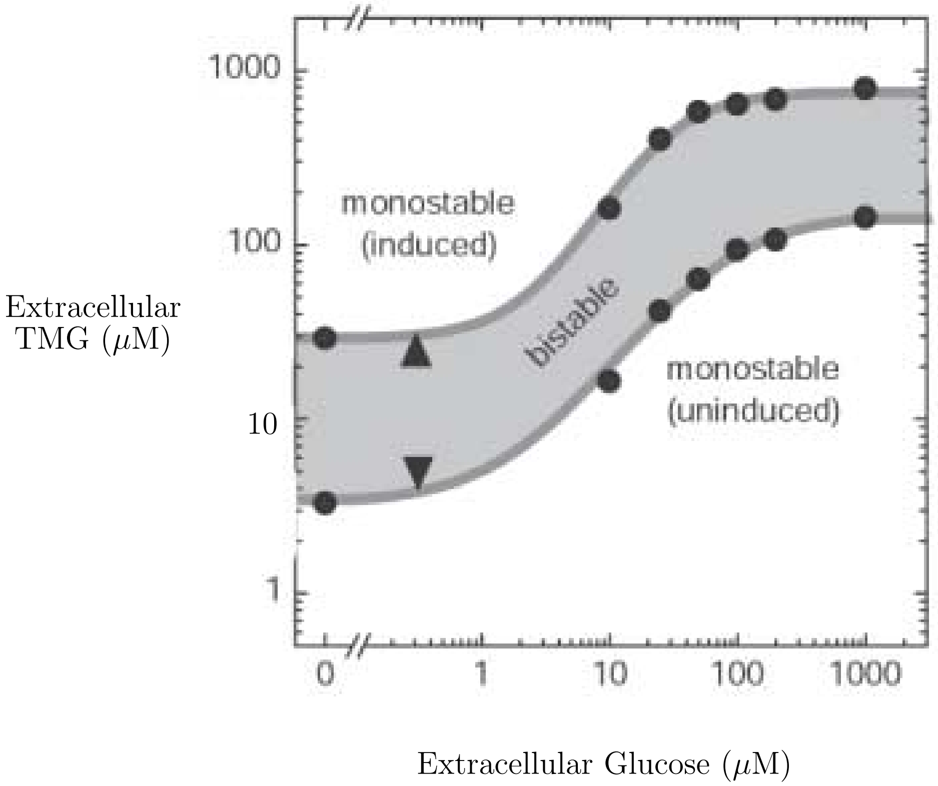

Figure 3-5: "The phase diagram of the wild-type lactose utilization network. When glucose is added to the medium, the hysteretic region moves to higher levels of TMG. At each glucose level, the lower (down arrow) and upper (up arrow) switching thresholds show those concentrations of TMG at which less than 5% of the cells are in their initial states."

$$\frac{\tau_y dy}{dt} = \alpha \frac{1}{1 + R/R_0} - y \tag{3.2}$$

$$\frac{\tau_x dx}{dt} = \beta y - x \tag{3.3}$$

Here, x is the intracellular TMG concentration, y is the concentration of LacY in green fluorescence units, R_T is the total concentration of LacI tetramers, and R is the concentration of active LacI. The active fraction of LacI is a decreasing sigmoidal function of the TMG concentration x, with half-saturation concentration x_0, and Hill coefficient n (Equation 3.1). This sigmoidal behavior arises from the fact that the binding of TMG to any one of four possible sites on the LacI tetramer is sufficient to interfere with LacI activity, while higher TMG occupancies cause even further impairment. There is extensive experimental evidence [75] showing that $n \approx 2$. The interaction of a single active LacI tetramer with multiple operator sites on the lac promoter generates a DNA loop which blocks transcription. The rate of generation of LacY is therefore a decreasing hyperbolic function of R, with maximal value α, half-saturation concentration R_0, and minimal value α/ρ achieved at $R = R_T$. The maximal activity, α, is the lac expression level that would be obtained if every repressor molecule were inactive. The repression factor, ρ, gives the ratio of maximal to basal activities, the latter being the expression level that would be obtained if every repressor molecule were active. The repression factor $\rho = 1 + R_T/R_0$ describes how tightly LacI is able to regulate lac expression. LacY is depleted in a first-order reaction with time constant τ_y, due to a combination of degradation and dilution (Equation 3.2). TMG enters the cell at a rate proportional to y, and is similarly depleted in a first-order reaction with time constant τ_x (Equation 3.3). The parameter β measures the TMG uptake rate per LacY molecule. Since we cannot directly measure x, we are free to choose its units so that $x_0 = 1$. Once inside the cell, TMG is able to inactivate LacI, completing the feedback loop. Combining these equations, one can obtain the steady state result:

$$y = \alpha \frac{1 + (\beta y)^2}{\rho + (\beta y)^2} \tag{3.4}$$

Here, ρ, α and β are allowed to be arbitrary functions of our external inputs, the extracellular glucose (G) and TMG (T) levels. As these parameters are varied, the system is capable of generating either one or two stable fixed points, with saddle-node bifurcations [64] separating these two behaviors (Figure 3-9d). The two fluorescence values at each fixed point do not by themselves provide enough information to uniquely specify the underlying parameters. However, by also applying the saddle-node condition at the switching thresholds, one is able to obtain three equations for the three unknowns ρ, α and β. We can therefore solve for these parameters at those values of G and T which lie on the boundaries of the bistable region. In this way, we obtain the complete functional dependence of these parameters on G and T as shown in the following section.

3.5 Theoretical phase diagram and calculation of parameters

The boundary between monostability (one stable fixed point) and bistability (two stable fixed points separated by one unstable fixed point) occurs when Equation 3.4 admits precisely two solutions: this signifies the onset of a saddle-node bifurcation. Rewriting Equation 3.4 as a cubic, we obtain:

$$y^3 - \alpha y^2 + (\rho/\beta^2)y - (\alpha/\beta^2) = 0 \tag{3.5}$$

On the other hand, a general cubic with two identical roots has the form:

$$(y - a)(y - a)(y - \theta a) = y^3 - (2 + \theta)ay^2 + (1 + 2\theta)a^2 y - \theta a^3 \tag{3.6}$$

where θ is the dimensionless ratio of roots. Comparing coefficients, we find:

$$\rho = (1+2\theta)(1+2/\theta) \tag{3.7}$$

$$\alpha\beta = (2+\theta)^{1.5}/\theta^{0.5} \tag{3.8}$$

These are the parametric equations describing the boundary of the bistable region (Figure 3-9d). The critical point occurs where all three roots coincide, so $\theta = 1$.

At every glucose concentration, we have two switching thresholds (OFF and ON). Given the two distinct fluorescence values at each threshold, we can calculate the values of all three parameters ρ, α and β. However, there are three caveats. First, we find that α is systematically higher (by about 15 %) at the OFF threshold than at the ON threshold, though it shows precisely the same linear behavior at both thresholds, over the entire range (a factor of 5 variation) of red fluorescence levels (Figure 3-4). This small discrepancy could arise from a systematic error in our estimate of the induced fluorescence, since our measurements are performed near but not precisely at the switching threshold. However, it might also arise due to a weak competitive interaction between CRP and LacI at the lac promoter, which we have neglected in our model. Second, the low fluorescence values at the OFF threshold are very close to background, introducing a large error in the calculation of ρ. We therefore estimate both α and ρ at the ON threshold alone, and use this information to calculate b at both thresholds. Third, we can decompose the net TMG uptake rate as β(T, G) = β_T(T) β_G(G), where T and G are the extracellular glucose and TMG concentrations. Assuming a power-law for β_T(T), we use a least-squares fitting routine to extract the functions β_T(T) and β_G(G). We have normalized the units to give β_T(G=0) = 100, and α(G=0) = 100. These calculations produce the following results (plotted in Figure 3-6, Figure 3-7 and Figure 3-8):

$$\alpha = \frac{84.4}{1+(G+8.1)^{1.2}} + 16.1 \tag{3.9}$$

$$\rho = 167.1 \tag{3.10}$$

$$\beta_T = (0.00123)T^{0.6} \tag{3.11}$$

$$\beta_G(G > 10) \cong 65 \tag{3.12}$$

where G and T are measured in μM. The uncertainities in the estimates of α and β_G are shown in Figure 3-8; the uncertainity in β_T is similar to that shown for β_G. There is a 50 % uncertainity in each measurement of ρ, and a 20 % uncertainity in its final fitted value.

We find that α is directly proportional to the red fluorescence level (upper panel in Figure 3-6), demonstrating that the lac and gat promoters respond identically to CRP-cAMP (section 3.7). By contrast, ρ is essentially independent of TMG levels, suggesting that LacI and CRP bind independently to the lac operator site. We find the repression factor ρ to be 170 ± 30 regardless of the glucose and TMG levels (lower panel in Figure 3-6). This confirms a strong prediction of our model that ρ should be a function of the LacI concentration alone. Assuming an effective LacI concentration in the nanomolar range cite [66], this value of ρ implies a dissociation constant between LacI and its major DNA-binding site of about 10^{-10} to 10^{-11} M, which is consistent with reported values [76].

The transport rate β is a product of two terms. The first term, β_T, which represents the TMG uptake rate per active LacY molecule, is purely a function of extracellular TMG levels (Figure 3-7). By fitting β_T to a hyperbola, we find that the half-saturation concentration for TMG uptake is 680 ± 25 μM, which agrees with previous measurements [77]. The second term, β_G, which represents the fraction of LacY molecules that are active, is purely a function of extracellular glucose levels, reflecting the inducer-exclusion effect [69]. This allows us to separate catabolite repression into its constituent parts (Figure 3-8). We find that by lowering CRP-cAMP levels glucose reduces operon expression by 80%, whereas by inactivating LacY molecules it reduces TMG uptake by 35%. However, the network's positive feedback architecture amplifies these effects, resulting in the observed hundred-fold difference in lac expression levels between induced and uninduced cells. This type of global information would be extremely difficult to obtain using standard molecular-genetic techniques and in vitro biochemical assays. Our approach allowed us to study the wild-type network in its entirety rather than isolated from other cellular systems or broken up into simpler subcomponents.

3.6 Transitions in the phase diagram

The phase diagram of the wild-type network (Figure 3-5) shows that lac induction always takes place hysteretically, with cells increasing their expression levels discontinuously as a switching threshold is reached. However, our theoretical phase diagram (Figure 3-9d) suggests that system response could also occur in a graded fashion, with the expression levels of individual cells moving continuously between low and high values. Such a response is predicted to occur when the degree of operon repression, ρ, is decreased below wild-type levels. Guided by this prediction, we sought to elicit a graded response from the lac system. We constructed two new strains, one containing a plasmid with an average copy number of 4, and the other containing a plasmid with an average copy number of 25. Each plasmid carried a single copy of the lac promoter. The introduction of extra LacI-binding sites had the expected titrating effect, reducing the effective LacI concentration and causing a drop in the operon-repression factor ρ from its wild-type value of 170, to 50 in the cells with 4 plasmids, and 5 in the cells with 25 plasmids. Graded behavior is expected to occur below a repression factor of nine, and this is precisely what we observe: cells with a repression factor of 50 have a discontinuous hysteretic response similar to that of the original cells (Figure 3-9b, bimodal histograms are shown in grey panels), whereas cells with a repression factor of 5 show a continuous graded response (Figure 3-9c).

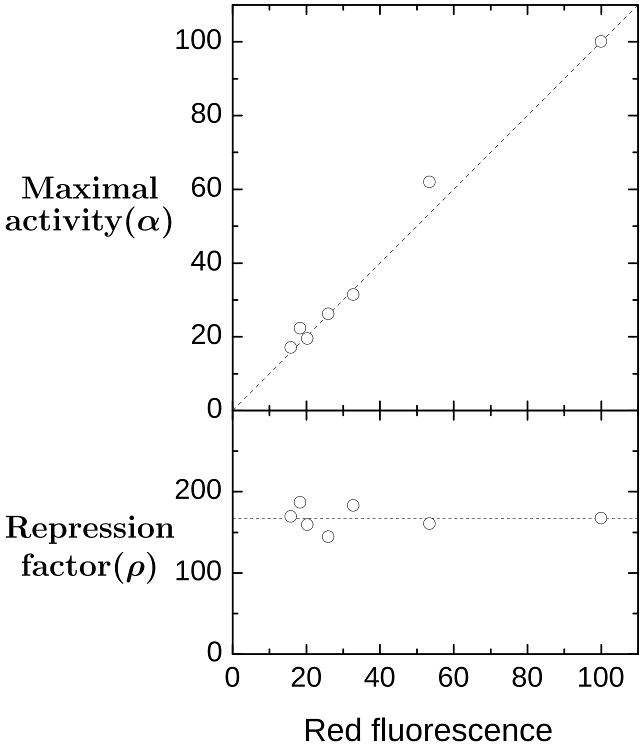

Figure 3-6: "Model parameters are extracted by fitting to measured fluorescence values at the switching thresholds. Upper panel: The maximal promoter activity, α, increases linearly with red fluorescence. Lower panel: The operon repression factor, ρ, is constant."

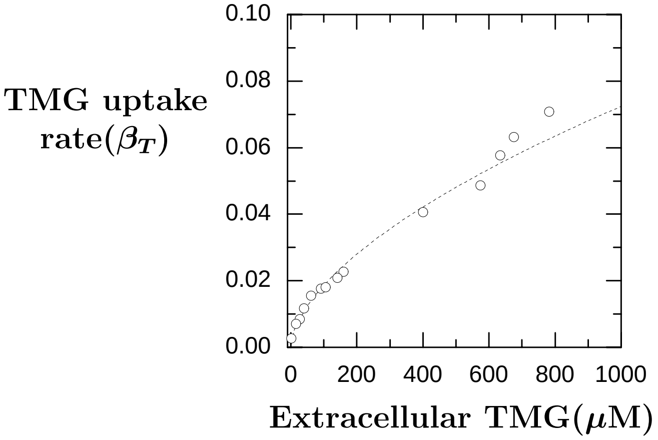

Figure 3-7: "The TMG uptake rate per active permease, β_T, increases with extracellular TMG levels. The dashed line shows a power-law fit with an exponent of 0.6. The data can also be fitted using a hyperbola, giving a half-saturation concentration of 680 μM."

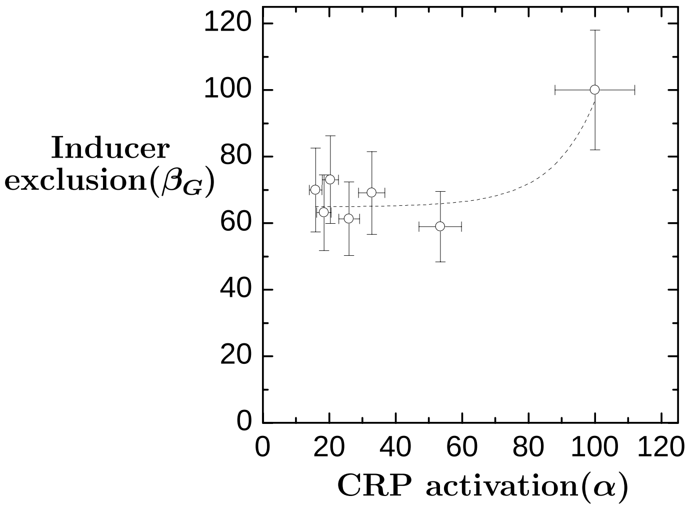

Figure 3-8: "The elements of catabolite repression. At each glucose concentration, we show the transport activity of LacY molecules (β_G, measuring inducer exclusion) versus the transcriptional activity of the lac operon (α, measuring CRP activation). We see that permease activity drops rapidly as glucose is added, falling to 65% of its maximal value. Operon activity drops more gradually, but falls to 20% of its maximal value. Error bars were determined by propagating the experimental error in measured fluorescence values."

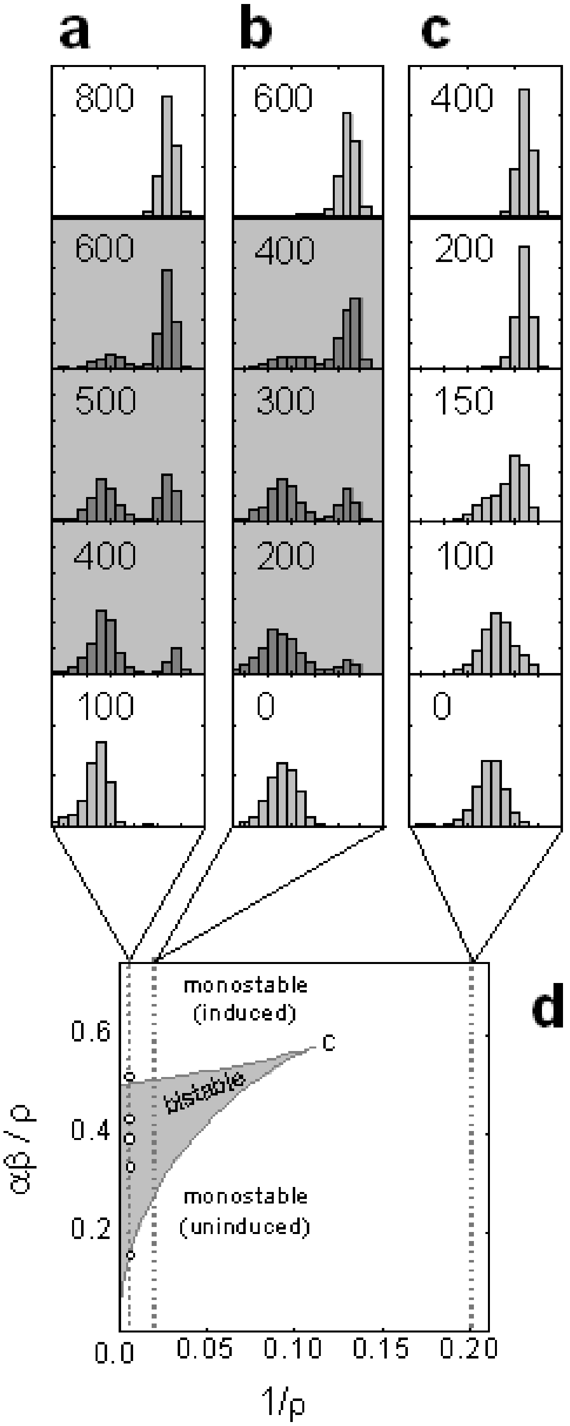

Figure 3-9: "Histograms of log(fluorescence) for cells that are initially uninduced, then grown in media containing 1 mM glucose and various levels of TMG (indicated in μM on each panel). a, Response of the wild-type network. b, c, Response of the network with extraneous LacI-binding sites on a 4-copy plasmid (b) and a 25-copy plasmid (c). d, Theoretical phase diagram of the lactose utilization network."

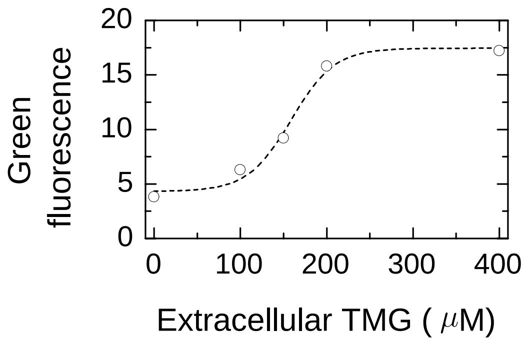

Figure 3-10: "The mean green fluorescence level of each cell population in Figure 3-9c is shown as a function of the TMG concentration. The response is highly sigmoidal (Hill coefficient ≈ 6) owing to positive feedback."

3.7 Correlation between green and red fluorescence values

The mean green fluorescence level of a population of induced cells was directly proportional to the mean red fluorescence level. This carried through to the behavior of single cells. Figure 3-11 shows a scatter plot of green and red fluorescence levels of single cells in a population that is initially uninduced, then grown in 10 mM glucose and 150 mM TMG. The population has a bimodal distribution of green fluorescence levels because it is close to a switching threshold. The green and red fluorescences of the induced subpopulation are highly correlated (correlation coefficient 0.71).

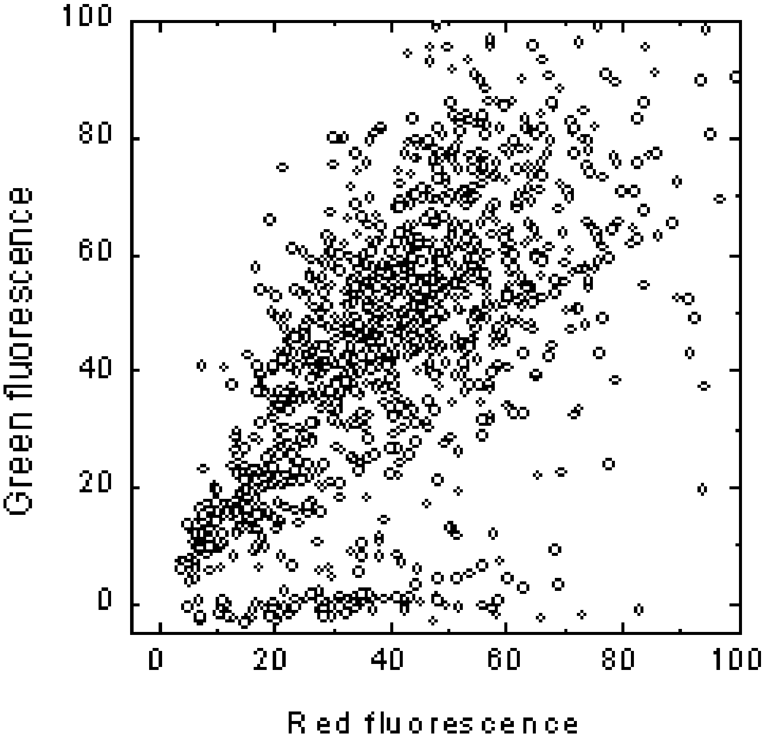

Figure 3-11: "Fluorescence levels of single cells in a bimodal population. Cells in the induced state show strongly correlated green and red fluorescence levels."

3.8 Population averaged measurements and mean fluorescences

Many previous studies of lac operon expression have focussed on population-averaged measurements [70], whereas we have used single cell measurements. These two approaches should produce similar results if the cell populations are homogeneous, but not if they are heterogeneous. In order to connect with previous studies, we incorporated our fitted parameters in a stochastic model [60], thus generating a map of mean lac expression levels as glucose and TMG concentrations were varied. In these calculations, we assumed that the system had reached a steady state distribution between the induced and uninduced sub-populations, corresponding to a timescale much longer than those at which hysteresis would be observed.

In a previous study [70], a series of population-averaged measurements of lac operon induction was conducted during growth on IPTG and cAMP. They found that the mean fluorescence of the cell population was a fairly intricate function of the two inducer concentrations. In particular, they found that the threshold for induction by IPTG depended on the cAMP level, and vice versa. In order to explain their results, the authors used a model of lac induction in which CRP and LacI bound competitively to the lac operator site. We show here that similar behavior can arise from a very different kind of model, in which the two regulators bind essentially independently, but positive feedback arises due to the uptake of inducer by LacY. We used such a model to analyze our results for growth on TMG; however, the fact that we find the system response to be bimodal during growth on IPTG implies that positive feedback must exist in that case as well. In Figure 3-12, we show the mean fluorescence as a function of TMG and glucose concentrations, using a stochastic model based on our fitted parameters. These results are qualitatively very similar to those of [70], with several thresholds and a strong interdependency between the effects of glucose and TMG.

The population-averaged fluorescence can increase either because the fluorescence level of induced cells increases (due to a decrease in glucose) or because the fraction

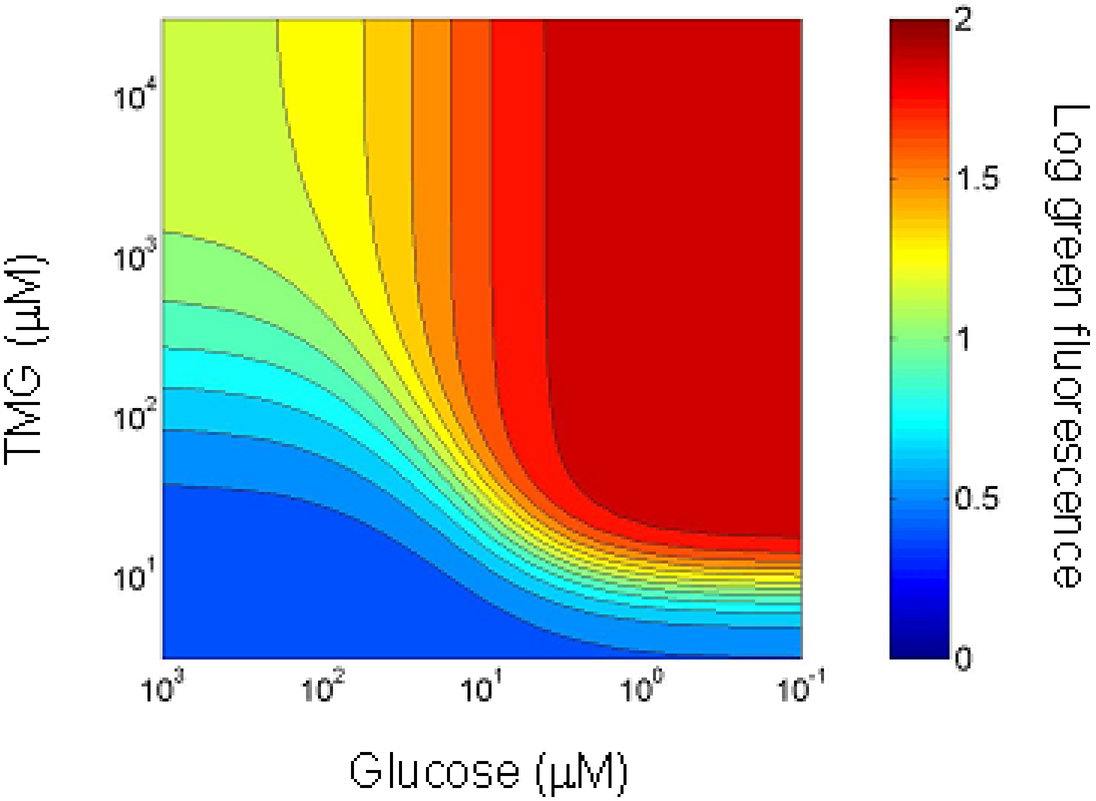

Figure 3-12: "Population averaged lac expression levels as a function of glucose and TMG concentrations. These results are obtained by incorporating our fitted parameters into a stochastic version of the positive feedback model. Axes are oriented so that cells are uninduced at the bottom left corner."

of cells in the induced state increases (due to an increase in TMG). It is only by performing single-cell measurements that these two effects can be distinguished.

3.9 Conclusion

The ability of a single system to produce both binary and graded responses has previously been observed but not explained [73]. We [78] presented a general mechanism by which this may be achieved and experimentally validated this mechanism in the context of a natural biological system, the lactose utilization network. We showed that the observed change from the wild-type binary response to the engineered graded response should be interpreted as a shift between different parts of a unified phase diagram. In this respect, the behavior of the lac system closely resembles that of thermodynamic systems [64]: the discontinuous transition from low to high induction is analogous to a first-order phase transition such as evaporation in a liquid-gas system, with chemical noise instead of thermal noise driving stochastic transitions between these states [60, 40]. The shift from a hysteretic to a graded response is analogous to a second-order phase transition across a critical point. More complex transitions, such as those between different types of multistability [79] or between stable and oscillatory behavior [80, 81], can be similarly analyzed. The response of any natural network may be regarded as a single realization in a vast but structured space of possible responses. By experimentally probing this space, we were able to gain quantitative insight into the architecture, dynamics and design constraints of biological systems.

Chapter 4

Conclusion and Future Directions

In the first half of my thesis, I discussed the regulation of fluctuations in the expression levels of a single gene. The results showed that cells can tune the mean and the variation in protein synthesis independently and the fluctuations in the protein levels can be regulated by changing the genetic parameters [44]. Other studies following our publication have focused on: the separation of gene intrinsic and extrinsic sources of fluctuations in protein synthesis [45] and noise in gene expression in eukaryotic cells [47].

All above mentioned studies focused on some of the first questions about the origin and regulation of the noise in gene expression. However, there are still a lot of unaddressed questions left for future studies. First of all, the time dynamics of fluctuations (temporal variability) in single cells has not been analyzed yet. Up to now, all the previous studies focused on the variability within a population of cells, not on the variability in a single cell at different time points. It was discussed in the introduction of this thesis that cells have complex interacting gene regulatory networks. Expression of many genes are controlled by regulatory proteins that are synthesized from an upstream gene in the network. How much noise is transferred from the gene expression of an upstream gene to the gene expression of a downstream gene in a gene regulatory cascade has not been studied yet. Although the intrinsic and extrinsic sources of noise in gene expression are identified, how much the fluctuations in the local environment of a gene (chromatin remodeling, changes in methylation or acethylation) affect the noise in the expression of that gene is left to be addressed. On the other hand, we have learnt that cells function very precisely even though they have inevitable fluctuations in gene expression. It has not yet studied in detail what common noise filtering mechanisms are used by multicellular organisms to achieve robustness during their development.

In the second half of my thesis, I studied a recurring gene network motif: a positive feedback loop. This feedback loop is implemented in the lactose transport network in *E. coli*. Hysteretic bistability is observed at the single cell level. A global analysis of the lactose transport network is done by modeling the system. The functional dependence of the key interaction parameters on the external inputs are obtained by doing in vivo single-cell experiments. The system response is changed from hysteretic bistability to graded response by the right perturbation as predicted by our model.

In the coming years, it will be possible to integrate the lessons we learned from the modular networks to understand bigger ones. Well-planned single-cell experiments will allow us to uncover how larger gene networks function so robustly in living cells. By doing these kinds of experiments one can reveal gene networks that has the same function in different species. Comparison of similar networks in different organisms will uncover which network connections are conserved in different species and allow us to identify core circuitry of gene networks. By looking at the differences in the interaction diagrams, one can identify specific advantages that these differences provide for each species to adapt to their environmental niches.

Appendix A

General Cloning Tools

A.1 PCR

The purpose of PCR (Polymerase Chain Reaction) is to make a large number of copies of a specific DNA sequence. PCR is a form of DNA cloning that is done outside of the cells using a purified, thermostable DNA polymerase enzyme. This type of DNA amplification requires a prior knowledge of the gene sequence (usually fewer than 5000 bp) that is to be amplified. Cloning by using PCR is much faster and easier than other standard cloning methods.

This technique, allows the DNA from a selected region of any genome to be amplified a billion-fold. Before starting a PCR reaction, the known sequence has to be used to design two synthetic DNA oligonucleotides (usually 25 nucleotides in length), one complementary to each strand of the DNA and positioned on opposite sides of the region to be amplified. These oligonucleotides serve as primers for in vitro DNA synthesis, which is catalyzed by a DNA polymerase, and they determine the ends of the final DNA fragment that is obtained.

There are three major steps in a PCR (Figure A-1), which are repeated about 20 to 30 cycles. This is done on an automated cycler, which can heat and cool the tubes with the reaction mixture in a very short time. Each cycle of the reaction requires a brief heat treatment (usually 94^{o}C) to separate the two strands of the genomic DNA (step 1: denaturation). The success of this technique depends on the use of

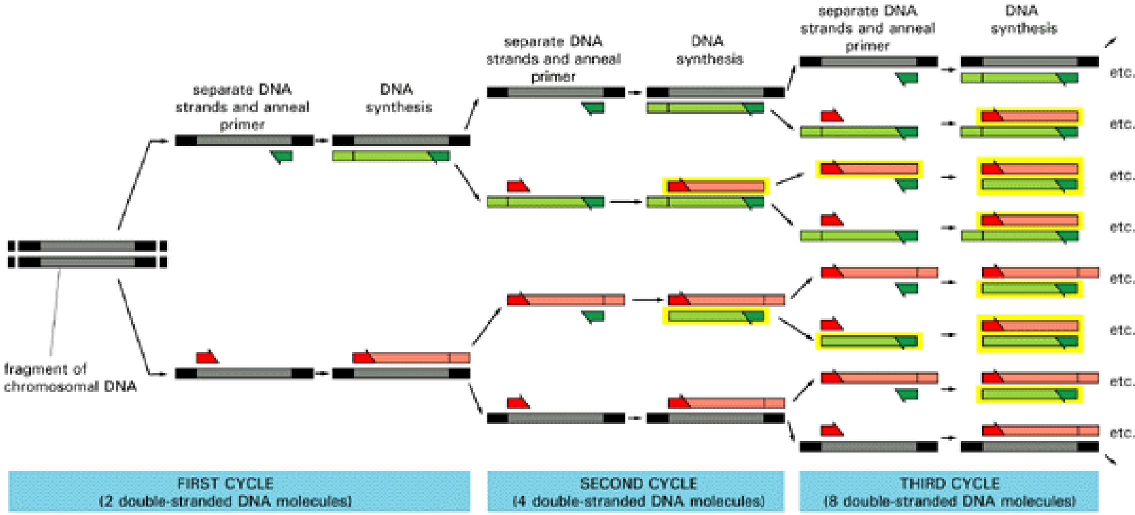
separate DNA strands and anneal primer
DNA synthesis
separate DNA strands and anneal primer
DNA synthesis
separate DNA strands and anneal primer
DNA synthesis
etc.
fragment of chromosomal DNA
FIRST CYCLE
(2 double-stranded DNA molecules)
SECOND CYCLE
(4 double-stranded DNA molecules)
THIRD CYCLE
(8 double-stranded DNA molecules)

Figure A-1: "PCR amplification. PCR produces an amount of DNA that doubles in each cycle of DNA synthesis and includes a uniquely sized DNA species. Three steps constitute each cycle, as described in the text. After many cycles of reaction, the population of DNA molecules becomes dominated by a single DNA fragment, X nucleotides long, provided that the original DNA sample contains the DNA sequence that was anticipated when the two oligonucleotides were designed. In the example illustrated, three cycles of reaction produce 16 DNA chains, 8 of which have this unique length (yellow); but after three more cycles, 240 of the 256 DNA chains would be X nucleotides long. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc."

a special DNA polymerase isolated from a thermophilic bacterium that is stable at much higher temperatures than normal, so that it is not denatured by the repeated heat treatments.

Normally, the primers in the reaction mixture are moving around, caused by the Brownian motion. Ionic bonds are constantly formed and broken between the single stranded primer and the single stranded DNA template. More stable bonds last a little bit longer (primers that perfectly match) and on that piece of double stranded DNA (template and primer), the polymerase can attach and starts copying the template. Once there are a few bases built in, the ionic bond is so strong between the template and the primer, that it does not break anymore. To start this attachment, the reaction mixture is cooled down to, usually, 55^{o}C to 65^{o}C. These temperature ranges allow oligonucleotides (primers) to hybridize to complementary sequences in the genomic DNA (step 2: annealing).

The annealed mixture is then extended by the polymerase till the regions of DNA in between each of the two primers are synthesized (step 3: extension). This step is carried at 72^{o}C, which is the ideal working temperature for the polymerase.

As these steps are repeated, the newly synthesized fragments serve as templates in their turn, and within a few cycles the predominant product is a single species of DNA fragment with the desired length. In practice, 20 to 30 cycles of reaction are required for effective DNA amplification. Each cycle doubles the amount of DNA synthesized in the previous cycle. A single cycle requires only about 3-5 minutes, and an automated procedure permits cell-free cloning of a DNA fragment within a few hours, compared with the several days required for standard cloning procedures [2].

Later, the PCR product has to be checked by running it on an agarose gel. DNA pieces of different length run at different speed on the gel. If the right length product is obtained, one can pass to the next stages of the cloning [82].

A.2 Insertion of DNA segments into plasmid vectors

The plasmid vectors are self-replicating shuttles, which are used for gene cloning. They are small circular molecules of double-stranded DNA derived from larger plasmids that occur naturally in bacterial cells. For use as cloning vectors, the purified plasmid DNA circles are first cut with a restriction nuclease to create linear DNA molecules. The PCR-amplified DNA is also cut with the same restriction nuclease. These digestion reactions are done by using specific enzymes that recognize different but well-determined sequences. Under their preferred salt concentrations and temperature conditions, they bind to their recognition sequences and cut DNA molecules into two pieces with cohesive ends.

The resulting restriction fragments (the gene to be cloned and plasmids) can be an-

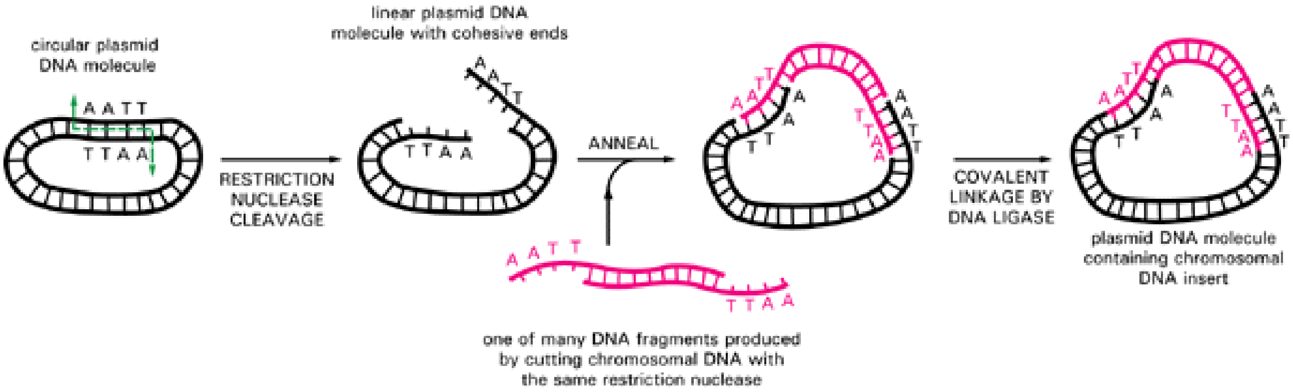
circular plasmid DNA molecule
A A T T
T T A A
RESTRICTION NUCLEASE CLEAVAGE
linear plasmid DNA molecule with cohesive ends
A A T T
T T A A
ANNEAL
A A T T
T T A A
one of many DNA fragments produced by cutting chromosomal DNA with the same restriction nuclease
COVALENT LINKAGE BY DNA LIGASE
plasmid DNA molecule containing chromosomal DNA insert

Figure A-2: "The formation of a recombinant DNA molecule. The cohesive ends produced by many kinds of restriction nucleases allow two DNA fragments to join by complementary base-pairing. DNA fragments joined in this way can be covalently linked in a highly efficient reaction catalyzed by the enzyme DNA ligase. In this example a recombinant plasmid DNA molecule containing a chromosomal DNA insert is formed. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc."

nealed to each other via their cohesive ends to form recombinant DNA circles. These recombinant molecules containing foreign DNA inserts are then covalently sealed by the enzyme "DNA ligase" (Figure A-2) [2, 82].

A.3 Transformation of plasmids into bacteria

In the next step of cloning, the recombinant DNA circles (plasmids) are introduced into bacterial cells. This step is called "plasmid transformation". Before this step, bacteria have to be made transiently permeable to foreign DNA. Typically, this is achieved by treating the cells with special media for a few hours and then by applying a mild heat-shock to these cells. As these cells grow and divide, doubling in number every 20 minutes, the recombinant plasmids also replicate to produce an enormous number of copies of DNA circles containing the foreign DNA (Figure A-3). The cells that accepted foreign DNA have to be separated from other cells that failed to do so. Many bacterial plasmids carry genes for antibiotic resistance, this property is used to select those cells that have been successfully transfected. When the bacteria are grown in the presence of the antibiotic, only cells containing plasmids will be able to develop antibiotic resistance and thus survive [2, 82].

A.4 Gene insertion into the chromosome of Bacillus subtilis

An integrable plasmid, which can replicate independently in Escherichia coli is used for inserting any gene of interest into the chromosome of B. subtilis as a single copy. This plasmid could only be integrated into the chromosome of B. subtilis if it contains sequences homologous to chromosomal sequences of the bacteria. These kinds of plasmids carry a selectable antibiotic resistance and unique sites for the ligation of gene of interest. The amyE locus, coding for a nonessential α-amylase, is used in most of the cases for integration. The antibiotic resistance marker and a multiple cloning site sandwiched between the two halves of the amyE gene, designated amyE-front and amyE-back. Upon transformation of B. subtilis cells, both amyE sequences will recombine at their homologous sites, thereby stably inserting the DNA sequences in between amyE-front and amyE-back into the B. subtilis chromosome via a double-crossover events [83, 84]. The cells stable integrated the foreign plasmid will loose the amyE. This can be checked by growing cells under conditions when they need α-amylase to digest starch and continue to grow.

A.5 Gene insertion into the chromosome of E. coli

Lambda InCh (”Into the Chromosome”) vectors are derived from bacteriophage lambda, which mediates the transfer of cloned DNA from pBR322-type or pUC-type plasmids into the E. coli chromosome. The transfer depends on three regions of homology. At these regions, recA dependent recombination is used as the mechanism for genetic exchange.

The first of these recombination events occurs during growth of the phage in a cell

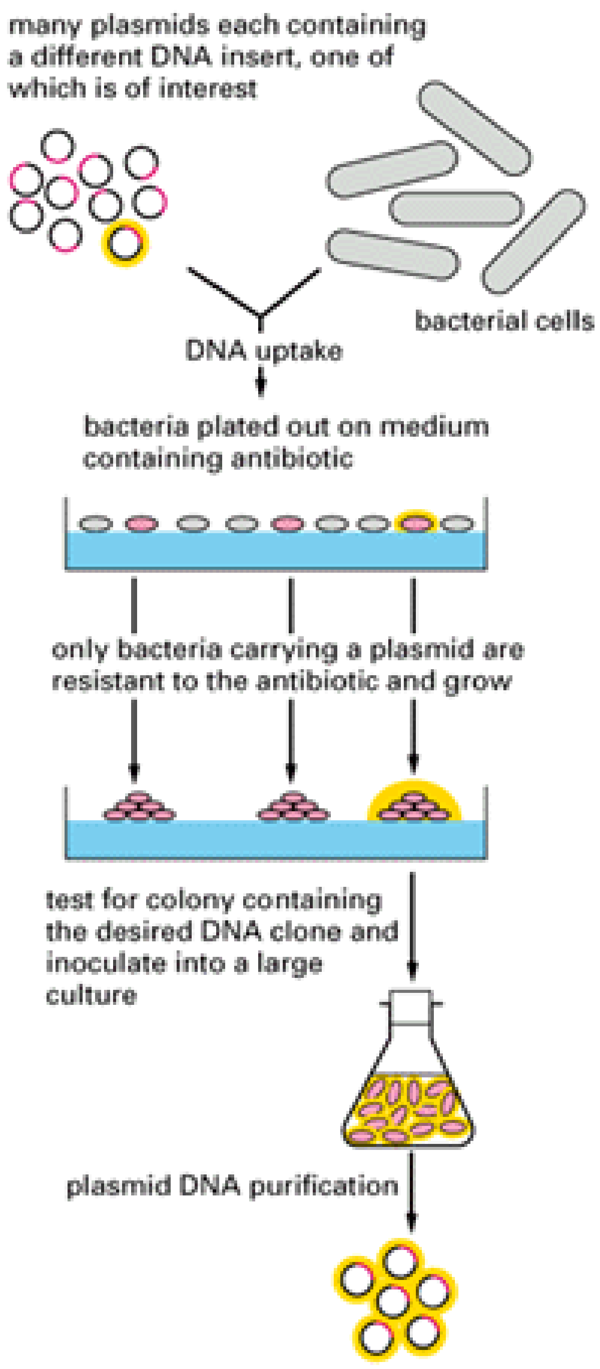
many plasmids each containing a different DNA insert, one of which is of interest
bacterial cells
DNA uptake
bacteria plated out on medium containing antibiotic
only bacteria carrying a plasmid are resistant to the antibiotic and grow
test for colony containing the desired DNA clone and inoculate into a large culture
plasmid DNA purification

Figure A-3: "Purification and amplification of a specific DNA sequence by DNA cloning in a bacterium. Each bacterial cell carrying a recombinant plasmid develops into a colony of identical cells, visible as a spot on the nutrient agar. By inoculating a single colony of interest into a liquid culture, one can obtain a large number of identical plasmid DNA molecules, each containing the same DNA insert. Copyright ©Molecular Biology of the Cell by B. Alberts and A. Johnson and J. Lewis and M. Raff and K. Roberts and P. Walter. Reproduced by permission of Garland Science/Taylor and Francis books, Inc."

containing a plasmid. Recombination at one of these regions shared by the plasmid and the phage results in a cointegrate formation. Resolution of the cointegrate then results in transfer of genetic material from the plasmid to the phage. The second step can happen either during growth of the phage or at a later step.

Once the plasmid insert has recombined into the phage, replacing the KanR allele with a complete bla allele along with whatever is cloned between the two homologous regions, an Ampicillin resistant lysogen can be selected. This involves site specific recombination at the lambda attachment site (att) on the E. coli chromosome.

In the last step, nearly all of the lambda DNA is removed by another homologous recombination event. An 800 bp fragment of the chromosome right next to the att site is cloned in the phage so that in the lysogen there is a direct repeat. Recombination between these regions loops out the intervening sequence deleting it. This is easily selected since the phage has a temperature sensitive repressor, cI857, and is induced at 42^oC killing any cell in which the deletion has not occurred. This results in a temperature independent (phage cured) strain with a stable single copy of the insert from the plasmid on the chromosome [85].

Appendix B

Methods 1

B.1 Strains, growth conditions and media

We placed the gene gfpmut2 under the control of the P_{spac} promoter and introduced mutations in the ribosome binding site, initiation codon and promoter region of gfpmut2 by PCR (see Appendix A.1). Mutations were verified by sequencing; spontaneous mutation frequencies were negligible over the timecourse of our experiments. We digested the PCR products and ligated them into the amyE integration vector pDR67, which contains a single copy of lacI downstream of the constitutive promoter P_{pen} (see Appendix A.2). We amplified the resulting recombinant plasmid in the E. coli AG1111 strain (see Appendix A.3) and inserted it into the chromosome of the B. subtilis JH642 strain by double-crossover at the amyE locus (see Appendix A.4). Cells of E. coli and B. subtilis were made competent and transformed according to standard procedures [82]. The resulting B. subtilis strain contained a single copy of gfpmut2 under the P_{spac} promoter and a single copy of lacI under the constitutive P_{pen} promoter. The P_{spac} promoter includes a binding site for Lac repressor, the product of the lacI gene; P_{spac} is externally inducible by IPTG, which binds to and inhibits the repressor function of LacI. The concentration of IPTG in the growth medium therefore determines the transcriptional efficiency of gfpmut2. Addition of IPTG is not expected to affect native operon expression in B. subtilis.

We grew cells overnight in Luria Bertani (LB) broth at 37^{o}C, diluted these cultures and induced them with varying amounts of IPTG for at least 5 h at 37^oC. We grew non-induced strains to determine the amount of background fluorescence due to autofluorescence. The background fluorescence is very similar to the fluorescence measured for the B. subtilis JH642 strain lacking gfpmut2. This implies that the P_{spac} promoter is tightly controlled.

B.2 Data acquisition and analysis

We collected cells from growth cultures at $OD_{600} \approx 1.0$, which corresponds to the late exponential phase. To eliminate cell aggregates, we centrifuged cells at 4,000 rpm for 1 min, pelleted the supernatant at 14,000 rpm for 1 min and resuspended the pellet in PBS. We independently confirmed the distributions of cell shapes using fluorescence microscopy. Single-cell fluorescence measurements were carried out on a Becton-Dickinson FACScan flow cytometer with a 488-nm Argon excitation laser and a 525-nm emission filter. FACScan data were analyzed on a Macintosh Quadra 650 using the Cell Quest program. During each flow-cytometer experiment, we collected data from $10^4 - 10^5$ single cells; each run typically lasted for 2 min and was conducted at room temperature. Cells from the same sample were often analyzed in two runs separated by 15 min or more. The measured fluorescence distribution was unchanged both during the course of a single run and between two such runs. To reduce noise in fluorescence values resulting from different cell sizes, we analyzed cells using the smallest allowed gate in the side-scattering and forward-scattering space.

B.3 Determination of transcriptional and translational efficiencies

For the translational mutants, we defined the transcriptional efficiency as the average fluorescence measured for a specific strain at a certain IPTG concentration normalized to the average fluorescence measured for that strain at full induction ([IPTG] = 1 mM). The translational efficiency of a strain was defined as the average fluorescence of the strain at full induction normalized to that of the wildtype strain (ERT25). For the transcriptional mutants, we defined transcriptional efficiency for each strain as the average fluorescence measured at full induction normalized to that of the ERT25 strain. We determined parameter error bars over at least 20 repeated measurements.

B.4 Monte Carlo simulations

Simulations were implemented using Gillespie's algorithm for stochastic coupled chemical reactions [86]. The reactions simulated are those schematically indicated in Figure 2-8. We assume individual reactions to be Poisson, so that the probability of a reaction with rate k happening in a time dt is given by kdt, and the waiting times between successive reactions are exponentially distributed. We assume that steady-state has been reached at a time equal to ten times the protein half-life. Each simulated histogram is the result of 5,000 trials.

B.5 Software

We converted data obtained in flowcytometer to ASCII format using MFI (E. Martz, Univ. of Massachusetts, Amherst, available at http://www.umass.edu/microbio/mfi.

Appendix C

Methods 2

C.1 Bacterial strains and plasmids

The gfp gene under the control of the wild-type lac promoter, obtained from plasmid pGFPmut3.1 (Clontech), was inserted into the chromosome of E. coli MG1655 at the lambda insertion site using the λ-InCH technique [85] (see Appendix A.5) to produce the strain MUK21. The gat promoter was amplified from the E. coli MG1655 chromosome by polymerase chain reaction (see Appendix A.1) using primers flanking the 2,175,231-2,175,531 chromosomal region. The HcRed gene was obtained from pHcRed1-C1 (Clontech) and was placed under the control of the gat promoter into a plasmid with a ColE1 replication origin (see Appendix A.2), which was transformed into MUK21 cells (see Appendix A.3) to obtain the strain ERT113. All measurements of wild-type network response were conducted in this strain. Two additional strains with lac operon repression factors at lower levels than in the wild-type were constructed by transforming MUK21 cells with multicopy plasmids, each incorporating a single copy of the lac promoter. The strain MUK21-pSC101* contains plasmids with a pSC101* replication origin (average copy number 4), and the strain MUK21-p15A contains plasmids with a p15A replication origin (average copy number 25)[37].

C.2 Growth conditions and media

Cells were grown at 37^oC in M9 minimal medium with succinate as the main carbon source, supplemented with varying amounts of glucose and TMG. Master cultures with cells induced for lac expression were prepared by overnight growth in 1 mM TMG, and master cultures with uninduced cells by overnight growth in the absence of TMG. During each experimental run, cells were transferred from these master cultures into media containing specified amounts of glucose and TMG. They were subsequently grown for 20 additional hours before they were harvested for measurement. The transfer volume was calculated to produce extremely low final cell densities (OD_{600} $\sim$ 0.001), thereby preventing the depletion of glucose and TMG from the medium. Cells were concentrated by filtration and centrifugation, and the resulting pellet was resuspended in 2.5 μl of the growth medium to prepare a microscope slide.

C.3 Data acquisition

Green and red fluorescence values of single cells were measured using a Nikon TE2000 microscope with automated stage and focus. For each experiment, images of about 1,000 cells on each slide were collected using a cooled back-thinned CCD camera (Micromax, Roper Scientific). These images were analyzed using Metamorph (Universal Imaging) to obtain the average fluorescence of each cell above the fluorescence background.

C.4 Data analysis

For each glucose concentration, the fraction of cells in the induced state was determined as a function of TMG concentration, and the switching thresholds (defined as the TMG concentrations at which less than 5% of the cells are in their initial states) were obtained by interpolation. We estimated the green fluorescence values of the induced (high) and uninduced (low) subpopulations at each switching threshold by averaging over two neighboring TMG concentrations. At each threshold, the high fluorescence was a linear function of the low fluorescence, with a small positive intercept comparable to the autofluorescence of E. coli MG1655 cells. We interpreted this intercept as the autofluorescence of the ERT113 strain. The repression factors for the MUK21-pSC101* and MUK21-p15A strains were estimated by taking the ratio of fully induced to uninduced fluorescence levels, assuming that these strains had the same autofluorescence background as ERT113.

C.5 Calculation of the repression factor

The most direct way to measure the operon repression factor is to take the ratio of fluorescence levels of fully induced cells (grown in saturating amounts of TMG) and uninduced cells (grown in the absence of TMG). However, because our reporter is present in single copy in the chromosome, the fluorescence levels of uninduced cells are very close to the measurement background of the camera, and comparable to cell autofluorescence. We therefore chose to determine the repression factor by the fitting technique discussed above. We obtained a repression factor $\rho = 170$ for the wild-type system, whereas previous studies [70] report much higher repression factors, of order 1000. This difference could be due to the following reasons. First, the wild-type lac promoter contains three operator sites (O_1, O_2 and O_3) to which LacI binds. The lac promoter used in our reporter construct is missing the O_2 operator site, leading to a decrease in repression efficiency due to a drop in DNA looping activity. Such a promoter has been reported [76] to have a repression factor of only 440. (Note that the native copy of the lac promoter still contains all three operator sites, so system response is unaffected.) Second, our fit is performed only at the switching thresholds, far from either the fully induced or uninduced limits. Our fitted value of ρ will probably be different from the value that would be obtained by direct measurement at the two limits, due to small differences between our approximate model and the true system response. However, the fact that we find ρ to be independent of glucose and TMG levels strongly suggests that the true repression factor is similarly independent of these parameters.

C.6 Growth in IPTG and lactose

For completeness, we conducted a series of experiments using lactose and isopropyl-β-D-thiogalactopyranoside (IPTG) as inducers in place of TMG.

During induction with IPTG, cells show a persistent bimodal response, but the fluorescence levels of uninduced cells are higher than with TMG. It is known that IPTG is able to enter cells independently of LacY, accounting for the increased fluorescence of the uninduced cells. However, the persistence of bimodality indicates that LacY continues to play a role in the active transport of IPTG [87], preserving positive feedback. Under these circumstances, we might expect the phase diagram during growth on IPTG to be very similar to the phase diagram we have measured using TMG. Indeed, as an indirect evidence of this, a population-averaged version of our TMG-glucose phase diagram bears a striking resemblence to the results of recent population-averaged measurements of lac expression using IPTG and cAMP as inducers [70] (Section 3.8).

During induction with lactose, initially uninduced cell populations show a transient bimodal distribution of green fluorescence levels at certain glucose concentrations, and a transient unimodal distribution at others. However, the steady state distribution after 4 hours of growth is always unimodal, and we never observe hysteresis. By performing extensive measurements, we confirmed this unimodal behaviour to occur for over fifty combinations of glucose and lactose concentrations, upto saturating quantities of each sugar. The difference between the observed responses to TMG and lactose could be due to several causes. First, because lactose is metabolized and therefore affects cell growth rate, it could happen that the induced sub-population of a bimodal population always grows to dominance. Second, since the metabolism of lactose leads to a drop in CRP-cAMP levels [69], the inducer activity of allolactose might be counteracted to some extent. Third, although an increase in operon expression leads to an increase in lactose uptake and allolactose production, it also leads to an increase in allolactose degradation by β-galactosiadase. Intracellular allolactose levels therefore depend very weakly on operon expression levels, reducing the strength of positive feedback and possibly eliminating bistability altogether.

Bibliography

[1] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys, and M. Tyers. Systematic identification of protein complexes in *saccharomyces cerevisiae* by mass spectrometry. *Nature*, 415(6868):180–183, 2002.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland/Taylor and Francis, 2002.

[3] L.J. Mota, P. Tavares, and I. Sa-Nogueira. Mode of action of arar, the key regulator of l-arabinose metabolism in *bacillus subtilis*. *Molecular Microbiology*, 33(3):476–489, 1999.

[4] S. Busby and R.H. Ebright. Transcription activation by catabolite activator protein (cap). *Journal of Molecular Biology*, 293(2):199–213, 1999.

[5] D.E. Clyde, M.S.G. Corado, X. Wu, A. Par, D. Papatsenko, and S. Small. A self-organizing system of repressor gradients establishes segmental complexity in *drosophila*. *Nature*, 426(6968):849–853, 2003.

[6] L. Giot, J.S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y.L. Hao, C.E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C.A. Stanyon, R.L.Jr. Finley, K.P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R.A. Shimkets, M.P. McKenna, J. Chant, and J.M. Rothberg. A protein interaction map of *drosophila melanogaster*. *Science*, 302(5651):1727–1736, 2003.

[7] A. Davy, P. Bello, N. Thierry-Mieg, P. Vaglio, J. Hitti, L. Doucette-Stamm, D. Thierry-Mieg, J. Reboul, S. Boulton, A.J. Walhout, O. Coux, and M. Vidal. A protein-protein interaction map of the *caenorhabditis elegans* 26s proteasome. *EMBO Rep.*, 2(9):821–828, 2001.

[8] B.L. Drees, B. Sundin, E. Brazeau, J.P. Caviston, G.C. Chen, W. Guo, K.G. Kozminski, M.W. Lau, J.J. Moskow, A. Tong, L.R. Schenkman, A.3rd McKenzie, P. Brennwald, M. Longtine, E. Bi, C. Chan, P. Novick, C. Boone, J.R. Pringle, T.N. Davis, S. Fields, and D.G. Drubin. A protein interaction map for cell polarity development. *J Cell Biol.*, 154(3):549–571, 2001.

[9] A. Arkin, J. Ross, and H.H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *escherichia coli* cells. *Genetics*, 149(4):1633–1648, 1998.

[10] B. Houchmanzadeh, E. Wieschaus, and S. Leibler. Establishment of developmental precision and proportions in the early *drosophila* embryo. *Nature*, 415(6873):798–802, 2002.

[11] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA*, 98(15):8614–8619, 2001.

[12] H.H. McAdams and A. Arkin. It's a noisy business! genetic regulation at the nanomolar scale. *Trends Genet*, 15(2):65–69, 1999.

[13] H.H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* , 94(3):814–819, 1997.

[14] M. Ptashne. *A Genetic Switch: Phage and Higher Organisms.* Cell Press and Blackwell Scientific Publications, 1992.

[15] J.D. Chung, G. Stephanopoulos, K. Ireton, and A.D. Grossman. Gene expression in single cells of *bacillus subtilis*: evidence that a threshold mechanism controls the initiation of sporulation. *J Bacteriol.*, 176(7):1977–1984, 1994.

[16] C.V. Rao, D.M. Wolf, and A.P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420(6912):231–237, 2002.

[17] A. Hernday, M. Krabbe, B. Braaten, and D. Low. Self-perpetuating epigenetic pili switches in bacteria. *Proc. Natl. Acad. Sci. USA*, 99, 2002.

[18] I. Connell, W. Agace W, P. Klemm, M. Schembri, S. Marild, and C. Svanborg. Type 1 fimbrial expression enhances *escherichia coli* virulence for the urinary tract. *Proc. Natl Acad. Sci. USA*, 93(18):9827–9832, 1996.

[19] I.R. Peak, M.P. Jennings, D.W. Hood, M. Bisercic, and E.R. Moxon. Tetrameric repeat units associated with virulence factor phase variation in *haemophilus* also occur in *neisseria spp.* and *moraxella catarrhalis. FEMS Microbiol. Lett.*, 137(1):109–114, 1996.

[20] C.F. Marrs, W.W. Ruehl G.K. Schoolnik, and S. Falkow. Pilin-gene phase variation of *moraxella bovis* is caused by an inversion of the pilin genes. *J. Bacteriol.*, 170(7):3032–3039, 1988.

[21] I.J. Mehr and H.S. Seifert. Differential roles of homologous recombination pathways in *neisseria gonorrhoeae* pilin antigenic variation, dna transformation and dna repair. *Mol. Microbiol.*, 30(4):697–710, 1998.

[22] V.d.P. Putte and N. Goosen. Dna inversions in phages and bacteria. *Trends Genet.*, 8(12):457–462, 1992.

[23] W. Ziebuhr, V. Krimmer, S. Rachid, I. Lößner, F. Götz, and Jorg Hacker. A novel mechanism of phase variation of virulence in *staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element is256. *Mol. Microbiol.*, 32(2):345–356, 1999.

[24] A.C. Wright, J.L. Powell J.B. Kaper, and J.G.Jr Morris. Identification of a group 1-like capsular polysaccharide operon for *vibrio vulnificus*. *Infect. Immun.*, 69(11):6893–6901, 2001.

[25] G. Neves, J. Zucker, M. Daly, and A. Chess. Stochastic yet biased expression of multiple dscam splice variants by individual cells. *Nature Genetics*, 36(3):240–246, 2004.

[26] M.D. Levi, C.J. Morton-Firth, W.N. Abouhamad, R.B. Bourret, and D. Bray. Origins of individual swimming behavior in bacteria. *Biophys. J.*, 74(1):175–181, 1998.

[27] J.L. Spudich and D.E.Jr. Koshland. Non-genetic individuality: chance in the single cell. *Nature*, 262(5568):467–471, 1976.

[28] H. Mayani, W. Dragowska, and P.M. Lansdorp. Lineage commitment in human hemopoiesis involves asymmetric cell division of multipotent progenitors and does not appear to be influenced by cytokines. *J. Cell. Physiol.*, 157(3):579–586, 1993.

[29] J.I. Elliott, R. Festenstein, M. Tolaini, and D. Kioussis. Random activation of a transgene under the control of a hybrid hcd2 locus control region/ig enhancer regulatory element. *EMBO Journal*, 14(3):575–584, 1995.

[30] R. Kemkemer, S. Schrank, W. Vogel, H. Gruler, and D. Kaufmann. Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro. *Proc. Natl Acad. Sci. USA*, 99(21):13783–13788, 2002.

[31] P.W. Sternberg and M.A. Felix. Evolution of cell lineage. *Curr. Opin. Genet. Dev.*, 7(4):543–550, 1997.

[32] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, 2000.

[33] T.M. Yi, Y. Huang, M.I. Simon, and J. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A.*, 97(9):4649–4653, 2000.

[34] D. Gonze, J. Halloy, and A. Goldbeter. Robustness of circadian rhythms with respect to molecular noise. *Proc Natl Acad Sci U S A.*, 99(2):673–678, 2002.

[35] J. Lewis. Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr Biol*, 13(16):1398–1408, 2003.

[36] D. Thieffry, A.M. Huerta, E. Pérez-Rueda, and J. Collado-Vides. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *escherichia coli*. *BioEssays*, 20(5):433–440, 1998.

[37] R. Lutz and H. Bujard. Independent and tight regulation of the transcriptional units in *escherichia coli* via the lacr/o, the tetr/o and arac/i1-i2 regulatory elements. *Nucleic Acids Res.*, 25(6):1203–1210, 1997.

[38] A. Løbner-Olesen. Distribution of minichromosomes in individual *escherichia coli* cells: implications for replication control. *EMBO J.*, 18(6):1712–1721, 1999.

[39] R.L. Vellanoweth and J.C. Rabinowitz. The influence of ribosome-binding-site elements on translational efficiency in bacillus subtilis and *escherichia coli* in vivo. *Mol. Microbiol.*, 6(9):1105–1114, 1992.

[40] T.B. Kepler and T.C. Elston. Stochasticity in transcriptional regulation: origins, consequences and mathematical representations. *Biophys. J.*, 81(6):3116–3136, 2001.

[41] N. Barkai and S. Leibler. Biological rhythms: Circadian clocks limited by noise. *Nature*, 403(6767):267–268, 2000.

[42] J. Paulsson, O.G. Berg, and M. Ehrenberg. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc. Natl Acad. Sci. USA*, 97(13):7148–7153, 2000.

[43] G. von Dassow, E. Meir, E.M. Munro, and G.M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, 2000.

[44] E.M. Ozbudak, M. Thattai, I. Kurtser, A.D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nature Genetics*, 31(1):69–73, 2002.

[45] M.B. Elowitz, A.J. Levine, E.D. Siggia, and P.S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[46] J. Paulsson. Summing up the noise in gene networks. *Nature*, 427(6973):415–418, 2004.

[47] W.J. Blake, M. KÆrn, C.R. Cantor, and J.J. Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633 – 637, 2003.

[48] J. Hasty, D. McMillen, F. Isaacs, and J.J. Collins. Computational studies of gene regulatory networks: in numero molecular biology. *Nature Rev. Genet*, 2(4):268–279, 2001.

[49] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in *escherichia coli*. *Nature*, 403(6767):339–342, 2000.

[50] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[51] J.E.Jr. Ferrell and E.M. Machleder. The biochemical basis of an all-or-none cell fate switch in *xenopus oocytes*. *Science*, 280(5365):895–898, 1998.

[52] U.S. Bhalla, P.T. Ram, and R. Iyengar. Map kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, 297(5583):1018–1023, 2002.

[53] J.R. Pomerening, E.D. Sontag, and J.E.Jr. Ferrell. Building a cell cycle oscillator: hysteresis and bistability in the activation of cdc2. *Nature Cell Biol.*, 5(4):346–351, 2003.

[54] W. Sha, J. Moore, K. Chen, A.D. Lassaletta, C.S. Yi, J.J. Tyson, and J.C. Sible. Hysteresis drives cell-cycle transitions in *xenopus laevis* egg extracts. *Proc. Natl Acad. Sci. USA*, 100(3):975–980, 2003.

[55] F.R. Cross, V. Archambault, M. Miller, and M. Klovstad. Testing a mathematical model of the yeast cell cycle. *Mol. Biol. Cell*, 13(1):52–70, 2002.

[56] A. Hernday, B.A. Braaten, and D. Low. The mechanism by which dna adenine methylase and papi activate the pap epigenetic switch. *Mol. Cell*, 12(4):947–957, 2003.

[57] A.R. Reynolds, C. Tischer, P.J. Verveer, O. Rocks, and P.I. Bastiaens. Egfr activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. *Nat. Cell Biol.*, 5(5):447–453, 2003.

[58] T.A. Blauwkamp and A.J. Ninfa. Physiological role of the glnk signal transduction protein of *escherichia coli*: survival of nitrogen starvation. *Mol. Microbiol.*, 46(1):203–214, 2002.

[59] D.A. Siegele and J.C. Hu. Gene expression from plasmids containing the arabad promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl Acad. Sci. USA*, 94(15):8168–8172, 1997.

[60] F.J. Isaacs, J. Hasty, C.R. Cantor, and J.J. Collins. Prediction and measurement of an augoregulatory genetic module. *Proc. Natl Acad. Sci. USA*, 100(13):7714–7719, 2003.

[61] A. Becskei, B. Seraphin, and L. Serrano. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.*, 20(10):2528–2535, 2001.

[62] J.E.Jr. Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.*, 14(2):140–148, 2002.

[63] J.S. Griffith. Mathematics of cellular control processes ii: Positive feedback to one gene. *J. Theor. Biol.*, 20(2):209–216, 1968.

[64] S.-K. Ma. *Modern Theory of Critical Phenomena.* Perseus Books, Reading, Massachusetts, 1976.

[65] S.H. Strogatz. *Nonlinear Dynamics and Chaos.* Perseus Books, Reading, Massachusetts, 1994.

[66] B. Muller-Hill. *The Lac Operon: A Short History of a Genetic Paradigm.* Walter de Gruyter, Berlin, 1996.

[67] A. Novick and M. Weiner. Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA*, 43, 1957.

[68] M. Cohn and K. Horibata. Inhibition by glucose of the induced synthesis of the -galactoside-enzyme system of *escherichia coli*: Analysis of maintenance. *J. Bacteriol.*, 78, 1959.

[69] J. Stulke and W. Hillen. Carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.*, 2(2):195–201, 1999.

[70] Y. Setty, A.E. Mayo, M.G. Surette, and U. Alon. Detailed map of a cis-regulatory input function. *Proc. Natl Acad. Sci. USA*, 100(13):7702, 2003.

[71] J.J. Tyson and H.G. Othmer. The dynamics of feedback control circuits in biochemical pathways. *Prog. Theor. Biol.*, 5, 1978.

[72] M. Louis and A. Becskei. Binary and graded responses in gene networks. *Science STKE [online]*, 2002.

[73] S.R. Biggar and G.R. Crabtree. Cell signaling can direct either binary or graded transcriptional responses. *EMBO J.*, 20(12):3167–3176, 2001.

[74] B. Nobelmann and J.W. Lengeler. Molecular analysis of the gat genes from *escherichia coli* and of their roles in galactitiol transport and metabolism. *J. Bacteriol.*, 178(23):6790–6795, 1996.

[75] G. Yagil and E. Yagil. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys J.*, 11(1):11–27, 1971.

[76] S. Oehler, E.R. Eismann H. Kramer, and B. Müller-Hill. The three operators of the lac operon cooperate in repression. *EMBO J.*, 9(4):973–979, 1990.

[77] J.D. Chung and G. Stephanopoulos. On physiological multiplicity and population heterogeneity of biological systems. *Chem. Eng. Sci.*, 51, 1996.

[78] E.M. Ozbudak, M. Thattai, H.N. Lim, B.I. Shraiman, and A. van Oudenaarden. Multistability in the lactose utilization network of *escherichia coli*. *Nature*, 427(6976):737–740, 2004.

[79] M. Thattai and B. Shraiman. Metabolic switching in the sugar phosphotransferase system of *escherichia coli*. *Biophys. J.*, 85(2):744–754, 2003.

[80] M.R. Atkinson, M.A. Savageau, J.T. Myers, and A.J. Ninfa. Development of genetic toggle circuitry exhibiting toggle switch or oscillatory behavior in *escherichia coli*. *Cell*, 113(5):597–607, 2003.

[81] P. Smolen, D.A. Baxter, and J.H. Byrne. Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. *Am. J. Physiol.*, 274(2), 1998.

[82] J. Sambrook, E.F. Fritsch, and T. Maniatis. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.

[83] D. Dubnau and R. Davidoff-Abelson. Fate of transforming dna following uptake by competent *bacillus subtilis.* i. formation and properties of the donor-recipient complex. *J. Mol. Biol.*, 56(2):209–221, 1971.

[84] H. Shimotsu and D.J. Henner. Construction of a single-copy integration vector and its use in analysis of regulation of the trp operon of *bacillus subtilis. Gene*, 43(1), 1986.

[85] D. Boyd, D.S. Weiss, J.C. Chen, and J. Beckwith. Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda inch, a simple *escherichia coli* plasmid-chromosome shuttle system. *J. Bacteriol.*, 182(3):842–847, 2000.

[86] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[87] L.H. Hansen, S. Knudsen, and S.J. Sorensen. The effect of the lacy gene on the induction of iptg inducible promoters, studied in *escherichia coli* and *pseudomonas fluorescens. Curr Microbiol.*, 36(6):341–347, 1998.