COMPUTATIONAL INVESTIGATIONS OF NOISE-MEDIATED CELL POPULATION DYNAMICS

Daniel A. Charlebois

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements

for the Doctor of Philosophy degree in Physics

Department of Physics

Faculty of Science

University of Ottawa

À mon père, André Charlebois, mon plus grand partisan.

SUMMARY

Fluctuations, or "noise", can play a key role in determining the behaviour of living systems. The molecular-level fluctuations that occur in genetic networks are of particular importance. Here, noisy gene expression can result in genetically identical cells displaying significant variation in phenotype, even in identical environments. This variation can act as a basis for natural selection and provide a fitness benefit to cell populations under stress.

This thesis focuses on the development of new conceptual knowledge about how gene expression noise and gene network topology influence drug resistance, as well as new simulation techniques to better understand cell population dynamics. Network topology may at first seem disconnected from expression noise, but genes in a network regulate each other through their expression products. The topology of a genetic network can thus amplify or attenuate noisy inputs from the environment and influence the expression characteristics of genes serving as outputs to the network.

The main body of the thesis consists of five chapters:

1. A published review article on the physical basis of cellular individuality.

2. A published article presenting a novel method for simulating the dynamics of cell populations.

3. A chapter on modeling and simulating replicative aging and competition

using an object-oriented framework.

4. A published research article establishing that noise in gene expression can facilitate adaptation and drug resistance independent of mutation.

5. An article submitted for publication demonstrating that gene network topology can affect the development of drug resistance.

These chapters are preceded by a comprehensive introduction that covers essential concepts and theories relevant to the work presented.

SOMMAIRE

Les fluctuations, ou le << bruit >>, peuvent être-utile pour les systèmes biologiques. En particulier, les fluctuations moléculaires dans les réseaux génétiques peuvent être d'importance particulière. Ici, la stochasticité dans l'expression d'un gène dans des cellules identiques peut introduire une source de variabilité phénotypique dans la population, même dans des environnements identiques. Cette variation peut fournir une base pour la selection naturelle et offrir un avantage reproductif aux populations cellulaires en état de stress.

Cette thèse vous présente les travaux de recherche sur l'effet de l'expression stochastique d'un gène et de la topologie d'un réseau génétique dans le développement du phénomène de résistance aux drogues, ainsi que de nouvelles méthodes pour simuler les dynamiques de populations cellulaires. La topologie d'un réseau et la stochasticité dans l'expression d'un gène peuvent sembler être déconnectés, mais elles ont un dénominateur commun : les gènes dans un réseau génétique se contrôle les uns les autres et la topologie du réseau peut amplifier ou atténuer le bruit dans les gènes servant comme source de sortie du réseau.

La thèse contient cinq chapitres:

1. Un article de revue publié sur la base physique de l'individualité cellulaire.

2. Un article publié qui présente une méthode innovatrice pour simuler les dynamiques de populations cellulaires.

3. Un chapitre sur la modélisation et simulation de l'effet de la sénescence et de la compétition sur les dynamiques des populations cellulaires.

4. Un article de recherche publié qui établis que les fluctuations de courte durée dans l'expression d'un gène peuvent assurer la survie à long terme d'une population résistante à une certaine drogue indépendamment des mutations génétiques.

5. Un article soumis pour publication qui démontre que la topologie d'un réseau génétique peut fournir la base du développement de la résistance aux drogues.

Ces chapitres sont précédés d'une introduction extensive couvrant les théories et les concepts essentiels à leur compréhension.

STATEMENT OF ORIGINALITY

I hereby certify that, to the best of my knowledge, the research presented in this thesis is original. I have defined the questions, conducted the studies, and wrote the papers under the supervision of Dr. Mads Kærn. Mr. Nezar Abdennur, a co-author of the *Physical Review Letters* article, was a master's student who, under my supervision, carried out the simulations of the burst model of gene expression. Mr. Abdennur also developed the object-oriented framework described in Chapter 4, together with its Python implementation, that I used to perform the simulations in this chapter. The models in Chapter 4 were co-developed by Mr. Abdennur and myself. Otherwise, I developed the algorithms, constructed the various mathematical models, and ran the simulations used in this thesis to investigate the effects of gene expression noise and network topology on population fitness, making new predictions, that are at the time of writing this thesis, being tested experimentally.

ACKNOWLEDGMENTS

I thank my supervisor, Dr. Mads Kærn, for his guidance over the past few years; our discussions and his honest appraisal of my work has been invaluable to me. I also acknowledge my former supervisors, Dr. Stuart Kauffman and Dr. André Ribeiro, for taking a chance on me as a young undergraduate student, inspiring me, and introducing me to the wonderful world of research. I should also mention Dr. Gregory Arkos, my first year physics professor, who first encouraged me to pursue studies in the physical sciences. Finally, I am grateful for my lab mates, who over the years, have made this experience a very rewarding one, especially Daniel Jedrysiak for our time, however brief, in the wet lab, and Nezar Abdennur for insightful discussions and our work together. I would also like to acknowledge Dr. Gábor Bálazsi, Hilary Phenix, and Afnan Azizi for helpful discussions. The quality of thesis was improved by the editing of Nezar Abdennur, Hilary Phenix, Afnan Azizi, and Stephanie Charlebois.

On a more personal note, I thank my family and friends for their encouragement and support. In particular, I am forever indebted to my wife Teresa Charlebois, who supported me in every way possible, and for whom a brief mention here does not do justice.

PREFACE

Science students, outside the physical sciences, are usually not taught much about noise other than it degrades the quality of a signal and thus is a hindrance to experimental design and measurement. Noise is in this context viewed as something negative, often external to the system and requiring filtration and removal. Fortunately, in physics, we also learn the fundamentals of noise, typically in courses on statistical mechanics and numerical methods for stochastic processes. These fundamental concepts are applicable to a broad range of systems, including living cells.

In the biological sciences, it is often assumed that populations of genetically identical cells of the same type are homogeneous and that any difference between these cells is either negligible or can be attributed to experimental error. However, the role of gene expression noise in generating beneficial cell-to-cell variability is beginning to attract considerable attention, largely as a result of the influx of physical scientists and mathematicians to areas such as biological physics, quantitative biology, and systems biology. As will be demonstrated in this thesis, this line of research is shedding new light on an issue of central importance to the treatment of disease: drug resistance. My work over the past several years adds new knowledge to our understanding of the development of drug resistance. This thesis is a contribution to the body of work that is transforming biology into a more quantitative discipline.

Contents

Table of Contents **x**

1 Introduction **1**
1.1 Gene Expression . . . 1
1.1.1 Central "Dogma" of Molecular Biology . . . 1
1.1.2 Stochastic Gene Expression . . . 2
1.1.3 Consequences of Noisy Expression . . . 4
1.1.4 Measuring Gene Expression . . . 6
1.1.5 Gene Regulatory Networks . . . 8
1.2 Modeling Gene Expression . . . 12
1.2.1 Ordinary Differential Equations . . . 13
1.2.2 Langevin Approach . . . 15
1.2.3 Chemical Master Equation . . . 18
1.3 Modeling Population Dynamics . . . 24
1.3.1 Cellular Growth and Division . . . 24
1.3.2 Constant-Number Monte Carlo . . . 25
1.3.3 Population Dynamics Algorithm . . . 27
1.3.4 Noise and Fitness . . . 29
1.4 Drug Resistance . . . 34
1.4.1 Mechanisms of Drug Resistance . . . 34
1.4.2 Genetic Basis of Drug Resistance . . . 35
1.4.3 Epigenetic Basis of Drug Resistance . . . 37
1.5 Thesis Overview . . . 42
1.6 References . . . 43

43

2 What all the Noise is About: The Physical Basis of Cellular Individuality **61**
2.1 A historical perspective . . . 62
2.2 Noise and biological systems . . . 63
2.3 Current research . . . 63
2.4 Conclusion . . . 65

3 An Accelerated Method for Simulating Population Dynamics 67
3.1 Introduction . . . 68
3.2 Algorithm . . . 69
3.2.1 Accelerated method for simulating population dynamics . . . 69
3.2.2 SSA . . . 72
3.2.3 Constant-number Monte Carlo method . . . 73
3.3 Numerical results and discussion . . . 73
3.4 Univariate model . . . 74
3.5 Multivariate model . . . 76
3.6 Environmental stress . . . 78
3.7 Parameter scans . . . 79
3.8 Conclusion . . . 80

4 Modeling and Simulating Replicative Aging and Cell Competition 84
4.1 Object-Oriented Framework for Simulating Heterogeneous Cell Populations . . . 84
4.2 Volume and Age-Dependent Growth in Yeast . . . 87
4.2.1 Background . . . 87
4.2.2 Model . . . 88
4.2.3 Results and Discussion . . . 89
4.3 Resource Competition . . . 91
4.3.1 Background . . . 91
4.3.2 Model . . . 91
4.3.3 Results and Discussion . . . 93
4.4 References . . . 95

5 Gene Expression Noise Facilitates Adaptation and Drug Resistance 97

6 Coherent Feedforward Transcriptional Regulatory Motifs Enhance Drug Resistance 103
6.1 Introduction . . . 104
6.2 Minimal Model . . . 105
6.2.1 FFLs accelerate and prolong transcriptional responses . . . 106
6.2.2 FFLs provide stable high expression in fluctuating environments 107
6.2.3 FFLs increase population heterogeneity and mixing times . . . 108
6.2.4 FFLs enhance drug resistance . . . 108
6.3 PDR5 Transcriptional Network Model . . . 109
6.4 Conclusion . . . 110

7 Other contributions 113

8 Conclusion **117**
8.1 Population Simulation Algorithms 117
8.2 Epigenetic Drug Resistance 120
8.3 Final Thoughts 123

Appendices **125**

Appendix A: Enlarged Figures - Chapter 3 **126**

Appendix B: Supplementary Figure - Chapter 4 **131**

Appendix C: Enlarged Figures - Chapter 6 **133**

Chapter 1

Introduction

All models are wrong, but some are useful. - George E.P. Box

1.1 Gene Expression

1.1.1 Central "Dogma" of Molecular Biology

The central dogma of molecular biology deals with the transfer of sequential information encoded in DNA, RNA, and protein molecules [31,32]. It is a negative statement asserting that information from a protein can never be transferred back to either protein or nucleic acid and generally flows from DNA to RNA, and from RNA to protein (Fig. 1.1).

More specifically, gene expression is the process by which a gene, a specific sequence of nucleotides in the DNA, is transcribed to produce messenger RNA (mRNA) and mRNA is translated into protein. To initiate transcription, an RNA polymerase (RNAp) must recognize and bind to the promoter region, a regulatory region of DNA that precedes the gene [5, 104]. Promoters have regulatory sites to which regulatory proteins (transcription factors) can bind to either activate or repress gene transcrip-

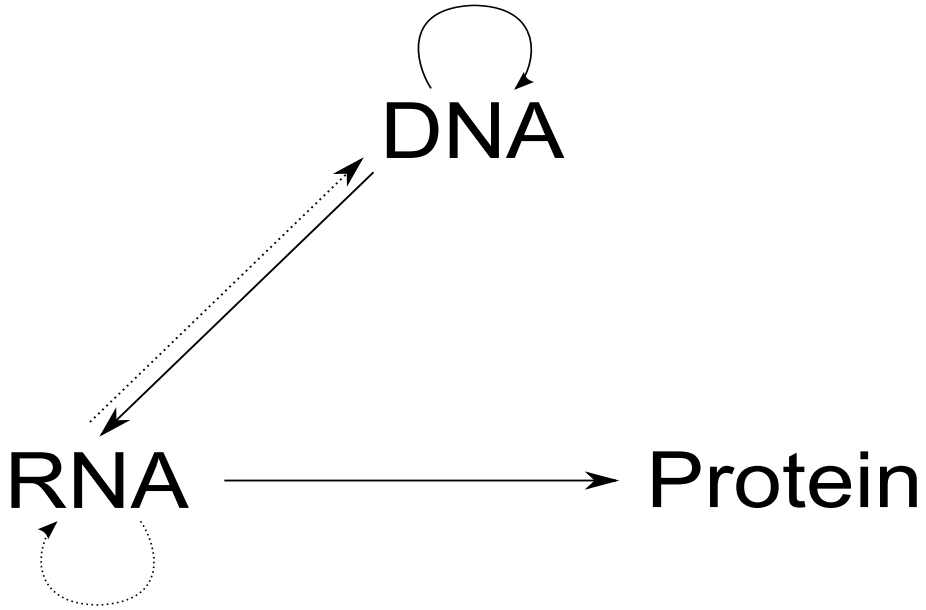


Figure 1.1 Central dogma of molecular biology. Solid arrows denote general information transfers which occur in all cells. Dotted arrows show special transfers that may occur in special circumstances (e.g., reverse transcription and RNA silencing).

tion. The promoter is followed by the coding sequence, which is transcribed by the RNAp into an mRNA molecule. Transcription stops when the RNAp reaches a termination sequence and unbinds from the DNA. Next, translation follows in which ribosomes read the mRNA sequence, and for each codon (a three-nucleotide sequence of DNA or mRNA), a corresponding amino acid is added to a polypeptide chain (a protein). The protein becomes capable of performing specific tasks after post-translational modification.

1.1.2 Stochastic Gene Expression

The expression of gene products is a noisy or stochastic process [40,65,72,98,107,114, 119,122,129,146]. Stochastic gene expression is ultimately a manifestation of the thermal nature of chemical reactions, which constitute probabilistic molecular events [50].

Specific physical reasons exist as to why the dynamics of chemically reacting systems are non-deterministic. Chemical systems are often in contact with a "heat bath" which keeps the system in thermal equilibrium at some temperature. When the system is in thermal equilibrium, the molecules are distributed randomly and uniformly throughout the containing volume. These molecules move with erratic motions as a result of the uneven bombardment of underlying fluid molecules (Brownian motion), which themselves have essentially random motions due to thermal fluctuations from the environment and collisions with other fluid molecules (self-diffusion). These erratic motions result in erratic collisions, and this results in the non-deterministic timing of individual reactions and an inherently noisy time evolution of molecular population levels [49, 51, 53].

The term "noise" when used in the context of gene expression is a broad reference to the observed variation in mRNA or protein content among apparently identical cells exposed to the same environment [42]. Noise-induced phenotypic heterogeneity has been observed in prokaryotes (e.g., bacteria [40]) and eukaryotes (e.g., yeast [107] and mouse cells [22]). This noise can be divided up into extrinsic and intrinsic components. Intrinsic gene expression noise refers to variation generated in the multistep processes that lead to the synthesis and degradation of mRNA and protein molecules [18,40,54,65,72,112,119,121] (Fig. 1.2). Extrinsic gene expression noise can be generally defined as fluctuations and variability that arise in a system due to disturbances originating from its environment, and therefore depends on how the system of interest is defined [133]. Extrinsic gene expression noise arises from several sources including: asymmetric and stochastic partitioning at cell division [54,61,106,126,168], cell age [121, 142, 157, 169], stage of the cellular replication cycle [107], fluctuations in the abundance of ribosomes [147], and variability in upstream signal transduction [115, 157].

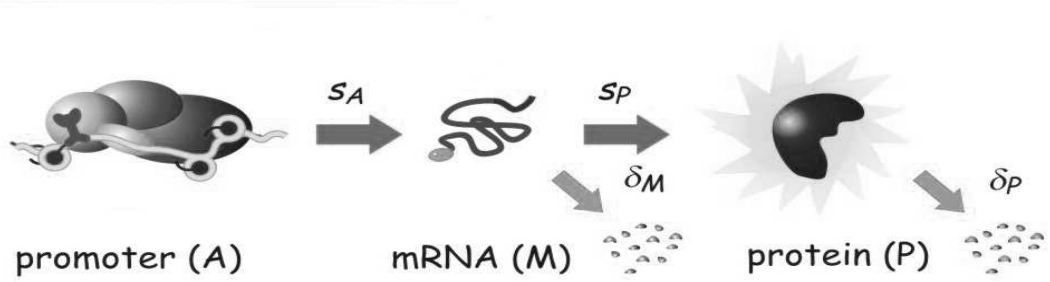


Figure 1.2 A phenomenological model for the expression of a single gene. mRNA (M) is synthesized from a single gene (represented by its promoter (A)) at a rate s_A, protein (P) is synthesized from an mRNA template at a rate s_P, and mRNA and protein molecules decay at rates δ_M and δ_P, respectively. All steps are modeled as first-order reactions with the indicated rate constants (units of inverse time) associated with these steps. Figure reproduced by permission from American Institute of Physics: Chaos 16: 026107, copyright 2006.

Several noise measures are used to quantify the degree of heterogeneity in gene expression, or noise η [65, 149]. The most common is the relative deviation from the average expression, which is determined by the ratio of the standard deviation σ to the mean μ. Another measure of noise, known as the fano factor ($\phi = \sigma^2/\mu$), can be used to uncover trends that might otherwise be obscured by the characteristic $1/\sqrt{\mu}$ scaling of the noise as described by Poisson statistics and observed experimentally in living systems (e.g., [148]).

1.1.3 Consequences of Noisy Expression

In molecular biology it is often assumed that clonal cell populations are uniform [58]. As a result, not much attention has been paid to the effects of noise-induced cell-to-cell variability. Variation in gene expression can be detrimental to cell function, necessitating minimization, as fluctuations in protein levels may disrupt intracellular signaling and cellular regulation [3, 15, 43]. However, noise-induced variability can

also be beneficial by providing the phenotypic diversity for natural selection to act upon [17, 42, 170, 171]. Noise-generated variability can have functionally significant consequences in response to perturbations such as drug treatment [21, 25, 29, 58, 171], and provide a fitness advantage in fluctuating environments [4, 42, 78, 156].

Evidence suggests that protein noise levels have been selected to reflect the costs and benefits of this variation [107]. Blake *et al.* established that noisy gene expression can be advantageous under conditions of high stress [17]. The authors genetically engineered two nearly identical *Saccharomyces cerevisiae* (budding yeast) strains such that one strain had relatively high noise and the other relatively low noise in the expression of a drug resistance gene. Next, they exposed the yeast populations to increasing concentrations of the antibiotic Zeocin. The authors found that for high levels of Zeocin the high-noise strain had a higher viability. Conversely, for low levels of Zeocin the low-noise strain had higher viability. Blake *et al.* attributed the differential impact of added noise to a change in the relative fraction of surviving cells at different levels of stress. While a high-noise population will have a higher number of cells above the protein threshold necessary for survival at high stress levels, the same applies for a low-noise population under a low level of stress (Fig.1.3).

Acar *et al.* found that cell populations can enhance their fitness by allowing individual cells to switch stochastically between phenotypes [4]. In this study, a budding yeast strain was engineered to randomly transition between two phenotypes, ON and OFF, characterized respectively by high- or low-expression of a gene encoding the Ura3 enzyme which is necessary for uracil synthesis. The drug 5-FOA is converted into a toxin (5-fluorouracil) by the Ura3 enzyme inside the cell which causes stress and reduces fitness. This allowed the researchers to design each phenotype to have a growth advantage over the other in one of the two environments. In the first environment which lacked uracil and FOA, cells with the ON phenotype had

an advantage. In the second environment, which contained uracil and FOA, cells with the OFF phenotype had the advantage. The authors changed environmental conditions by adding or removing uracil and FOA from the media. It was found that fast switching populations outgrew slow switching populations when the environment fluctuated rapidly, whereas slow-switching phenotypes outgrew fast switchers when the environment rarely changed. These results suggest that tuning inter-phenotypic switching rates may constitute a simple strategy to cope with fluctuating environments. Thus an isogenic population can improve its fitness by bet-hedging, namely optimizing phenotypic diversity such that, at any given time, an optimal fraction of the population is prepared for an unexpected environmental stress. This phenotypic diversity is introduced naturally through the stochastic process of gene expression.

1.1.4 Measuring Gene Expression

Classic mRNA and protein experiments (e.g., immunoblots or microarrays) involve disrupting the cell membrane and releasing the contents of all the cells in the sample [58]. The trait X of interest, here the abundance of a specific mRNA or protein, is isolated and its average level in the sample is measured. As useful as these techniques are, they do not provide information at a single-cell level which is essential for studying the effects of noise-generated cell-to-cell variability.

Flow cytometry is one commonly employed technique for measuring the biochemical and physical properties of a cell population at a single-cell resolution. This technique, however, cannot be used to monitor temporal changes within an individual cell as it generates a population "snapshot" by a recording specific properties of each cell in the population once at a specific point in time. Gene expression within a population of a single cell type can be measured experimentally using this technique. Specifically, one can obtain a histogram of a given protein in individual cells across a

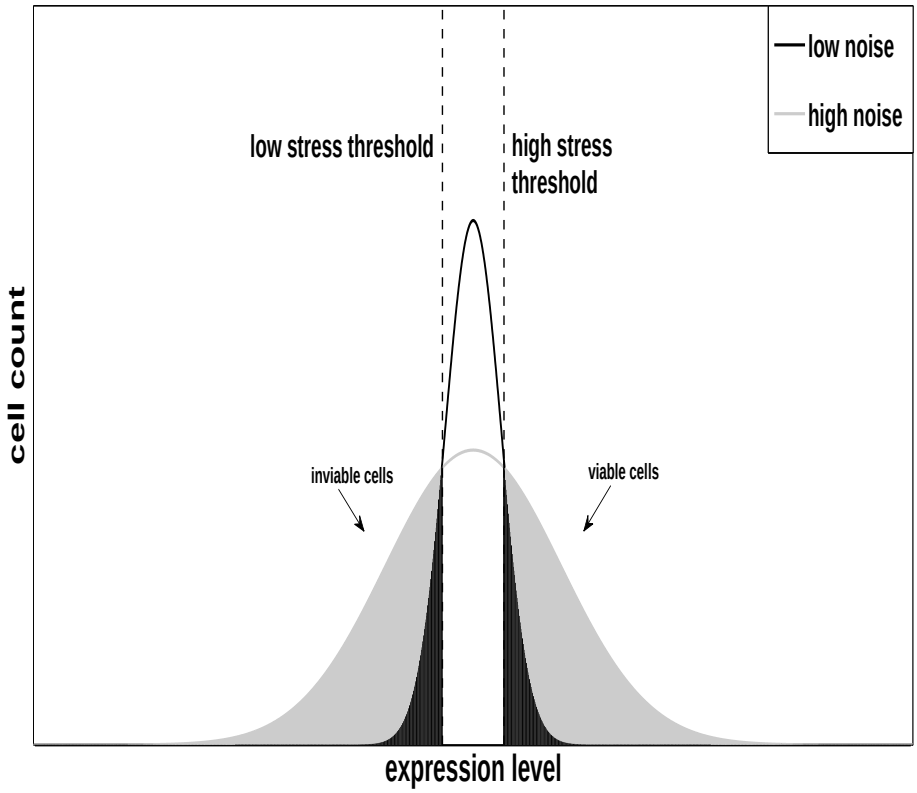


Figure 1.3 Gene expression noise confers survival in clonal cell populations. Schematic illustration of protein expression distributions for low- and high-noise populations. The fraction of cells that express above (viable cells) and below (inviable cells) a stress threshold depends on the position of the threshold and the level of noise in the population. Figure reproduced by permission from Blackwell Publishing Ltd: Molecular Microbiology 71: 13331340, copyright 2009.

large cell population. Variability in gene expression can be observed within the peak of a protein histogram generated from flow-cytometry data. For a genetically identical (clonal) cell population the abundance of the protein in the cells with the lowest and highest expression level typically differs by three or more orders of magnitude. This spread far exceeds signal measure error which suggests that gene expression noise is not simply a measurement artifact [22].

A microfluidics device can be assembled together with a fluorescence microscope and a charge-coupled device (CCD) camera to perform fluorescence live cell imaging experiments (e.g., [159]). In a microfluidics device, cells are fixed between a membrane and the coverslip which constrains them to grow as a 2D sheet. This facilitates imaging while simultaneously allowing for fluid exchange. Additionally, this device provides a controlled environment for drug resistance experiments as the media is constantly flowing over the cells, ensuring that all cells are exposed to the same nutrient and drug concentrations. This setup is very practical for conducting drug resistance experiments as the cells do not float around during media changes or when the media is stirred so as to maintain uniform experimental conditions. This technique monitors the dynamics of trait X, and the resulting time series delivers kinetic information on the temporal structure of the fluctuations of gene expression that evade flow cytometry and other techniques. This technique also allows the tracking of cell fate history and the construction of cell lineages.

1.1.5 Gene Regulatory Networks

In one of the most highly cited papers in physics, Barabási and Albert [12] proposed that a general feature of large networks is that connectivity of the nodes follows a

scale-free power law distribution

$$p(k) \propto k^{-\gamma}, \tag{1.1}$$

where k is the number of inputs and γ is the degree exponent which describes the importance of hubs (a node with many connections) in the network. This distribution follows a straight line on logarithmic scales. The term "scale-free" indicates the absence of a typical node in the network that can be used to characterize the connectivity of the rest of the nodes. This feature was found to be a consequence of a network expanding continuously by the addition of new nodes and new edges attaching preferentially to sites that were already highly connected [12]. Most networks within the cell, including metabolic, protein-protein, and genetic networks, approximate a scale-free topology (see [13] for a review).

Stuart Kauffman first modeled a gene in 1969 as a binary (ON-OFF) device and the genome as a network of N interconnected genes [69, 70]. In this model, each gene is described by k and one of among the 2^{2^k} logical (boolean) functions on its k inputs. The output of the network is the gene expression profile. In these studies, Kauffman was considering random networks so k and the assignment of the boolean function were random variables. A gene network state (one of 2^N states) was defined as a list of the present value (0 or 1, representing OFF or ON, respectively) for each of the N genes. The subsequent state of the network is determined by the network structure and the present state. Several different models have since been employed to extend the boolean network model and more realistically describe the dynamics of gene expression and regulation (e.g., [124, 127, 133, 146]). Some of these modeling approaches were used in the works presented in Chapters 2.4-6.4, and are described in Section 1.2.

A genetic regulatory network (GRN) is broadly defined as an ensemble of molecules

and interactions that control gene expression and thereby regulate cellular dynamics [23]. Gene expression is regulated by transcription factors (TFs), which are themselves proteins encoded by genes [5]. Hence, the production (and activity) of TFs may be regulated by other TFs. It is this set of physical protein-protein and protein-DNA interactions that forms a transcriptional regulatory network. The inputs to transcriptional networks are signals that transmit information about the cellular or extracellular environment. The input signals can activate biochemical signal-transduction pathways, resulting in a chemical modification of a TF. In other cases, another molecule may bind the TF directly altering its activity.

GRNs are comprised of simple recurring patterns of interconnections known as network motifs, including feedback loops, cascades, and feedforward loops (Fig. 1.4) [5, 103], commonly found in different species (e.g., [80,103,136]). These motifs are found at frequencies much higher than those found in randomized networks and can be considered as building blocks necessary to assemble entities of more complex functionality.

GRN topology can influence the properties of gene expression noise. For example, a transcriptional cascade (Fig. 1.4b) can significantly increase the noise in the expression of a downstream gene by transmitting variability in the expression of upstream genes [27,115,126]. Cascades can also be designed to produce output that is less noisy than the input [150]. Positive feedback (Fig. 1.4a) increases noise by amplifying fluctuations about a mean and can result in bistability [27,111]. The latter occurs in gene expression when the rate of positive regulation is strong compared to the degradation/dilution rate of the expression products [5]. Negative feedback (Fig. 1.4a) on the other hand reduces noise by attenuating fluctuations about a mean [15, 27, 39, 138]. Feedforward loops (Fig. 1.4b) may similarly filter out fluctuations in the input signal [5]. For example, a coherent feedforward loop with OR logic at the output buffers

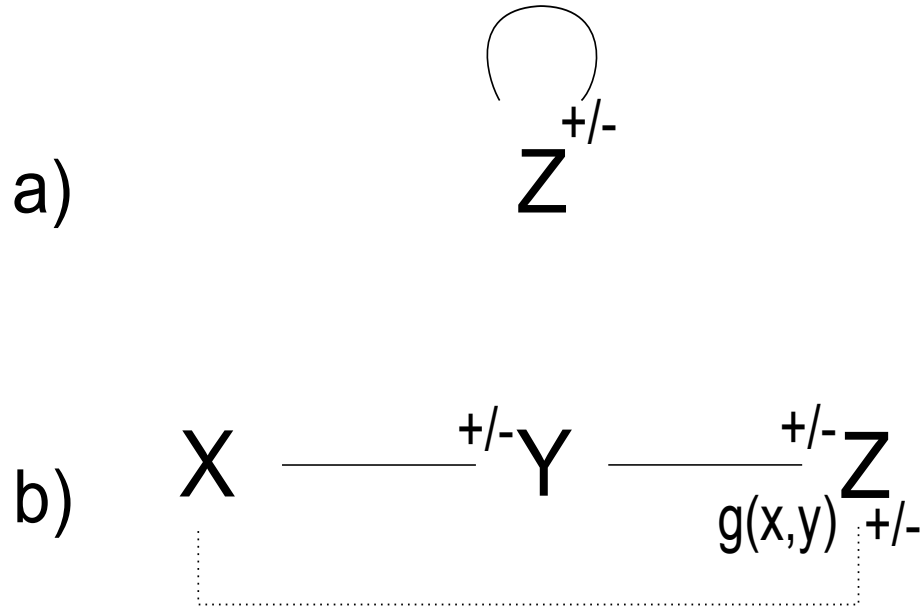


Figure 1.4 Common gene regulatory network motifs. (a) Autoregulation. (b) Feedforward loop, or if dotted line is absent, a cascade. Activation or repression are respectively denoted by + and - at the inputs and g(x,y) is the gate function which depends on the logic governing the inputs to Z (e.g., OR, AND, or, SUM) [99]. For example, if all inputs are + in (b) then the schematic depicts a coherent feedforward network.

against OFF fluctuations in the input signal and a coherent feedforward loop with AND logic at the output buffers ON fluctuations in the input signal.

In Chapter 6, we use mathematical modeling to study the pleiotropic (multiple) drug resistance (PDR) network in budding yeast (see also Section 1.4.3), which is interesting because it consists of a positive feedback loop embedded in a coherent feedforward loop (Fig. 1.4b). The increased relaxation time (defined as the time for the level of gene expression to fall to 50% of the steady-state value when the input to the network is turned ON [5]) and increased noise of this network relative to other network motifs resulting from the same genes [5, 66, 134], led me to hypothesize that the structure of the PDR5 transcriptional network is not random but rather evolved due to the fitness advantage it confers yeast in stressful environments.

1.2 Modeling Gene Expression

The information that a gene encodes is processed by the machinery of the cell to execute the instructions it contains. Understanding how this information is produced, processed, and propagated is vital for understanding cellular behaviour [27]. The gene expression process can be modeled at multiple scales, from detailed physical descriptions of molecular interactions to relatively simple phenomenological representations. A phenomenological model of this process is shown in Fig. 1.2. Although this depiction is a drastic simplification of the gene expression process (see e.g., [97,124,127,146] for more complex models of gene expression), it captures the essential features, namely the synthesis of mRNA from a single gene, the synthesis of protein from an mRNA template, and the decay of mRNA and protein molecules [133].

In this Section, we present the approximate ordinary differential equation (ODE) and Langevin approaches, and the exact chemical master equation (CME) formalism, together with the corresponding numerical methods, to model the gene expression process shown in Fig. 1.2. Each method has advantages and disadvantages and thus are appropriate under different circumstances. Obtaining an analytical and numerical solution to a system of ODEs is relatively straightforward. However, this approach is only valid when molecular species concentrations are high, and even in this case, it only captures averages. The Langevin approach incorporates additive noise terms into a system of ODEs to capture randomly fluctuating concentrations of chemical species. Analytically, this case is described using a Fokker-Planck equation, and although more challenging than for ODEs, can for simple cases be solved exactly to obtain the moments of the distribution. Numerical methods to solve Langevin equations involve random numbers and are more costly in terms of computational resources than corresponding numerical ODE methods. The CME approach describes

the time evolution of the probability distribution of every chemical species in the system. Since the chemical species inside the cell appear in integer amounts and are often present at low numbers, the CME is more exact and appropriate than the aforementioned methods which describe concentrations. The CME is generally intractable analytically and numerically. The *de facto* standard is to simulate each reaction in the system using a stochastic simulation algorithm (SSA). This approach is thus more computationally intensive than the other methods considered in this thesis.

1.2.1 Ordinary Differential Equations

Traditionally, the time evolution of a chemical system has been modeled as a deterministic process using a set of ordinary differential equations (ODEs). This approach is based on the empirical law of mass action, which provides a relation between reaction rates and molecular concentrations [153]. Namely, the instantaneous rate of a reaction is directly proportional to the product of the reactant concentrations (which is in turn proportional to mass) raised to the power of their stoichiometric coefficients. In the deterministic description of the model shown in Fig. 1.2, the cellular mRNA and protein concentrations ($[M]$ and $[P]$, respectively) are governed by the rate equations

$$\frac{d[M]}{dt} = s_A - \delta_M[M], \tag{1.2}$$

$$\frac{d[P]}{dt} = s_P[M] - \delta_P[P], \tag{1.3}$$

where the terms $\delta_M[M]$ and $\delta_P[P]$ are the degradation rates for mRNA and protein, respectively; the term $s_p[M]$ is the rate of protein synthesis, and mRNA production occurs at a constant rate (s_A) due to the presence of a single promoter. The steady-

state concentrations found by setting Eqs. (1.2) and (1.3) equal to zero are

$$[M^s] = \frac{s_A}{\delta_M}, \tag{1.4}$$

$$[P^s] = \frac{[M]s_P}{\delta_P} = \frac{s_A s_P}{\delta_M \delta_P}, \tag{1.5}$$

and are related to the average steady-state number of M and P by the cell volume V.

Note that the deterministic mathematical model (Eqs. (1.2) and (1.3)) was obtained by treating each step as a first-order chemical reaction and applying the law of mass action. The law of mass action was developed to describe chemical reactions under conditions where the number of each chemical species is so large that concentrations can be approximated as continuous variables without introducing a significant error [133].

Deterministic ODEs can in certain cases accurately describe the mean of gene expression (Fig. 1.5a). They cannot, however, capture the fluctuations about the mean and therefore the resulting probability distributions that are available when using a stochastic approach (Fig. 1.5b). Furthermore, when reaction rates depend nonlinearly on randomly fluctuating components, macroscopic rate equations may be far off the mark even in their estimates of averages [113].

In order for the deterministic approach to provide a valid approximation of an exact stochastic description (see Section 1.2.3), the system size must be large in terms of the numbers of each species and the system volume (e.g., for the model we are considering here s_A and V must be large so that the number of expressed mRNA and protein molecules is high with the ratio s_A/V remaining constant) [65]. When this condition is not satisfied, the effects of molecular noise can be significant. The high molecular number condition is not satisfied for gene expression, due to low copy number of genes, mRNAs, and transcription factors within the cell [165].

Runge-Kutta Method

There are several numerical techniques available to solve an ODE [92]. The fourth order Runge Kutta (RK4) method is the best practical compromise between accuracy (error at each step on the order of $O(h^5)$ and total accumulated error of order $O(h^4)$) and computational effort (four evaluations of f per time step). Let an initial value problem be specified by $y(t_0) = y_0$ and $\dot{y}(t) = f(t, y)$, where y is a scalar or a vector. The RK4 algorithm is as follows

$$\begin{aligned} k_1 &= f(t_n, y_n), \\ k_2 &= f((t_n + h/2), y_n + hk_1/2), \\ k_3 &= f((t_n + h/2), y_n + hk_2/2), \\ k_4 &= f(t_n + h, y_n + hk_3), \\ t_{n+1} &= t_n + h, \\ y_{n+1} &= y_n + h(k_1 + 2k_2 + 2k_3 + k_4)/6, \end{aligned} \tag{1.6}$$

where n is a non-negative integer and h is the step-size. To solve Eqs. (1.2) and (1.3), as shown in Fig. 1.5, $y(t_0) = [0, 0]$ and $f(t, [M, P]) = [s_A - M\delta_M, s_P M - P\delta_P]$.

1.2.2 Langevin Approach

Langevin theory can be used to model intrinsic and extrinsic noise in gene expression (see [134] and Chapter 5; a historical introduction to the Langevin equation is provided in Chapter 2). A Langevin approach assumes that fluctuations do not drive the system far from steady-state, and thus, in the case considered here, is only strictly valid when the number of molecules is sufficiently large [145]. However, this method often works well for very small numbers of molecules, and at the very least, almost always provides useful qualitative information about the system. To model intrinsic

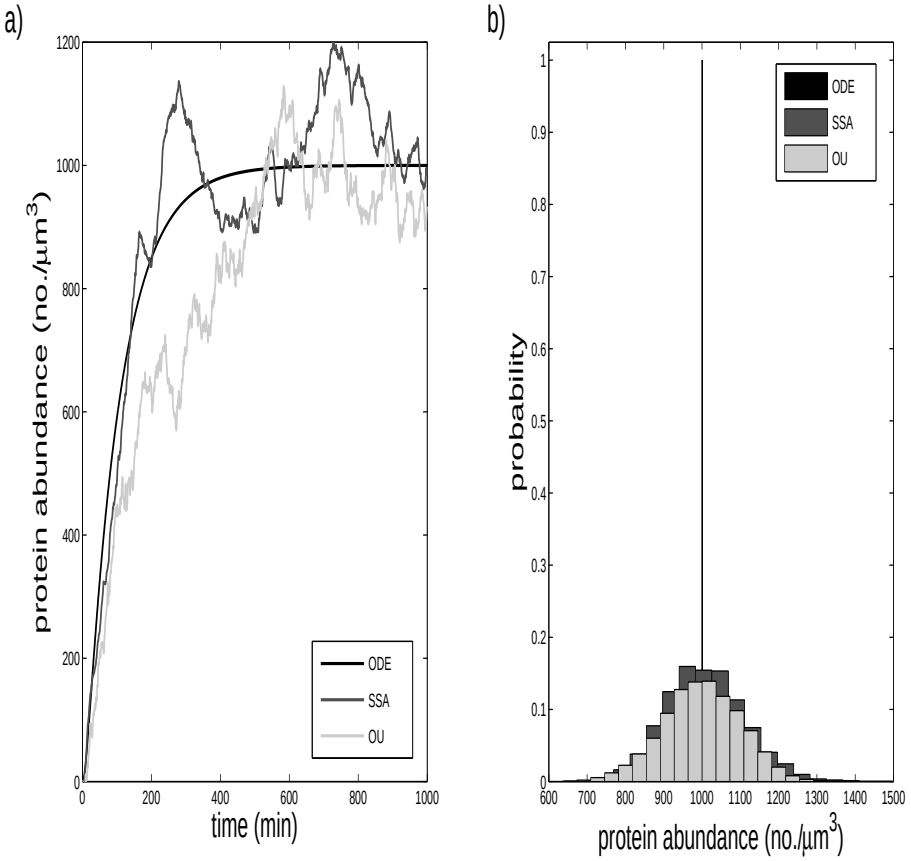


Figure 1.5 Simulating gene expression dynamics. (a) Time series of protein number generated by deterministic (ODE) and stochastic methods (OU process and SSA). (b) Histograms corresponding to steady-state numerical and stochastic simulations and shows the probability that a cell will have a given intracellular protein level. Parameters were set to (units min^{-1}): $s_A = 1$, $s_P = 1$, $\delta_M = 0.1$, $\delta_P = 0.01$ (ODE and SSA), and $c = 100$, and $\tau = 200$ (OU). In (a) and (b), cells are assumed to have a fixed volume of 1 μm^3 and in (b) steady-state distributions were obtained after 10^6 minutes of simulation time.

noise, white noise terms (ξ_M and ξ_P) are added to the ordinary differential Eqs. (1.2) and (1.3)

$$\frac{d[M]}{dt} = s_A - \delta_M[M] + \xi_M, \tag{1.7}$$

$$\frac{d[P]}{dt} = s_P[M] - \delta_P[P] + \xi_P. \tag{1.8}$$

Since $\langle \xi_M(t) \rangle = \langle \xi_P(t) \rangle = 0$ (angle brackets denote ensemble average), M^s and P^s obtained from these equations are equivalent to Eqs. (1.4) and (1.5). The stochastic variables ξ_M and ξ_P, have been shown to satisfy [145]

$$\begin{aligned} \langle \xi_M(t_1)\xi_M(t_2) \rangle &= 2\delta_M M^s \delta(t_1 - t_2), \\ \langle \xi_P(t_1)\xi_P(t_2) \rangle &= 2\delta_P P^s \delta(t_1 - t_2), \end{aligned} \tag{1.9}$$

where M^s and P^s are respectively given in Eqs. (1.4) and (1.5), $\delta(t1-t2)$ is the Dirac delta function, and $\langle \xi_M(t_1)\xi_P(t_2) \rangle = 0$.

To model extrinsic noise, a variable ϵ can be added to any of the rate constants in the model of Fig. 1.2. The dynamics of ϵ satisfy the Ornstein-Uhlenbeck (OU) process, a continuous, mean-reverting stochastic process that can be represented as a Langevin equation

$$\frac{d\epsilon(t)}{dt} = \frac{1}{\tau}(\mu - \epsilon(t)) + \sqrt{c}\xi_0(t), \tag{1.10}$$

where $\xi_0(t)$ is Gaussian white noise ($\langle \xi_0(t) \rangle = 0$, $\langle \xi_0(t_1)\xi_0(t_2) \rangle = \delta(t)$). Note that $\xi_0(t)$ is uncorrelated with $\xi_M(t)$ and $\xi_P(t)$. Correlated extrinsic fluctuations acting on different parameters of the system correspond to the same ϵ changing each of the relevant parameters. Multiple uncorrelated extrinsic fluctuations require an ϵ_i for each uncorrelated fluctuation. The parameter τ is called the relaxation time, and c is called the diffusion constant. The diffusion constant determines the strength of random fluctuations, while the relaxation time determines how rapidly on average a fluctuation will dissipate back to the mean (specifically, $1/e$ of the mean).

The probability density function (p) for the OU process is given by [52]

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{\tau}\frac{\partial [xp(x,t)]}{\partial x} + \frac{c}{2}\frac{\partial^2 p(x,t)}{\partial x^2}, \tag{1.11}$$

which is the Fokker-Planck equation for the OU process. The stationary probability distribution of an OU process is a Gaussian distribution with mean μ and variance $c\tau/2$. The OU process thus allows a fixed degree of heterogeneity to be imposed over a long period of time in a single cell, or at a given instant in time across a large population of cells. The effects of gene expression dynamics on fitness can be investigated by varying the values of c and τ (see Chapter 5).

Exact Updating Formula

An exact updating formula for Ornstein-Uhlenbeck process was developed by D.T. Gillespie [49, 52]

$$x(t+\Delta t) = x(t)\exp(-\Delta t/\tau) + [c\tau/2(1-\exp(-2\Delta t/\tau))]^{1/2}\, n, \tag{1.12}$$

where Δt is a positive finite variable and n is a unit normal random number.

The advantage of using a stochastic framework to simulate the present model of gene expression can be seen in Fig. 1.5. Specifically, the stochastic method captures not only the mean concentrations, but also the fluctuations in molecular species abundance (Fig. 1.5a). These fluctuations provide the information necessary for the histograms that describe the probability that a cell will have a given level of a particular molecular species (Fig. 1.5b), which can play a significant role in cellular dynamics and fitness (discussed Section 1.1.3).

1.2.3 Chemical Master Equation

In the stochastic formulation of chemical kinetics, the time evolution of a chemical system is described analytically by a finite differential-difference equation in which

time t and N distinct and reacting species populations, where $X_i : i \in \{1, 2, \ldots, N\}$ is the population of species $S_i : i \in \{1, 2, \ldots, N\}$, all appear as independent variables [50, 51]. The differential-difference equation in this context is referred to as the chemical master equation, and the function which satisfies it, namely $p(X_1, \ldots, X_N; t)$, is known as the grand probability function (GPF).

The GPF describes the probability that there will be X_1 molecules of S_1, X_2 molecules of S_2, ... , and X_N molecules of S_N, in a volume V at t; the CME is the equation governing the time-evolution of this function [50,51]. The state of the chemically reacting system can be described by the integer vector $\mathbf{x} = [X_1, \ldots, X_N]^T$. The system's state can change through any one of the M reactions $R_\mu : \mu \in \{1, 2, \ldots, M\}$. An R_μ reaction results in a state transition from $\mathbf{x}$ to $\mathbf{x} + \mathbf{s}_\mu$, where $\mathbf{s}_\mu$ is a vector which represents the changes in molecular species numbers that occurred as a result of R_μ. The propensity for a reaction to occur a_μ (units T^{-1}) is given by $a_\mu = c_\mu x_i$, where c_μ is a reaction parameter that characterizes reaction R_μ, and x_i is the number of distinct molecular reactant combinations for reaction R_μ found to be present in V at t. The fundamental hypothesis of the stochastic formulation of chemical kinetics is that the probability that a particular combination of reactant molecules in a particular reaction R_μ will react within the next infinitesimal time interval dt is given by $a_\mu dt$.

The CME can be expressed as follows [50, 51]

$$\frac{\partial}{\partial t} p(\mathbf{x}, t) = \sum_{\mu=1}^{M} \left[a_\mu(\mathbf{x} - s_\mu) p(\mathbf{x} - s_\mu, t) - p(\mathbf{x}, t) a_\mu(\mathbf{x}) \right]. \tag{1.13}$$

The first and second moments of $p(\mathbf{x}, t)$ with respect to a species i are the average number $\langle X_i(t) \rangle$ of that molecule and variance $\sigma_i^2(t) = \langle X_i^2 \rangle - \langle X_i \rangle^2$. The intrinsic noise, which unless otherwise indicated will be hereafter denoted by η_i, for species i, is defined by $\eta_i(t) = \sigma_i(t) / \langle X_i(t) \rangle$, and is thus directly related to the moments of

$p(\mathbf{x}, t)$ [133]. Note that an approximation of the CME is the Fokker-Planck equation, which describes the time evolution of a continuous probability distribution. If we set $\frac{\partial}{\partial t}p(X_i, t) = 0$, then in some cases it is possible to directly obtain the stationary probability distribution $p^s(X_i)$. In the steady-state, the probability of transition from a state with X_i molecules to the state with $X_i + 1$ molecules must be equal to the probability of transition from a state with $X_i + 1$ molecules to the state with X_i molecules.

In order to construct the CME associated with the model of gene expression shown in Fig. 1.2 we proceed as follows. To write the CME in a compact form, we introduce a *step operator* $\mathbf{E}_i^k$ which describes the addition or removal of k molecules of species i when a particular reaction occurs [133]. For a function $f(X_i, X_j)$ with two integer arguments, $\mathbf{E}_{X_i}^k$ increments X_i by an integer k, such that

$$\mathbf{E}_{X_i}^k f(X_i, X_j) = f(X_i + k, X_j). \tag{1.14}$$

Now we consider a change in the system due to degradation of mRNA. The change in probability is given by

$$\frac{dp(M, P, t)}{dt} = \delta_M(M + 1)p(M + 1, P, t) - \delta_M M p(M, P, t), \tag{1.15}$$

where the first and second terms describe, respectively, the flux in and out of state M, P, t due to the removal of one mRNA. Using the step operator and letting $p(M, P, t) = p$, the above equation can be expressed in a more compact form as

$$\frac{dp}{dt} = \delta_M(\mathbf{E}_{\mathbf{M}}^1 - 1)Mp. \tag{1.16}$$

If we incorporate the change in probability due to the production of mRNA then we obtain

$$\frac{dp}{dt} = s_A(\mathbf{E}_{\mathbf{M}}^{-1} - 1)p + \delta_M(\mathbf{E}_{\mathbf{M}}^1 - 1)Mp. \tag{1.17}$$

Similarly, the contributions to the probability flux due to the production and decay of mRNA and protein is

$$\begin{aligned}\frac{dp}{dt} &= s_A(\mathbf{E}_{\mathbf{M}}^{-1}-1)p + s_P(\mathbf{E}_{\mathbf{P}}^{-1}-1)Mp \\ &\quad + \delta_M(\mathbf{E}_{\mathbf{M}}^{1}-1)Mp + \delta_P(\mathbf{E}_{\mathbf{P}}^{1}-1)Pp. \end{aligned} \tag{1.18}$$

Because the CME is linear in the state variables M and P the moments of the probability distribution can be calculated using moment generating functions [155]. The first moment of p yields the average steady-state numbers of mRNA (for a derivation see supplemental materials in [135])

$$\langle M^s \rangle = \frac{s_A}{\delta_M} \tag{1.19}$$

and protein molecules

$$\langle P^s \rangle = \langle M^s \rangle \frac{s_P}{\delta_P} = \frac{s_A s_P}{\delta_M \delta_P}. \tag{1.20}$$

Note that Eqs. (1.19) and (1.20) are in agreement with Eqs. (1.4) and (1.5). The variances in the state variables can be obtained from the second moment of p, and thus an expression for the intrinsic noise in the steady-state protein abundance can be determined. The noise in steady-state protein number can be expressed as [135]

$$\eta_P^s = \frac{\sigma_P^s}{\mu_P^s} = \left(\frac{1}{P^s} + \frac{1}{1+\phi}\frac{1}{M^s}\right)^{1/2}, \tag{1.21}$$

where $\phi = \delta_M/\delta_P$. The production and decay of mRNA molecules can modeled as a birth-death process and thus the noise in steady-state mRNA number η_M^s is simply $1/\left\langle\sqrt{M^s}\right\rangle$ (since from Poisson statistics $(\sigma_M^s)^2 = \langle M^s \rangle$).

Although there are a few specific cases were the CME can be solved exactly, in general, analytical solving or numerically simulating the master equation for a system of realistic size and complexity is not possible [50,51,165]. The approach used in practice is to simulate the very process that the CME describes using Monte Carlo (MC) methods.

Stochastic Simulation Algorithm

Models involving the stochastic formulation of chemical kinetics are increasingly being used to simulate and analyze the dynamics of cellular systems [48]. Analytical solutions to these models are often intractable due to the nonlinearity of the corresponding system of equations. Thus, Monte Carlo (MC) simulation procedures for the number of each molecular species are commonly employed. Among these procedures, the Gillespie stochastic simulation algorithm is the *de-facto* standard for simulating biochemical systems in situations where a deterministic formulation may be inadequate [50, 51].

For every simulation step the SSA determines which reaction in the system will occur next and the time at which this reaction will occur (τ) [50, 51]. These items are determined from the following probability density function

$$p(\tau, \mu) = a_\mu \exp(a_0 \tau), \tag{1.22}$$

where $a_\mu = h_\mu c_\mu$, $a_0 = \sum a_\mu$, and $p(\tau, \mu)dt$ is the probability that the given reaction μ will occur in the infinitesimal time interval dt. After determining μ and τ, the system time and integer number of molecules in the system products are updated. This process is reiterated until the simulation is finished. The direct method SSA can be implemented via the pseudocode provided in Chapter 3. This algorithm, unlike most procedures for solving deterministic rate equations, never approximates an infinitesimal time interval dt by a finite time step Δt. The algorithm determines the exact time at which individual molecular reactions occur.

The following reaction equations are required to stochastically simulate the model of gene expression under consideration (Fig. 1.2)

$$A \xrightarrow{s_A} A + M, \tag{1.23}$$

$$M \xrightarrow{s_P} M + P, \tag{1.24}$$

$$M \xrightarrow{\delta_M} \oslash, \tag{1.25}$$

$$P \xrightarrow{\delta_P} \oslash. \tag{1.26}$$

Equations (1.23) and (1.24) respectively describe the transcription and translation processes. The degradation of M and P are accounted for by Eqs. (1.25) and (1.26), respectively. A single SSA time series realization and the steady-state distribution corresponding to Eqs. (1.23)-(1.26) are shown in Fig. 1.5a and Fig. 1.5b, respectively.

There is a simple relationship between deterministic and stochastic rate constants. This relationship is important because it allows experimentally determined rate constants to be incorporated into the SSA [74]. For zero-order reactions (reactions that proceed at a rate that is independent of the reactant concentration) and first-order reactions (reactions that proceed at a rate that depends linearly on only one reactant concentration) the deterministic rate constant k_μ and stochastic rate constant c_μ are equal [50]. The reason for the parity is that zero-order and first-order reactions are independent of the reaction volume, whereas the reaction rates for higher-order reactions depend on the reaction volume as the reactant molecules must collide in order to react. For second order reactions (reactions that proceed at a rate that depends on the concentration of one second-order reactant or two first-order reactants), the stochastic rate constant equals the deterministic rate constant divided by the volume of the reaction environment

$$c_\mu = \frac{nk_\mu}{N_A V}, \tag{1.27}$$

where N_A is Avogadro's number, and $n = 1$ if the reaction involves two different species of reactants and $n = 2$ if the reaction involves two reactants of the same species [74].

1.3 Modeling Population Dynamics

1.3.1 Cellular Growth and Division

The origins and consequences of cell-to-cell variability is often investigated analytically or computationally using single-cell models (e.g., [28, 146]). Such models typically ignore or idealize the effects of cell growth and division, and rarely capture the population-level effects of differential reproduction [2]. In ODE models, a first-order effective degradation term is often used to account for dilution due to exponential growth [64]. Similarly, in discrete stochastic models, the effect of growth and/or partitioning of cellular contents at division can be approximated by increasing the degradation rates of all components [113]. However, these methods tend to average away the dynamics resulting from growth and division that can play an important role in cellular dynamics (e.g., asymmetric division [61, 106, 142, 168]). Another approach is to model the cellular growth and division explicitly.

Based on observations in prokaryotes (e.g., [77]), cellular growth has been modeled using a linear function [7, 74, 75, 146]

$$V(t) = V_0(1 + t_d/T), \tag{1.28}$$

where V_0 is the volume of the cell at the time of its birth, t_d is the time since last division, and T is the interval between volume doubling. Exponential growth has been observed in eukaryotes (e.g., [148]) and has been described using an exponential growth law [26, 96, 157]

$$V(t) = V_0 \exp\left[\ln 2(t_d/T)\right]. \tag{1.29}$$

Modeling cell division involves determining when a particular cell will divide and how the cell volume and contents will be distributed. In models that do not incorporate the effects of cell size, division can be based on a periodic or random time (see,

e.g., Chapter 5). When cell size is explicitly incorporated into the model, the simplest option is to assume that division occurs once the cell has exceeded a critical size V_{div} corresponding to one doubling of its initial volume V_0. Another option is "sloppy cell-size control" [154], where cell division is treated as a discrete random event that takes place with a volume-dependent probability. Asymmetric cell division can be modeled in either of these cases by setting $V_{daughter} < V_{mother}$ such that total cell volume prior to cell division is conserved. Age dependent replication is described in Chapter 4.

When cell division is triggered, additional rules must be specified to model the partitioning of cellular content between mother and daughter cells. When the cellular contents are assumed to partition independently, they can be partitioned deterministically between the two volumes (e.g., symmetrically [74]) or probabilistically (e.g., using binomial distribution to partition non-DNA molecules [146]). Models of more disordered segregation (e.g., clustering due to packaging in vesicles) can be modeled as a multinomial process [61].

1.3.2 Constant-Number Monte Carlo

Quantitative modeling plays an important role in bridging the relationship between single-cell and cell population dynamics. An approach known as population balance modeling addresses this using partial differential equations [41, 44, 45, 63, 120, 152]. In this approach, cells are described as a continuous density flowing through a multi-dimensional state space that quantifies different physiological attributes (e.g., mass and chemical composition), and integral terms are used to account for birth and death processes along with a function describing the partitioning of cellular contents at division. The time-evolution is uniquely determined by the initial population distribution, and all cell densities change according to the same deterministic rules.

Correspondingly, information about individual cell trajectories is lost in this formalism. Population balance models also quickly become very difficult to formulate and solve when more than a few variables are incorporated [57].

Another approach is to simulate a sufficiently large ensemble of individual cells serving as a representative sample of the "true" population. This is advantageous because it puts the available methods for single-cell simulation at our disposal without the difficulty of integrating complicated and heterogeneous single-cell behaviour into a broader mathematical framework. Notably, the use of individual-based models places virtually no constraints on the biologically relevant details that can be formulated and simulated (e.g., [67]).

Models of non-interacting and non-dividing cells have been used extensively to study population variability arising from the process of gene expression. These models are typically simulated by performing numerous independent realizations of individual cells. To incorporate cell division, one could simulate the time courses of single cells, randomly choosing one of the two newborn cells to follow when a cell divides. The result is lineages (or cell chains) containing a single individual per generation (e.g., [146]). One problem with the cell chain approach is that it does not take into account the proliferative competition between cells in different gene expression or cellular states, and so will not provide the correct joint distribution of cell properties except in special cases. Another approach is to simulate the time courses of single cells, and continue to simulate all newborn cells produced. The result is a complete lineage tree (e.g., [123]). This approach can be used when dealing with a model in which cell proliferation can vary with a number of intrinsic variables such as age, metabolic state, and cell type. The problem here is that the size of the simulation ensemble rapidly grows to the point of intractability. This can be addressed using a technique called the constant-number Monte Carlo (CNMC) method [79, 93, 139],

originally developed to approximate the solution of population balance models of particulate processes. The idea is that the total number of particles being simulated is kept fixed while the composition of different particles in the finite sample still reflects the true number density within the full population. Note that as the CNMC method is using a fixed number of randomly chosen individuals to represent the population as a whole, there is an error associated with the method. This error was shown to be $(\ln x)^{0.89}/\sqrt{2N}$, where x is the extent of growth and N is the size of the sample population [139].

The CNMC method has been applied to the simulation of heterogeneous populations [26, 100, 101]. Mantzaris used the method in conjunction with deterministic and stochastic Langevin models of single-cell dynamics [100], along with methods to determine the timing of cell divisions and partitioning of cell contents that agree with the population balance formalism [101].

The SSA and CNMC method were combined in a cell population dynamics algorithm [26] discussed in the next section. In this implementation of the CNMC method, the individual mother and daughter cells are stored in two separate arrays [26]. Each time a cell divides, the daughter cell is placed in the daughter array and the time of birth recorded. Then, at specified intervals, cells within the mother array are replaced one at a time, with the oldest daughter cells being inserted first. Because every mother cell is equally likely to be replaced during the sample update, the joint distribution of the population remains intact for sufficiently large N [139].

1.3.3 Population Dynamics Algorithm

For my master's thesis [24], I developed a population dynamics algorithm (PDA) that combines the SSA (Section 1.2.3) with the CNMC method (Section 1.3.2) to accurately and efficiently simulate gene expression dynamics across growing and dividing

cell populations at single-cell resolution [26]. The PDA was the algorithm employed to perform the cell population simulations in Chapters 5 and 6.

Individual-based simulations provide an attractive alternative to rigorous mathematical analysis in investigations of phenomena such as drug resistance resulting from heterogeneous dynamics of cell populations. This is because of the relative ease of incorporating biologically relevant details, which generally make the mathematical model equations difficult to formulate and solve. Individual-based simulations can also capture the trajectories, lineages or fates of individual cells, thus enabling a more direct comparison of simulation and experimental results. The main drawback is that comprehensive analyses require significant computational resources.

The SSA in particular can be very computationally intensive since the step size τ becomes very small when the total number of molecules is high or the fastest reaction occurs on a time-scale that is much shorter than the time-scale of interest. It is therefore useful to develop techniques that can be used to speed up simulations. An exact method known as the "Next Reaction Method" was developed by Gibson and Bruck [47]. This method minimizes the number of random numbers required (to one per time step) and reduces the number of redundant time calculations through the use of dependency graphs that denote the dependencies between reactions. The reaction times are stored in a data structure known as an indexed priority queue. However, the same reasons that make the Next Reaction Method faster, also make it much more challenging to code and implement within the framework of the PDA. SSA simulations can also be sped up using approximate simulation methods such as the tau-leaping procedure, in which each time step τ advances the system through possibly many reaction events [53]. In order to preserve statistical accuracy of the simulation results while decreasing simulation run-times, the SSA was used together with parallel computing methods [26]. Though, in principle, any numerical or sim-

ulation method can be implemented in the PDA (e.g., in Chapter 5 the OU process was used to simulate the gene expression dynamics).

The accuracy of the PDA was benchmarked against exact and approximate analytical solutions for several scenarios of increasing biological complexity, including steady-state and time-dependent gene expression, and the effects on population heterogeneity of cell growth, division, and DNA replication [26]. This comparison demonstrated that the PDA provides an accurate approach to simulate how complex biological features influence gene expression. To further benchmark the algorithm, we implement a coarse-grained two-state model of a bet-hedging yeast population in order to simulate fitness (growth-rate) dynamics under environmental stress and the results were found to be qualitatively in agreement with experimental data. Finally, for the same yeast population, we simulated a fine-grained model which explicitly incorporates both gene expression and fitness in order to capture environmental effects on phenotype distributions. These simulations exemplified the utility of the PDA in a case where deterministic methods were unable to account for the full dynamics of the system.

1.3.4 Noise and Fitness

The stochastic nature of gene expression has led to the hypothesis that evolution by natural selection has fine-tuned noise-generating mechanisms and genetic architectures to derive beneficial population diversity [95, 140, 156]. Direct evidence that genome sequence contributes to cell-to-cell variability indicates that gene expression noise, like other genome-encoded traits, is inheritable and subject to selective pressures, and therefore evolvable. Large-scale proteomic studies in yeast have shown that genes associated with stress response pathways have elevated levels of intrinsic noise [11, 43, 107]. Stress-response genes have likely experienced positive pressure

toward high expression variability, presumably because this provides a selective advantage during periods of stress. By broadening the range of environmental stress resistance across a population, added gene expression noise could increase the likelihood that some cells within the population are better able to endure environmental assaults [20, 144]. Experimental results providing support for this hypothesis were obtained in a study by Bishop *et al.* [16], which demonstrated a competitive advantage of stress-resistant yeast mutants under high stress due to increased phenotypic heterogeneity.

In a qualitative explanation, Blake *et al.* [18] attribute the differential impact of added noise to a change in the relative fraction of surviving cells at different levels of stress (Fig. 1.3). In a quantitative model, the size of the fraction of viable cells depends on the probability distribution function associated with the spread of protein content among individual cells. Consequently, if it is assumed that cells are either unaffected or completely affected by the stress, the population fitness (reproductive rate) and differential fitness (difference in reproductive rates between two populations, e.g., a high and a low noise cell population) for a given stress level can be calculated (Fig. 1.6a and 1.6b, respectively) [42]. This provides a very simple quantitative framework that captures the observed impact of population heterogeneity on population fitness following acute stress.

The impact of acute stress on the fitness of the cell population W (macroscopic fitness) can be calculated theoretically by evaluating the integral

$$W = \int w(x)f(x)dx, \tag{1.30}$$

where the fitness of an individual cell $w(x)$ (microscopic fitness) is the relative reproductive rate of cells expressing a stress-related gene at a level given by x, and $f(x)$ describes the population distribution of gene expression when cells are exposed to

stress [171]. In a study by Fraser *et al.* [42], this distribution was approximated by the lognormal distribution

$$f(x) = \frac{1}{x\beta\sqrt{2x}} \exp\left[\frac{(\ln(x) - \alpha)^2}{2\beta^2}\right], \tag{1.31}$$

where α and β are defined by the average gene expression level μ and gene expression noise η through the relationships $\beta^2 = \ln(1+\eta^2)$ and $\alpha = \ln(\mu) - 0.5\beta$. The impact of acute stress was approximated by a step function such that cells expressing a stress-resistance gene below a certain threshold would have a reproductive rate of zero, that is, fitness $w(x) = 0$ for $x < s_{thr}$ and are otherwise unaffected, that is, $w(x) = 1$ for $x \geq s_{thr}$.

If it is assumed that the level of stress s experienced by the population is related to the most likely level of gene expression (i.e., the mode of the distribution in Eq. (1.31)), then the noise-dependency of population fitness in Eq. (1.30) for a threshold model is given by the error function (erf) describing the cumulative lognormal distribution

$$\begin{aligned} W(\eta, s) &= \int w(x)f(x)dx = \int_{s_{thr}}^{\infty} f(x)dx \\ &= \frac{1}{2} + \frac{1}{2}erf\left[\sqrt{\frac{\ln(1+\eta^2)}{2}}\left(\frac{\ln(s_{thr}/s)}{\ln(1+\eta^2)} - 1\right)\right], \end{aligned} \tag{1.32}$$

where cells with high expression of a stress-resistant gene have high fitness, and cells with low expression have low fitness (positive selection scheme) [27, 171]. Eq. (1.32) was used to calculate the fitness curves displayed in Fig. 1.6a using $s_{thr} = 6.91$ and $\eta = 0.1$ or $\eta = 0.4$, for the low and high noise populations, respectively. Correspondingly, the differential fitness curves displayed in Fig. 1.6b were obtained by evaluating the quantity $\Delta W(\eta, s) = W(\eta, s) - W(\eta_0, s)$, where $W(\eta, s)$ is the fitness of the high noise population ($\eta = 0.2$, 0.3, or 0.4) and $W(\eta_0, s)$ is a reference population with low noise ($\eta_0 = 0.1$).

The model of fitness presented in this section assumes that the gene expression level in each cell is fixed and is therefore appropriate only when considering an acute

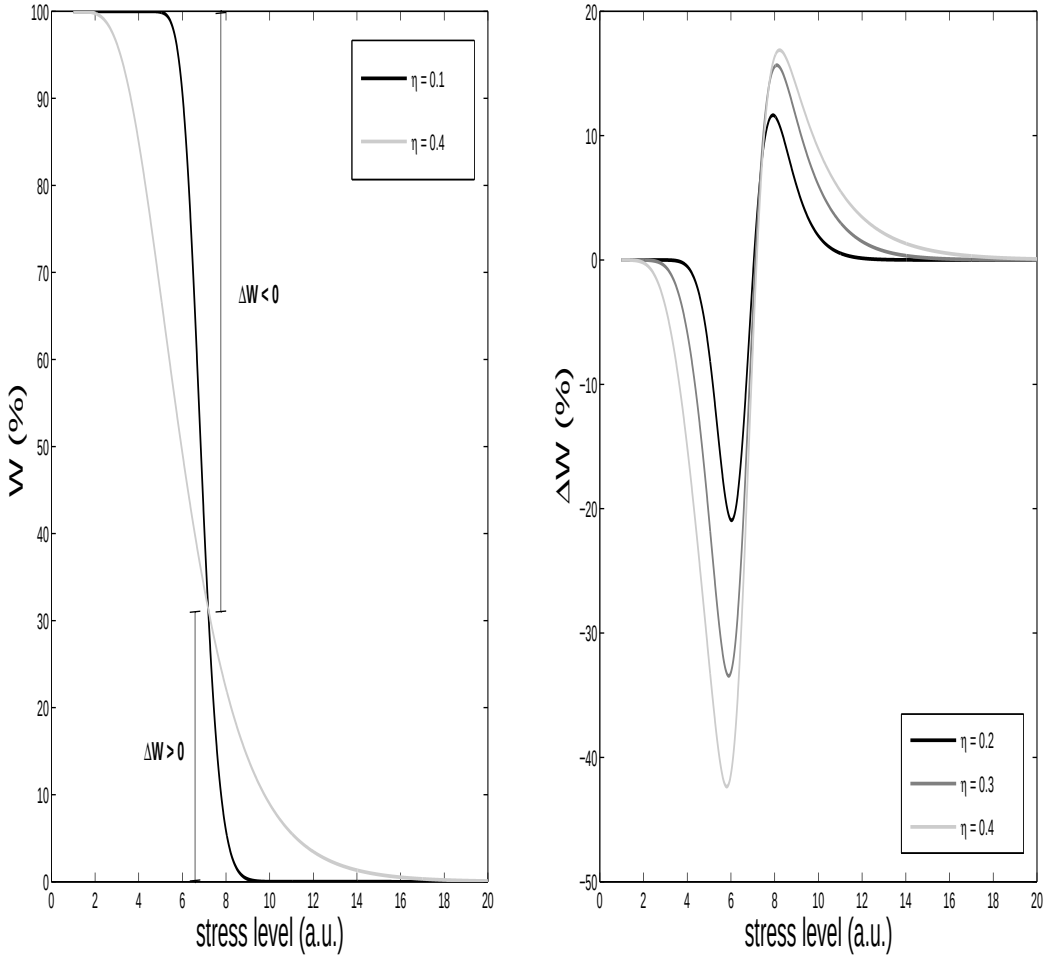


Figure 1.6 The effects of noise in the expression of a stress-resistant gene. (a) The effect of varying the stress level on fitness (W) for low and high noise (η) cell populations. Stress levels where noise is beneficial and disadvantageous are defined by positive and negative values of the differential fitness (ΔW), respectively. (b) Differential fitness at varying stress levels for three populations with elevated noise relative to a low noise ($\eta_0 = 0.1$) reference population. Figure reproduced from Zhuravel *et al.* [171] by permission of the Creative Commons Attribution Noncommercial License.

stress. In reality, the level of expression in each cell fluctuates and thus the fitness of the cell can change even in a constant stressful environment. In order to describe the effects of expression noise, prolonged stress exposure, and fitness, we developed a general theoretical framework incorporating the timescale of gene expression fluctuations (see Chapter 5). Specifically, we generalized Eq. (1.30) by incorporating the first-passage time distribution of gene expression fluctuations. When this equation cannot be solved analytically (e.g., when an equation for the first-passage time distribution is unavailable) or numerically (e.g., when it is unknown how $f(x)$ evolves in time), we resort to computer simulations to incorporate the effect of gene expression fluctuations on fitness. Each generation, macroscopic fitness is determined by the ratio of the number of cell divisions that occur in a population under stress (cells with sufficiently long first-passage times such that they can maintain their expression level x above a fitness threshold and divide during that generation) to the number of cell divisions in the corresponding unstressed population. The simulations are performed such that mother cells can only divide once and daughter cells not at all (daughter cells may become mother cells in the next generation if they are not replaced themselves by other daughter cells during the CNMC substitution). Whether or not a mother cell divides during a particular generation depends on their microscopic fitness. For a given generation, the population has maximum fitness if all the mother cells divide once, and minimum fitness if none of the mother cells divide. Although the total number of cells in the population is fixed by the CNMC method, microscopic fitness, and thus macroscopic fitness, are dynamic variables.

1.4 Drug Resistance

1.4.1 Mechanisms of Drug Resistance

Drug resistance was first observed in hospitals where, at the time, most antibiotics were being used [86]. For example, sulfonamide resistant *Streptoccoccus pyogenes* was discovered in the late 1930's [82], followed by penicillin resistant *Staphylococcus aureus* [14] and streptomycin resistant *Mycobacterium tuberculosis* [33] in the 1940's. These resistant strains emerged shortly after the introduction of the antibiotics [81, 86, 88, 105, 167]. Although drug resistance is not a recent phenomenon, the number of resistant organisms and range of organisms resistant to multiple drugs is on the rise [30, 81, 86, 88, 105, 151]. Thus, elucidating the underlying mechanisms of drug resistance is of ever increasing importance in medicine [56].

There are many biochemical mechanisms which allow microorganisms to resist drugs. These mechanisms can be broadly divided into three categories: reduced drug delivery [56, 104, 108], reduced drug activity (via drug inactivation, and alteration of drug target sites and metabolic pathways) [56,62,88,141,161,162], and increased drug efflux [84, 94, 109]. These mechanisms are influenced by both genetic and epigenetic factors (discussed in Sections 1.4.2 and 1.4.3, respectively).

Drug efflux pumps are of particular relevance to this thesis (Chapter 6). This mechanism allows cells to resist drugs by pumping cytotoxic drugs out of the cell. For example, drug efflux mediates resistance to tetracyclines, chloramphenicol, and fluoroquinolones [84, 109]. The most well known class of efflux pumps are transmembrane transport proteins of the ATP-binding cassette (ABC) superfamily. ABC-transporters are characterized by the presence of a cytoplasmic adenosine triphosphate (ATP) binding domain [94]. It harnesses energy from ATP hydrolysis that the transporters require to pump out various intracellular drugs. ABC-transporters also

contain a transmembrane domain. The transmembrane domain offers the binding site for substrates or drugs for translocation from the cytoplasm to the extracellular environment.

More than one type of mechanism may provide resistance to the same drug. For example, tetracycline resistance can be affected by either efflux or ribosome protection [102]. On the other hand, pleiotropic drug resistance (PDR) can result from a single mechanism such as the ABC-transporter PDR5 that provides yeast with resistance to several structurally and functionally unrelated drugs [10, 34–36, 68]. In some cases, PDR can be reversed by a variety of pharmacological agents which promote drug accumulation (e.g., calcium channel blockers) [56]. However, in other cases, PDR strains can require the use of more than six different drugs [19] and patients may succumb to the PDR infection because all available drugs have failed [81].

1.4.2 Genetic Basis of Drug Resistance

Drug resistance is often associated with two components: the drug which inhibits sensitive microorganisms and selects the resistant ones, and the genetic resistance determinant selected by the drug [85, 87]. At a molecular level, the biochemical mechanisms mentioned in the previous section can be a result of, though not limited to, mutations, horizontal gene transfer, and gene amplification or deletion [6]. All these effects can occur directly on resistance genes or on genes involved in their regulation [56].

Decreased drug activity resulting from a modification of the drug target can occur due to mutations in a structural gene encoding a target protein and is usually associated with drugs whose target is well-defined [56]. For example, mutations in the enzyme thymidylate synthetase can result in resistance to 5-fluorouracil. In the absence of plasmids and transposons (which generally mediate high-level resistance), a

step-wise progression from low-level to high-level resistance occurs in bacteria through sequential mutations in chromosomes [81, 132, 163]. This process was responsible for the initial emergence of penicillin and tetracycline resistance in strains of *Neisseria gonorrhoeae*.

The long term use of a single antibiotic can not only select for bacteria that are resistant to the antibiotic but also to several others [83, 88]. This phenomenon reflects the linkage of different resistance genes on the same transposon or plasmid [143]. Bacteria that are already resistant to one growth inhibitory agent can recruit additional resistance traits from other bacteria sharing the environment. It was from the doubly resistant (penicillin and tetracycline) strains of *Neisseria gonorrhoeae* that the new fluoroquinolone-resistant strains emerged [88]. Many of the known resistance genes can be transferred from resistant bacteria to sensitive bacteria of the same or different species. Resistance genes can be transfered by horizontal gene transfer in several ways: cell-to-cell conjugation, transformation by naked DNA (as linear DNA or on plasmids) that is released by dead cells, or bacteriophage-mediated transduction [6,56,81,88,89]. There is evidence of the transfer of resistance elements to human commensal bacteria and pathogens [1,166], and extensive gene transfer in the human intestinal microbiome [128].

A direct consequence of gene amplification (increase in gene copy number) is protein overexpression. This can result in drug resistance due to an increase in the concentration of the drug target. For example, a methionine sulphoximine-resistant Chinese hamster ovary cell line has been described that over-produces the target enzyme glutamine synthetase and was associated with an increase in the copy number of the gene encoding the target protein [130]. Conversely, drug resistance can occur from the under-production of a protein resulting from a gene deletion. For example, it was found that deletion of the genes regulating the MSH2 enzyme required for DNA

mismatch repair increased resistance to thiopurine chemotherapy in human leukemia cells [38].

The types of genetic change discussed in this section are not mutually exclusive. In fact, examination of the multiple genetic changes that are frequently observed in drug resistant tumour cell lines suggests that they can operate simultaneously [56]. Furthermore, mutations causing a genetic change do not necessarily occur independently of their phenotypic consequences (i.e., mutations are not necessarily independent random events). Paradigm shifting research has shown that mutations facilitating adaptation can be induced by a stressor (for a review see Rosenberg [125]).

1.4.3 Epigenetic Basis of Drug Resistance

Defining the term "Epigenetic"

The classical definition of "epigenetic" is: "a change in the state of expression of a gene that does not involve a mutation, but that is nevertheless inherited in the absence of the signal (or event) that initiated that change" [118]. There are, however, in the literature several specific uses of this term.

The term epigenetic was coined by Waddington in the 1940s to describe phenotypes that arise from the interactions between genes, rather than those directly encoded by a gene [158]. This meaning is captured in the conceptual epigenetic landscape framework that Waddington proposed to explain the phenotypes of discrete cell types (Fig. 1.7). This idea allowed scientists to think conceptually about the process of cell development from zygote to adult [71]. In particular, the zygote has daughter cells which roll down the branching valleys into the final resting "wells" that specify the terminal cell types. The epigenetic landscape can be altered by an environment change or mutation which can lead to developmental errors. Physicists still

use the term epigenetic in the same manner in the context of genetic networks [8,160]. Mathematically, a stable gene expression pattern can be represented by an "attractor state" [60] at the bottom of a pseudo-potential well, similar to a minimal energy equilibrium state. The large number of such stable attractor states of a genomic network collectively comprise the epigenetic landscape in which valleys represent the "basins of attraction" of the attractor states. In this framework, an attractor state determines a cell-type specific gene expression profile [59,60,69]. Being at the bottom of a valley explains the inherent robustness of the cell-type specific gene expression profiles that continuously face temporal noise or stochastic perturbations [60].

Molecular biologists have borrowed the term epigenetic from Waddington to describe covalent modifications of DNA or histones (proteins found in eukaryotic cell nuclei that function like a spool to compact DNA into a structure called a nucleosome) [73, 131]. Gene expression can be upregulated or downregulated when DNA is methylated or histones are modified via methylation, acetylation, deacetylation, and so on [118]. DNA methylation, a modification found in some eukaryotes (not in yeast and flies) and bacteria can be self-perpetuating over many generations: maintenance methylases recognize hemi-methylated DNA, the product of replication of fully methylated DNA, and add methyl groups to the unmethylated DNA strand [117,137]. It is unknown if histone modifications, found in eukaryotes, can be transmitted upon replication. However, a recent experimental study of fly embryos suggests that histone modifying proteins, and not methylated histones, remain associated with DNA during replication [116]. It should be noted that histone modifiers and enzymes that trigger DNA methylation must be recruited to genes by specific DNA binding proteins. Therefore, it can be argued that the dynamics of a gene network are orchestrated by the genes themselves and not the other way around.

"Epigenetic memory" can be thought of broadly as the timescale associated with

an epigenetic phenotype. We have recently used the notion of a fluctuation timescale in gene expression to refer to epigenetic memory (see Chapter 5). Fluctuations in gene expression occur over a wide range of timescales; while intrinsic transcriptional bursting (the production of many mRNA molecules within a short time) may occur on the order of milliseconds to minutes, extrinsic sources of variation can lead to long range fluctuations with timescales on the order of the cell cycle [76, 119]. Thus, the relaxation time (discussed in Section 1.2.2) can be used to quantify epigenetic memory. Our definition of epigenetic memory is akin to the mean first-passage time of a Brownian particle returning to the attractor after a perturbation placed it somewhere in the basin of attraction.

Epigenetic Drug Resistance

It is well established that an acquired genetic change can cause drug resistance (e.g., [46, 55, 164]). Less appreciated is that drug resistance can emerge from epigenetic effects.

Zhuravel *et al.* demonstrated that budding yeast under drug treatment reversibly change expression levels in the direction that increases fitness [171]. Specifically, the authors showed that budding yeast can survive by adapting during drug exposure [171]. This study used the Ura3-FOA system described in Section 1.1.3 for the Acar *et al.* experiments. Here, the yeast strain was genetically engineered to express the Ura3 gene fused with a gene that produces a fluorescence protein to experimentally measure expression levels of the Ura3 protein. The mode of the gene expression distribution was observed to shift to a lower level under FOA treatment. The gene expression distribution relaxed back to its pretreatment distribution when the FOA was removed from the media. These results are characteristic of epigenetic drug resistance.

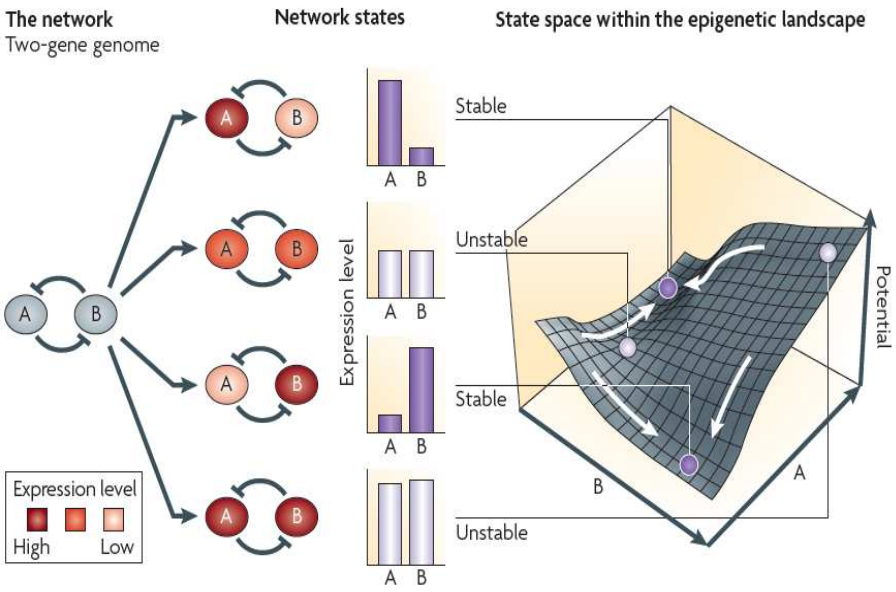


Figure 1.7 Schematic of the epigenetic landscape corresponding to a genetic toggle switch. The state space here is projected onto a two-dimensional plane in which each position represents a state corresponding to a particular gene expression pattern. Changes in gene expression translate into the movement of the network state in the state space. Each network state in the state space plane can be assigned to a quasi-potential, the magnitude of which is inversely related to the probability of the network being found in that state, which in turn reflects its stability when the system is at equilibrium. There are only two stable states in this network as gene A and gene B inhibit each other, and so all expression patterns in which A and B are equally expressed are highly unstable. For example, a slight excess of gene B (so that $B > A$) would suppress gene A, reducing its own inhibition and, hence, promoting its own expression, which further increases the excess of B over A expression. This would continue until the network reaches an equilibrium state at $B >> A$ when the expression of B cannot increase further owing to other limitations. Figure reproduced by permission from Macmillan Publishers Ltd: Nature Reviews Genetics 10: 336-342, copyright 2009.

Another example is the development of drug resistance during chemotherapy. When the drug Imatinib is used to treat chronic myeloid leukemia, the disease recurs with a frequency of 20-30% [21]. Even though numerous genetic mutations have been shown to render the drug ineffective [46, 55, 164], in two-thirds of cases no mutations have been found [21]. Instead, elevated levels of survival pathway proteins in Imatinib-resistant leukaemia cell lines were detected [110]. The rapid rate of resistance development, its dose dependence and high frequency of upregulation of the correct pathways are consistent with nongenetic heterogeneity, that is, variation in gene expression across a population of genetically identical cells. This mechanism generates enduring outlier cells with distinct phenotypes, some of which may be subject to selection.

Recently, it was found that the antibiotic Isoniazid (INH) failed to eliminate a portion of a clonal population of *Mycobacterium smegmatis* [159]. This phenomenon was previously attributed to infrequently dividing or nondividing persister cells (dormant cells that can withstand antimicrobial drugs) [9,37,91]. However, Wakamoto *et al.* detected cells dividing in the drug environment via single-cell measurements [159]. Pre-INH elongation rates between persister and nonpersister cells were found to be not significantly different, ruling out preexisting persister cells. Persistence was attributed to a reversible tolerance which resulted from low levels of gene expression (expressed catalase-peroxidase activates INH), rather than stable genetic mutations.

Epigenetic memory can contribute to the development of genetic mutations [21, 25]. It was recently argued that epigenetic memory may accelerate tumor progression by contributing to the development of drug resistant cancer cells [21]. In this hypothesis, phenotypic variability from the noisy expression of a resistance gene renders some cells and their offspring temporarily insensitive to the drug, thereby increasing the probability of acquiring a mutation conferring permanent immunity.

In budding yeast, the expression of the ABC-transporter PDR5 is controlled homologous transcriptional regulators PDR1 and PDR3. The PDR5 protein provides yeast with generalized resistance to a broad spectrum of functionally and structurally unrelated drugs. The topology of the PDR5 transcriptional network, as my thesis research has revealed (see Chapter 6), may significantly enhance the development of nongenetic drug resistance.

1.5 Thesis Overview

The unifying concept of the research presented in this thesis is that it has a physical basis. The physics of noise is introduced through a review article in Chapter 2. The physics of stochastic processes guided the use of a stationary approximation which yields a novel way to speedup population simulations in Chapter 3. Incorporating more biological detail results in both insights and the need for more sophisticated computational techniques (Chapter 4). To help make those techniques tractable, the principles of physics are employed to make simplifying assumptions. In Chapter 5, a simple physical model of a mean-reverting process reveals that gene expression noise can render a population drug resistant. Incorporating network topology demonstrates that such increased time scales can be achieved and suggests that they are tuned during evolution (Chapter 6).

In the final chapter of this thesis (Chapter 8), the main conclusions of the work presented in Chapters 3-6 are summarized, ongoing research discussed, and future research directions suggested.

The Appendices include a supplementary figure as well as enlarged versions of figures that are difficult to read in the main body of the text due to the direct incorporation of journal manuscripts into the thesis.

1.6 References

[1] F.M. Aarestrup. Veterinary drug usage and antimicrobial resistance in bacteria of animal origin. *Basic Clin. Pharmacol. Toxicol.*, 96:271–281, 2005.

[2] N. Abdennur. A framework for individual-based simulation of heterogeneous cell populations. Master's thesis, University of Ottawa, 2012.

[3] M. Acar, A. Becskei, and A. van Oudenaarden. Enhancement of cellular memory by reducing stochastic transitions. *Nature*, 435:228–232, 2005.

[4] M. Acar, J.T. Mettetal, and A. van Oudenaarden. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.*, 40:471–475, 2008.

[5] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton: Chapman & Hall/CRC.

[6] D.I. Andersson. Veterinary drug usage and antimicrobial resistance in bacteria of animal origin. *Basic Clin. Pharmacol. Toxicol.*, 96:271–281, 2005.

[7] A. Arkin, J. Ross, and H.H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda infected *Escherichia coli* cells. *Genet.*, 149:1633–1648, 1998.

[8] E. Aurell and K. Sneppen. Epigenetics as a first exit problem. *Phys. Rev. Lett.*, 88:048101, 2002.

[9] N.Q. Balaban. Persistence: mechanisms for triggering and enhancing phenotypic variability. *Curr. Opin. Genet. Dev.*, 21:768–775, 2011.

[10] E. Balzi and A. Goffeau. Yeast multidrug resistance: The pdr network. *J. Bioenerg. Biomembr.*, 27:71–76, 1995.

[11] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, and *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.*, 38:636–643, 2006.

[12] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[13] A.-L. Barabási and Z.N. Oltvai. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.*, 5:101–113, 2004.

[14] M. Barber. Infection by penicillin resistant staphylococci. *Lancet*, 2:641–644, 1948.

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

[16] A.L. Bishop, F.A. Rab, E.R. Sumner, and S.V. Avery. Phenotypic heterogeneity can enhance rare-cell survival in stress-sensitive yeast populations. *Molec. Microbiol.*, 63:507–520, 2007.

[17] W.J. Blake, G. Balazsi, M.A. Kohanski, F.J. Issacs, and *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Molec. Cell*, 24:853–865, 2006.

[18] W.J. Blake, M. Kaern, C.R. Cantor, and J.J. Collins. Noise in eukaryotic gene expression. *Nature*, 422:633–637, 2003.

[19] B.R. Bloom and C.J.L. Murray. Tuberculosis: commentary on a re-emergent killer. *Science*, 257:1055–1064, 1992.

[20] I.R. Booth. Stress and the single cell: intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. *Int. J. Food Microbiol.*, 78:19–30, 2002.

[21] A. Brock, H. Chang, and S. Huang. Non-genetic heterogeneity - a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.*, 10:336–342, 2009.

[22] H.H. Chang, M. Hemberg, M. Barahona, E. Ingber, and *et al.* Transcriptomewide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453:544–547, 2008.

[23] D. Charlebois. A biophysicist ponders the application of hidden metric spaces to genetic networks. *Nature*, 458:811, 2009.

[24] D.A. Charlebois. An algorithm for the stochastic simulation of gene expression and heterogeneous population dynamics. Master's thesis, University of Ottawa, 2010.

[25] D.A. Charlebois, N. Abdennur, and M. Kaern. Gene expression noise facilitates adaptation and drug resistance independently of mutation. *Phys. Rev. Lett.*, 107:218101, 2011.

[26] D.A. Charlebois, J. Intosalmi, D. Fraser, and M. Kaern. An algorithm for the stochastic simulation of gene expression and heterogeneous population dynamics. *Commun. Comput. Phys.*, 9:89–112, 2011.

[27] D.A. Charlebois and M. Kaern. *Information Processing and Biological Systems*. Springer-Verlag, 2011.

[28] K.C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, and *et al.* Kinetic analysis of a molecular model of the budding yeast cell cycle. *Molec. Biol. Cell*, 11:369–391, 2000.

[29] A.A. Cohen, N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, and *et al.* Dynamic proteomics of individual cancer cells in response to a drug. *Science*, 322:1511–1516, 2008.

[30] M.L. Cohen. Changing patterns of infectious disease. *Nature*, 406:762–767, 2000.

[31] F. Crick. On protein synthesis. *Symp. Soc. Exp. Biol.*, XII:139–163, 1958.

[32] F. Crick. Central dogma of molecular biology. *Nature*, 227:561–563, 1970.

[33] J. Crofton and D.A. Mitchison. Streptomycin resistance in pulmonary tuberculosis. *Br. Med.*, 2:1009–1015, 1948.

[34] A. Delahodde, T. Delaveau, and C. Jacq. Positive autoregulation of the yeast transcription factor pdr3p, which is involved in control of drug resistance. *Mol. Cell. Biol.*, 15:4043–4051, 1995.

[35] T. Delaveau, A. Delahodde, A. Carvajal, and E. Subnik and *et al.* Pdr3, a new yeast regulatory gene, is homologous to pdr1 and controls the multidrug resistance phenomenon. *Mol. Gen. Genet.*, 224:501–512, 1994.

[36] D. Dexter, W.S. Moye-Rowley, A.L. Wu, and J. Golin. Mutations in the yeast pdr3, pdr4, pdr7 and pdr9 pleiotropic (multiple) drug resistance loci affect the transcript level of an atp binding cassette transporter encoding gene, pdr5. *Genetics*, 136:505–515, 1994.

[37] N. Dhar and J. D. McKinney. Microbial phenotypic heterogeneity and antibiotic tolerance. *Curr. Opin. Microbiol.*, 10:30–38, 2007.

[38] B. Diouf, Q. Cheng, N.F. Krynetskaia, and W. Yang and *et al.* Somatic deletions of genes regulating msh2 protein stability cause dna mismatch repair deficiency and drug resistance in human leukemia cells. *Nat. Med.*, 17:1298–1303, 2011.

[39] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, and M. Foglierini *et al.* Noise in transcription negative feedback loops: simulation and experimental analysis. *Molec. Syst. Biol.*, 2, 2006.

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

[41] H.V. Foerster. *In The kinetics of cellular proliferation.* New York: Grune and Stratton, 1959.

[42] D. Fraser and M. Kaern. A chance at survival: gene expression noise and phenotypic diversification strategies. *Molec. Microbiol.*, 71:1333–1340, 2009.

[43] H.B. Fraser, A.E. Hirsh, G. Giaever, and J. Kumm and *et al.* Noise minimization in eukaryotic gene expression. *PLoS Biol.*, 2:e137, 2004.

[44] A.G. Fredrickson, D. Ramkrishna, and H.M. Tsuchiya. Statistics and dynamics of procaryotic cell populations. *Math. Biosci.*, 1:327–374, 1967.

[45] A.G. Fredrickson and H.M. Tsuchiya. Continuous propagation of microorganisms. *AIChE J.*, 9:459–468, 1963.

[46] C.B. Gambacorti-Passerini, R.H. Gunby, R. Piazza, A. Galietta, and *et al.* Molecular mechanisms of resistance to imatinib in philadelphia-chromosome-positive leukaemias. *Lancet. Oncol.*, 4:75–85, 2003.

[47] M.A. Gibson and J. Bruck. Exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, 105:1876–1889, 2000.

[48] D.T. Gillespie. *Handbook of Materials and Modeling.*

[49] D.T. Gillespie. *Markov processes: an introduction for physical scientists.* New York: Academic Press Limited.

[50] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.

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

[52] D.T. Gillespie. Exact numerical simulation of the ornstein-uhlenbeck process and its integral. *Phys. Rev. E.*, 54:2084–2091, 1996.

[53] D.T. Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.

[54] I. Golding, J. Paulsson, S.M. Zawilski, and E.C. Cox. Realtime kinetics of gene activity in individual bacteria. *Cell*, 123:1025–1036, 2005.

[55] M.E. Gorre, M. Mohammed, K. Ellwood, N. Hsu, and *et al.* Clinical resistance to sti-571 cancer therapy caused by bcr-abl gene mutation or amplification. *Science*, 293:876–880, 2001.

[56] J.D. Hayes and C.R. Wolf. Molecular mechanisms of drug resistance. *Biochem. J.*, 272:281–295, 1990.

[57] M.A. Henson. Dynamic modeling of microbial cell populations. *Curr. Opin. Biotech.*, 14:460–467, 2003.

[58] S. Huang. Non-genetic heterogeneity of cells in development: More than just noise. *Development*, 136:3853–3862, 2009.

[59] S. Huang, G. Eichler, Y. Bar-Yam, and D.E. Ingber. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.*, 94:128701, 2005.

[60] S. Huang and S. Kauffman. *Encyclopedia of Complexity and Systems Science.*

[61] D. Huh and J. Paulsson. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.*, 43:95–102, 2011.

[62] H.F. Jenkinson. Ins and outs of antimicrobial resistance: era of the drug pumps. *J. Dent. Res.*, 75:736–742, 1996.

[63] H.M. Tsuchiya J.M. Eakman, A.G. Fredrickson. Statistics and dynamics of microbial cell populations. *Chem. Eng. Prog. S. Ser.*, 62:37–49, 1966.

[64] M. Kaern, W.J. Blake, and J.J. Collins. The engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng.*, 5:179–206, 2003.

[65] M. Kaern, T.C. Elston, W.J. Blake, and J.J. Collins. Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.*, 6:451–464, 2005.

[66] S. Kalir, S. Mangan, and U. Alon. A coherent feed-forward loop with a sum input function prolongs flagella expression in escherichia coli. *Mol. Sys. Biol.*, 2005.

[67] J.R. Karr, J.C. Sanghvi, D.N. Macklin, M.V. Gutschow, and *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell*, 150:389–401, 2012.

[68] D.J. Katzmann, P.B. Burney, J. Golin, Y Mahe, and *et al.* Transcriptional control of the yeast pdr5 gene by the pdr3 gene product. *Mol. Cell. Biol.*, 14:4653–4661, 1994.

[69] S. Kauffmann. Homeostasis and differentiation in random genetic control networks. *Nature*, 224:177–178, 1969.

[70] S. Kauffmann. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theoret. Biol.*, 22:437–467, 1969.

[71] S.A. Kauffmann. The epigenetic landscape and clinical implications. *J. Crit. Care*, 26:e15, 2011.

[72] B.B. Kaufmann and A. van Oudenaarden. Stochastic gene expression: from single molecules to the proteome. *Curr. Opin. Genet. Dev.*, 17:107–112, 2007.

[73] S. Khorasanizadeh. The nucleosome: from genomic organization to genomic regulation. *Cell*, 116:259–272, 2004.

[74] A.M. Kierzek. Stocks: Stochastic kinetic simulations of biochemical systems with gillespie algorithm. *Bioinf.*, 18:470–481, 2002.

[75] A.M. Kierzek, J. Zaim, and P. Zielenkiewicz. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *J. Biol. Chem.*, 276:8165–8172, 2001.

[76] E. Klipp. Timing matters. *FEBS Lett.*, 583:4013–4018, 2009.

[77] H.E. Kubitschek. Cell volume increase in escherichia coli after shifts to richer media. *J. Bacteriol.*, 172:94–101, 1990.

[78] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309:2075–2078, 2005.

[79] K. Lee and T. Matsoukas. Simultaneous coagulation and break-up using constant-n monte carlo. *Powder Technol.*, 110:82–89, 2000.

[80] T.I. Lee, N.J. Rinald, F. Robert, D.T. Odom, and *et al.* Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, 298:799–804, 2002.

[81] S.B. Levy. *The Antibiotic Paradox: How Misuse of Antibiotics Destroys their Curative Powers*. Cambridge: Perseus Publishing.

[82] S.B. Levy. Microbial resistance to antibiotics. an evolving and persistent problem. *Lancet*, 2:83–88, 1982.

[83] S.B. Levy. *Engineered Organisms in the Environment: Scientific Issues*. Washington DC: ASM Press, 1985.

[84] S.B. Levy. Active efflux mechanisms for antimicrobial resistance. *Antimicrob. Agents Chemother.*, 36:695–703, 1992.

[85] S.B. Levy. Balancing the drug resistance equation. *Trends Microbiol.*, 2:341–342, 1992.

[86] S.B. Levy. The challenge of antibiotic resistance. *Sci. Am.*, 278:46–53, 1998.

[87] S.B. Levy. The 2000 garrod lecture. factors impacting on the problem of antibiotic resistance. *J. Antimicrob. Chemother.*, 49:25–30, 2002.

[88] S.B. Levy and B. Marshall. Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med.*, 10:S122–S129, 2004.

[89] S.B. Levy and R.V. Miller. *Gene Transfer in the Environment.* New York: McGraw Hill.

[90] S.F. Levy, N. Ziv, and M.L. Siegal. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLOS Biol.*, 10, 2012.

[91] K. Lewis. Persister cells. *Annu. Rev. Microbiol.*, 64:357–372, 2010.

[92] I. L'Heureux. Lecture notes: Computational physics i, 2012.

[93] Y. Lin, K. Lee, and T. Matsoukas. Solution of the population balance equation using constant-number monte carlo. *Chem. Eng. Sci.*, 57:2241–2252, 2002.

[94] F.S. Liu. Mechanisms of chemotherapeutic drug resistance in cancer therapy - a quick review. *Taiwan J. Obstet. Gynecol.*, 48:239–1252, 2009.

[95] L. Lopez-Maury, S. Marguerat, and J. Bahler. Tuning gene expression to changing environments: from rapid response to evolutionary adaptation. *Nat. Rev. Gen.*, 9:583–593, 2008.

[96] T. Lu, D. Volfson, L. Tsimring, and J. Hasty. Cellular growth and division in the gillespie algorithm. *Syst. Biol.*, 1:121–128, 2004.

[97] L. Ma, J. Wagner, J.J. Rice, and H. Wenwei and *et al.* A plausible model for the digital response of p53 to dna damage. *PNAS*, 102:14266–14271, 2005.

[98] N. Maheshri and E.K. O'Shea. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, 36:413–434, 2007.

[99] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *PNAS*, 100:11980–11985, 2003.

[100] N.V. Mantzaris. Stochastic and deterministic simulations of heterogeneous cell population dynamics. *J. Theor. Biol.*, 241:690–706, 2006.

[101] N.V. Mantzaris. From single-cell genetic architecture to cell population dynamics: Quantitatively decomposing the effects of different population heterogeneity sources for a genetic network with positive feedback architecture. *Biophys. J.*, 92:4271–4288, 2007.

[102] L.M. McMurry and S.B. Levy. *Tetracycline resistance in gram-positive bacteria, in Gram-Positive Pathogens*. Washington DC: ASM Press.

[103] R. Milo, S. Shen-Orr, S. Itzkovitz, and N. Kashtan *et al.* Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.

[104] L.G. Mitchell N.A. Campbell, J.B. Reece. *Biology*. Menlo Park: Benjamin/Cummings.

[105] H.C. Neu. The crisis in antibiotic resistance. *Science*, 257:1064–1073, 1992.

[106] R.A. Neumuller and J.A. Knoblich. Dividing cellular asymmetry: asymmetric cell division and its implications for stem cells and cancer. *Genes Dev.*, 23:2675–2699, 2009.

[107] J.R.S. Newman, S. Ghaemmaghami, J. Ihmels, and D.K. Breslow and *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, 441:840–846, 2006.

[108] H. Nikaido. Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science*, 264:382–388, 1994.

[109] H. Nikaido. Multidrug efflux pumps of gram-negative bacteria. *J. Bacteriol.*, 178:5853–5859, 1996.

[110] S. Okabe, T. Tauchi, and K. Ohyashiki. Characteristics of dasatinib and imatinibresitant chronic myelogenous leukemia cells. *Clin. Cancer Res.*, 14:6181–6186, 2008.

[111] D. Orrell and H. Bolouri. Control of internal and external noise in genetic regulatory networks. *J. Theor. Biol.*, 230:301–312, 2004.

[112] E.M. Ozbudak, M. Thattai, I. Kurtser, and A.D. Grossman and *et al.* Regulation of noise in the expression of a single gene. *Nat. Genet.*, 31:69–73, 2002.

[113] J. Paulsson. Noise in a minimal regulatory network: plasmid copy number control. *Quart. Rev. Biophys.*, 34:1–59, 2001.

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

[115] J.M. Pedraza and A. van Oudenaarden. Noise propagation in gene networks. *Science*, 307:1965–1969, 2005.

[116] S. Petruk, Y. Sedkov, D.M. Johnston, and J.W. Hodgson *et al.* Trxg and pcg proteins but not methylated histones remain associated with dna through replication. *Cell*, 150:922–933, 2012.

[117] M. Ptashne. On the use of the word epigenetic. *Curr. Biol.*, 17:R1–R4, 2005.

[118] M. Ptashne and A. Gann. *Genes and Signals*. New York: Cold Spring Harbor Press.

[119] A. Raj and van A. Oudenaarden. Nature, nuture, or chance: stochastic gene expression and its consequences. *Cell*, 135:216–226, 2008.

[120] D. Ramkrishna. *Population Balances: Theory and Applications to Particulate Systems in Engineering*. London UK: Academic Press.

[121] J.M. Raser and E.K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304:1811–1814, 2004.

[122] J.M. Raser and E.K. O'Shea. Noise in gene expression: origins, consequences, and control. *Science*, 309:2010–2013, 2005.

[123] A.S Ribeiro, D.A Charlebois, and J. Lloyd-Price. *CellLine*, a stochastic cell lineage simulator. *Bioinf.*, 23:3409–3411, 2007.

[124] A.S Ribeiro, R. Zhu, and S.A. Kauffman. A general modeling strategy for gene regulatory networks with stochastic dynamics. *J. Comp. Biol*, 13:1630–1639, 2006.

[125] S. Rosenberg. Evolving responsively: Adaptive mutation. *Nat. Rev. Genet.*, 2:504–515, 2001.

[126] N. Rosenfeld, J.W. Young, U. Alon, P.S. Swain, and *et al.* Gene regulation at the single-cell level. *Science*, 307:1962–1965, 2005.

[127] M. Roussel and R. Zhu. Validation of an algorithm for the delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Phys. Biol.*, 3:274–284, 2006.

[128] A.A. Salyers, A. Gupta, and Y. Wang. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends Microbiol.*, 12:412–416, 2004.

[129] M.S. Samoilov, G. Price, and A.P. Arkin. From fluctuations to phenotypes: The physiology of noise. *Sci. STKE*, 366:re17, 2006.

[130] P. G. Sanders and R.H. Wilson. Amplification and cloning of the chinese hamster glutamine synthetase gene. *EMBO J.*, 3:65–71, 1984.

[131] C.B. Schaefer, S.K. Ooi, T.H. Bestor, and D. Bourc'his. Epigenetic decisions in mammalian germ cells. *Science*, 316:398–399, 2007.

[132] T. Schneiders, S.G.B. Amyes, and S.B. Levy. Role of acrr and rama in fluoroquinolone resistance in clinical klebsiella pneumoniae isolates from singapore. *Antimicrob. Agents Chemother.*, 47:2831–2837, 2003.

[133] M. Scott, B. Ingalls, and M. Kaern. Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. *Chaos*, 16:026107, 2006.

[134] V. Shahrezaei, J.F. Ollivier, and P. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, 4, 2008.

[135] V. Shahrezaei and P.S. Swain. Analytical distributions for stochastic gene expression. *PNAS*, 105:17256–17261, 2008.

[136] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nat. Genet.*, 31:64–68, 2002.

[137] A.B. Silveira, C. Trontin, S. Cortijo, and J. Barau *et al.* Extensive natural epigenetic variation at a de novo originated gene. *PLoS Genet.*, 9:e1003437, 2013.

[138] M.L. Simpson, C.D. Cox, and G.S. Sayler. Frequency domain analysis of noise in autoregulated gene circuits. *PNAS*, 100:4551–4556, 2003.

[139] M. Smith and T. Matsoukas. Constant-number monte carlo simulation of population balances. *PNAS*, 53:1777–1786, 1998.

[140] W.K. Smits, O.P. Kuipers, and J.W. Veening. Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.*, 4:259–271, 2006.

[141] B.G. Spratt. Resistance to antibiotics mediated by target alterations. *Science*, 264:388–393, 1994.

[142] E.J. Stewart, R. Madden, G. Paul, and F. Taddei. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.*, 3:e45, 2005.

[143] A.O. Summers. Generally overlooked fundamentals of bacterial genetics and ecology. *Clin. Infect. Dis.*, 34 Suppl 3:S85–S92, 2002.

[144] E.R. Sumner and S.V. Avery. Phenotypic heterogeneity: differential stress resistance among individual cells of the yeast *Saccharomyces cerevisiae*. *Microbiology*, 148:345–351, 2002.

[145] P.S. Swain. Efficient attenuation of stochasticity in gene expression through posttranscriptional control. *J. Mol. Biol.*, 344:965–976, 2004.

[146] P.S. Swain, M.B. Elowitz, and E.D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *PNAS*, 99:12795–12800, 2002.

[147] J.J. Tabor, T.S. Bayer, Z.B. Simpson, and M. Levy *et al.* Engineering stochasticity in gene expression. *Mol. Biosyst.*, 4:754–761, 2008.

[148] S. Di Talia, J.M. Skotheim, J.M. Bean, and E.D. Siggia *et al.* The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature*, 448:947–952, 2007.

[149] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *PNAS*, 98:8614–8619, 2001.

[150] M. Thattai and A. van Oudenaarden. Attenuation of noise in ultrasensitive signaling cascades. *Biophys. J.*, 82:29432950, 2002.

[151] J. Travis. Reviving the antibiotic miracle? *Science*, 264:360–363, 1994.

[152] H.M. Tsuchiya, A.G. Fredrickson, and R. Aris. Dynamics of microbial cell populations. *Adv. Chem. Eng.*, 6:125–206, 1966.

[153] T.E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling in vivo reactions. *Comput. Biol. Chem.*, 28:165–178, 2004.

[154] J.J. Tyson and O.J. Diekmann. Sloppy size control of the cell division cycle. *Theor. Biol.*, 118:405–426, 1986.

[155] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Amsterdam: North-Holland.

[156] J.W. Veening, W.K. Smits, and O.P. Kuipers. Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.*, 62:193–210, 2008.

[157] D. Volfson, J. Marciniak, W.J. Blake, and N. Ostroff and *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 439:861–64, 2006.

[158] C.H. Waddington. The epigenotype. *Endeavour*, 1:18–20, 1942.

[159] Y. Wakamoto, N. Dhar, R. Chait, and K. Schneider and *et al.* Dynamic persistence of antibiotic-stressed mycobacteria. *Science*, 339:91–95, 2013.

[160] A.M. Walczak, J.N. Onuchic, and P. G. Wolynes. Absolute rate theories of epigenetic stability. *Proc. Natl. Acad. Sci. USA*, 102:18926–18931, 2005.

[161] C. Walsh. *Antibiotics: Actions, Origins, Resistance*. Washington: ASM Press.

[162] C. Walsh. Molecular mechanisms that confer antibacterial drug resistance. *Nature*, 406:775–781, 2000.

[163] H. Wang, J.L. Dzink-Fox, M. Chen, and S.B. Levy. Genetic characterization of highly fluoroquinolone-resistant clinical escherichia coli strains from china: role of acrr mutations. *Antimicrob. Agents Chemother.*, 45:1515–1521, 2001.

[164] Y. Wei, M. Hardling, B. Olsson, and R. Hezaveh and *et al.* Not all imatinib resistance in cml are bcr-abl kinase domain mutations. *Ann. Hematol.*, 85:841–847, 2006.

[165] D.J. Wilkinson. *Stochastic Modelling for Systems Biology*. Boca Raton: Chapman & Hall.

[166] W. Witte. Medical consequences of antibiotic use in agriculture. *Science*, 279:996–997, 1998.

[167] M.J. Wood. Microbial resistance: bacteria and more. *Clin. Infect. Dis.*, 36:S2–S3, 2003.

[168] P.S. Wu, B. Egger, and A.H. Brand. Asymmetric stem cell division: lessons from drosophila. *Semin. Cell Dev. Biol.*, 19:283–293, 2008.

[169] R. Zadrag-Tecza, M. Kwolek-Mirek, G. Bartosz, and T. Bilinski. Cell volume as a factor limiting the replicative lifespan of the yeast *Saccharomyces cerevisiae*. *Biogerontology*, 10:481–488, 2009.

[170] Z. Zhang, W. Qian, and J. Zhang. Positive selection for elevated gene expression noise in yeast. *J. Mol. Syst. Biol.*, 5, 2009.

[171] D. Zhuravel, D. Fraser, S. St-Pierre, and L. Tepliakova *et al.* Phenotypic impact of regulatory noise in cellular stress-response pathways. *Syst. Synth. Biol.*, 4, 2010.

Chapter 2

What all the Noise is About: The Physical Basis of Cellular Individuality

Daniel A. Charlebois, Mads Kærn. *Can. J. Phys.* (2012).

919

REVIEW / SYNTHÈSE

What all the noise is about: the physical basis of cellular individuality

Daniel A. Charlebois and Mads Kærn

Abstract: Noise has been traditionally viewed as undesirable in biology, resulting in disorder, distortion, and disruption, and ultimately as something that needs to be filtered and removed. More recently, it has been shown that noise can also be beneficial. We briefly review historical developments pertaining to noise in biological physics, and some of the current research in the field of molecular and cellular biophysics.

PACS Nos: 87.18.Tt, 05.40.Jc, 87.10.Mn, 87.10.Rt, 87.16.Yc, 87.23.Kg

Résumé : Traditionnellement, le bruit a été vu comme étant indésirable en biologie, résultant en désordre, distorsion et perturbation, et ultimement comme quelque chose qui doit être filtré et éliminé. Dernièrement, la recherche a démontré que le bruit peut être utile. Dans ce travail nous passons brièvement en revue les développements historiques touchant le bruit dans le domaine de la physique biologique et la recherche actuelle en biophysique moléculaire et cellulaire.

A historical perspective

In 1827 a botanist named Robert Brown observed pollen particles moving about randomly in a fluid [1, 2]. He observed this with all sorts of nonliving materials including minerals, woods, and century-old dried out plants, and concluded that the erratic movements were not a property of living organisms.

Many years later, physicist Georges Gouy [3] conceived that the motion observed by Brown was a result of the irregular thermal fluctuations of the molecules in the liquid. This phenomenon is now called Brownian motion (BM). It was a young Albert Einstein who worked out the now famous result describing the mean-square displacement (in one dimension) of a particle undergoing BM

$$\langle x^2 \rangle = 2Dt \tag{1}$$

where t is the time and D is the diffusion constant. The key idea behind Einstein's equation is that the random motion of a large particle occurs because it is being constantly bombarded by other "invisible" smaller particles in the fluid. Ultimately, Einstein used the concept of such "noise" to predict the existence of atoms [4–7]. Paul Langevin arrived at the same result for the mean-square displacement a few years later using a different approach, namely, a differential equation with a random force term, or as it is now known in statistical physics, a Langevin equation [8].

In the previous framework by Einstein, the position of a Brownian particle undergoing BM is nowhere differentiable and its instantaneous velocity is correspondingly undefined [9]. To avoid this, Ornstein and Uhlenbeck described the velocity of a Brownian particle, instead of the position, as the main random quantity using a Langevin equation [10]

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{1}{\tau}[\mu - x(t)] + c^{1/2}\xi_t \tag{2}$$

where μ is the mean, c the diffusion constant, τ the relaxation time, and ξ_t describes a Gaussian white noise process with zero mean and fixed variance.

In classical biology, genetically identical cells in an identical environment are expected to have identical phenotypes (i.e., observable chemical and physical properties). Any observed difference is attributed to experimental error. However, in 1945 a biophysicist named Max Delbrück found that the number of virus particles released from infected bacteria showed reproducible variations and, accordingly, is best described by a probability distribution rather than a single value [11]. These experiments were inspired by earlier theoretical work where Delbrück wrote down a master equation (ME; see later in text) describing the statistical

Received 10 May 2012. Accepted 25 June 2012. Published at www.nrcresearchpress.com/cjp on 21 August 2012.

D.A. Charlebois. Ottawa Institute of Systems Biology, University of Ottawa, 451 Smyth Road, Ottawa, ON K1H 8M5, Canada; Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, ON K1N 6N5, Canada.
M. Kærn. Ottawa Institute of Systems Biology, University of Ottawa, 451 Smyth Road, Ottawa, ON K1H 8M5, Canada; Department of Cellular and Molecular Medicine, University of Ottawa, 451 Smyth Road, Ottawa, ON K1H 8M5, Canada; Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, ON K1N 6N5, Canada.

Corresponding author: Daniel Charlebois (e-mail: daniel.charlebois@uottawa.ca).

fluctuations in the number of particles for an autocatalytic chemical reaction [12]. When Delbrück solved these equations, he obtained the well-known noise scaling relationship

$$\eta = \frac{1}{\sqrt{N}} \tag{3}$$

in which the magnitude of the fluctuations or noise, η, is equal to the reciprocal of the square root of the number of particles, N, that initiate the reaction. Correspondingly, Delbrück hypothesized that the variation he observed arose from the variation in the number of the initial infection. About four decades later, Spudich and Koshland demonstrated that bacterial cells grown in homogeneous conditions showed characteristic behavioural differences that persisted over their lifespans [13]. They attributed this nongenetic individuality to poissonian fluctuations in the small numbers of generator molecules, and suggested that it may also apply to other processes, such as differentiation and asynchrony of cell cultures.

Noise and biological systems

It is not altogether surprising that variation exists in biological systems when one considers the random (stochastic) nature of biochemical reactions. These reactions are stochastic as they result from collisions between Brownian particles, which lead to the nondeterministic timing of individual reactions and an inherently noisy time evolution of molecular population levels [14, 15]. The relative amplitudes of these fluctuations are effectively averaged out of systems, such as test tubes, with a large number of molecules (see (3)). These systems are appropriately described using deterministic equations.

A noisy system can formally be described using the so-called ME approach [16]. A ME is a set of first-order differential equations governing the time evolution of the probability of a system to occupy each one of a discrete set of states. A ME usually takes the form

$$\frac{\mathrm{d}p_k(t)}{\mathrm{d}t} = \sum_{k'} \{W_{k'\to k} p_{k'}(t) - W_{k\to k'} p_k(t)\} \tag{4}$$

Here, p_k is the time-dependent probability associated with state and $W_{k\to k'}$ is the transitional probability per unit time from k' to k. In this form, it is clear that the ME is a gain–loss equation for the probabilities of the separate states, k. The gain of state k due to the transitions from other states k' is represented by the first term, and the loss due to transitions from k into other states k' is represented by the second term.

The time evolution of the probability distribution for a continuous variable can be described using a Fokker–Plank equation. The Fokker–Plank equation for a single variable x has the form

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}[\gamma(x)p(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[D(x)p(x,t)] \tag{5}$$

where γ is the deterministic drift term and D is the stochastic diffusion term. The Fokker–Plank equation is often used as an approximation of the ME (4). We refer the reader to ref. 16 for a thorough introduction to the subject.

A chemical master equation (CME) accounts for the random timing in the birth and death of individual molecules caused by the nondeterministic timing of individual reactions. As elegant as the CME formalism is, it usually cannot be solved analytically. Consequently, one either has to resort to approximations or simulate every individual state transition occurring in the system. Daniel Gillespie's stochastic simulation algorithm is a Monte Carlo simulation of the very process that the CME describes, and is the gold standard for simulating biochemical reaction systems [14, 15].

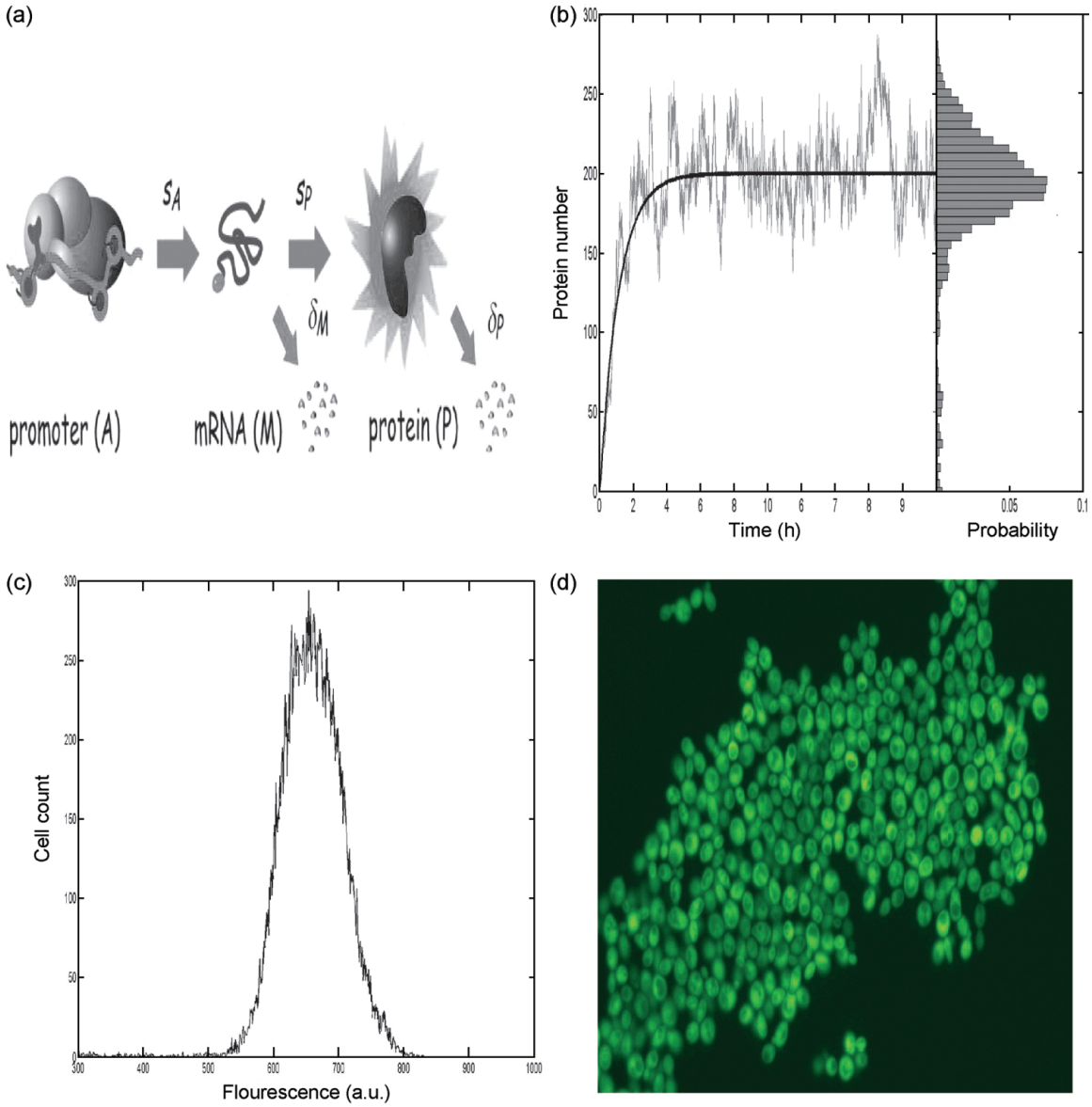
The Gillespie algorithm is often applied to simulate the gene expression process inside living cells. This process is fundamental to all life and is one of the most actively researched topics in science today. Physicists have long been interested in genetics; in his famous book *What is Life?*, Erwin Schrödinger introduced the idea of a gene as an aperiodic structure that stored genetic information in its configuration of covalent chemical bonds [17]. He also predicted that due to the order present in living organisms, DNA must be made up of a large number of atoms to counter the property of increasing randomness with smaller numbers of atoms. Within a decade, Watson and physicist Francis Crick deduced the double helical model for the structure of DNA [18]. Crick subsequently proposed the central hypothesis of molecular biology, namely that the gene expression process involves copying DNA into mRNA (transcription) and the production of a protein from this mRNA template (translation) (Fig. 1*a*) [19, 20]. Importantly, gene expression involves the collisions of small numbers of particles. Usually only one or two copies of DNA are found in a cell along with small numbers of mRNAs and transcription factors [21], and thus gene expression is an inherently noisy process (see (3) and Fig. 1*b*) like the processes observed by Delbrück and Spudich and Koshland [11, 13].

Gene expression can be measured experimentally using fluorescent proteins. More precisely, the gene coding for the fluorescent protein is placed beside a gene of interest such that they are transcribed and translated together. The degree of fluorescence, which indicates the level of gene expression, can then be measured in individual cells by flow cytometry to produce a population "snapshot" in the form of a gene expression distribution (Fig. 1*c*), or by time-lapse microscopy to produce a time series (Fig. 1*d*).

Current Research

The noise in gene expression allows for variation to exist among genetically identical cells in the same environment (for a comprehensive review see refs. 22 and 23). This is of particular interest because it can allow some members of a population to survive while others perish [24] (Fig. 2*a*). For instance, Blake et al. [25] observed that genetically identical yeast populations engineered to have higher noise (more cell-to-cell variation) reproduced faster than low noise populations when exposed to high levels of an antibiotic. Noise in gene expression also allows for "elastic adaptation", which occurs when the noise-generated distribution of a phenotype changes reversibly due to an environmental stress such that the reproductive fitness of a population in the new environment is optimized. For example, populations of yeast cells have been observed to adapt to long-term exposure to a drug

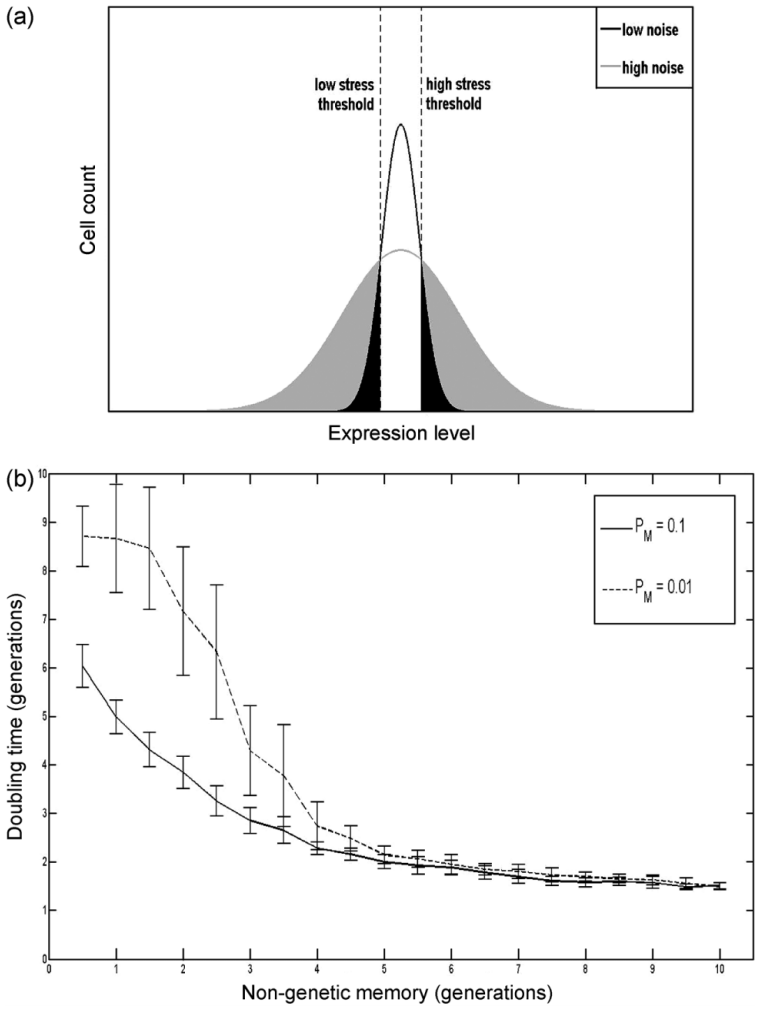
Fig. 1. Gene expression is a stochastic process. (*a*) A simple two-step model of gene expression. The schematic shows the synthesis of mRNA (M) from a gene with an active promoter (A) at a rate S_A, and the synthesis of protein (P) from an M template at a rate S_P, and the decay of M and P molecules at rates δ_M and δ_P, respectively. Reprinted with permission from (Scott et al. *Chaos*, **16**, 026107-2, (2006)). Copyright 2006, American Institute of Physics. (*b*) Time series of protein number generated by deterministic (solid black line) and stochastic (gray line) simulations. The histogram in the right-hand panel corresponds to the stochastic simulation and shows the probability that a cell will have a given intracellular protein level. Reprinted by permission from Macmillan Publishers Ltd. (Kaern et al. *Nat. Rev. Genet.* **6**, 453, copyright 2005. (*c*) Experimental green fluorescent protein (GFP) distribution for a clonal population of budding yeast obtained via flow cytometry (unpublished data). (*d*) GFP expression for a clonal population of budding yeast obtained using a microfluidics device (unpublished data).



by shifting the gene expression distribution in the direction that minimizes the impact of a drug [26]. When the drug is removed, the shift in gene expression can revert back to the distribution observed before the drug was applied [26]. This phenomenon has been attributed to nongenetic memory, as opposed to genetic memory where a mutation in the DNA would result in a permanent shift in gene expression. The term "nongenetic memory" can generally be defined as any mechanism that produces an enduring phenotype without altering the DNA sequence.

Genetic networks can store nongenetic memory in two or more discrete, stable states of network activity (see ref. 27

Fig. 2. Gene expression noise confers survival in clonal cell populations. (*a*) Schematic illustration of distributions for a low- and high-noise population. A greater number of cells in the high-noise population express above (cell survival) and below (cell death) the high- and low-stress thresholds, respectively. Reproduced with permission from (Fraser et al. *Mol. Microbiol.* **71**, 1335 (2009)). In the case shown here, the high-noise population has a higher fitness than the low-noise population when the stress is high, vice versa when the stress is low. (*b*) Effect of nongenetic memory and probability of mutation (P_M) per generation on the time for a simulated cancer cell population undergoing prolonged drug treatment to double. Note that the doubling time is more or less unaffected by P_M when the nongenetic memory is roughly above four generations and that in both cases a drug-resistant cell population develops. Reprinted figure with permission from (Charlebois et al. *Phys. Rev. Lett.* **107**, 218101-4 (2011)). Copyright 2011 by the American Physical Society.



for a review). For instance, in yeast, nongenetic memory can be enhanced by reducing the rate of stochastic transitions between two stable gene expression states [28]. In human cells, Brock et al. [29] proposed that nongenetic memory conferring temporary drug resistance contributes to tumour development by increasing the chance that some cells acquire a mutation conferring permanent immunity to the treatment regime. Nongenetic memory can also be stored in the lifetime of the gene expression fluctuations. That is, the lower the frequency of the noise the higher the level of the nongenetic memory, as the previous state is "remembered" by the cell for a longer period of time than at higher frequency noise. This was shown using an Ornstein and Uhlenbeck process [10] (see (2)) to be sufficient for the development of long-term drug resistance, independent of genetic memory conferring resistance [30] (Fig. 2*b*). This hypothesis is currently being investigated experimentally.

Conclusion

This is a new era for biology, one where more and more physicists are playing leading roles and driving the field to become more quantitative. The mathematical models being developed are helping to better explain the data gathered in the laboratory and to predict novel behaviour. In particular, stochastic models are being used increasingly in preference to deterministic models to describe biochemical networks and elucidate dynamics at the single-cell level [21]. Due to

the randomness inherent in living systems, an understanding of the source of this randomness and its effects is of fundamental importance. Fortunately, due to the foundations laid by early physicists, and the familiarity of many physical scientists today with the theory of stochastic processes, we can obtain a deeper understanding of biological systems.

Acknowledgments

The authors would like to acknowledge D. Jedrysiak for carrying out the experiments together with DC to obtain the data for Figs. 1*c* and 1*d*. We would also like to thank N. Abdennur for helpful discussions.

References

1. R. Brown. Philos. Mag. **4**, 161 (1828).
2. R. Brown. Philos. Mag. **6**, 161 (1829).
3. L.-G. Gouy. C.R. Hebd. Seances Acad. Sci. **CIX**, 102 (1889).
4. A. Einstein. Ann. Phys. **322**, 549 (1905). doi:10.1002/andp.19053220806.
5. A. Einstein. Ann. Phys. **324**, 371 (1906). doi:10.1002/andp.19063240208.
6. A. Einstein. Angew. Phys. Chem. **13**, 41 (1907).
7. A. Einstein. Zeitschrift für Elektrochemie und Angewandte Physikalische Chemie, **14**, 235 (1908). doi:10.1002/bbpc.19080141703.
8. P. Langevin. C.R. Acad. Sci. (Paris), **146**, 530 (1908).
9. W. Horsthemke and R. Lefever. Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology. Springer-Verlag, Berlin. 2006.
10. G. Uhlenbeck and L. Ornstein. Phys. Rev. **36**, 823 (1930). doi:10.1103/PhysRev.36.823.
11. M. Delbrück. J. Bacteriol. **50**, 131 (1945).
12. M. Delbrück. J. Chem. Phys. **8**, 120 (1940). doi:10.1063/1.1750549.
13. J.L. Spudich and D.E. Koshland, Jr. Nature, **262**, 467 (1976). doi:10.1038/262467a0. PMID:958399.
14. D.T. Gillespie. J. Comput. Phys. **22**, 403 (1976). doi:10.1016/0021-9991(76)90041-3.
15. D.T. Gillespie. J. Phys. Chem. **81**, 2340 (1977). doi:10.1021/j100540a008.
16. N.G. van Kampen. Stochastic Processes in Physics and Chemistry. North-Holland Personal Library, Amsterdam, North-Holland. 1992.
17. E. Schrödinger. What is Life?: The Physical Aspect of the Living Cell, Based on lectures delivered under the auspices of the Dublin Institute for Advanced Studies at Trinity College, Dublin. February 1943. 1944.
18. J.D. Watson and F.H.C. Crick. Nature, **171**, 737 (1953). doi:10.1038/171737a0. PMID:13054692.
19. F.H.C. Crick. Symp. Soc. Exp. Biol. **XII**, 138 (1958).
20. F. Crick. Nature, **227**, 561 (1970). doi:10.1038/227561a0. PMID:4913914.
21. D.J. Wilkinson. Nat. Rev. Genet. **10**, 122 (2009). doi:10.1038/nrg2509. PMID:19139763.
22. M. Kaern, T.C. Elston, W.J. Blake, and J.J. Collins. Nat. Rev. Genet. **6**, 451 (2005). doi:10.1038/nrg1615. PMID:15883588.
23. M.S. Samoilov, G. Price, and A.P. Arkin. Sci. STKE, **2006**, re17 (2006). doi:10.1126/stke.3662006re17. PMID:17179490.
24. D. Fraser and M. Kaern. Mol. Microbiol. **71**, 1333 (2009). doi:10.1111/j.1365-2958.2009.06605.x. PMID:19220745.
25. W.J. Blake, G. Balázsi, M.A. Kohanski, F.J. Isaacs, K.F. Murphy, Y. Kuang, C.R. Cantor, D.R. Walt, and J.J. Collins. Mol. Cell, **24**, 853 (2006). doi:10.1016/j.molcel.2006.11.003. PMID:17189188.
26. D. Zhuravel, D. Fraser, S. St-Pierre, L. Tepliakova, W.L. Pang, J. Hasty, and M. Kaern. Syst. Synth. Biol. **4**, 105 (2010). doi:10.1007/s11693-010-9055-2. PMID:20805931.
27. S. Huang. Prog. Biophys. Mol. Biol. Available online 10 May 2012. doi:10.1016/j.pbiomolbio.2012.05.001.
28. M. Acar, A. Becskei, and A. van Oudenaarden. Nature, **435**, 228 (2005). doi:10.1038/nature03524. PMID:15889097.
29. A. Brock, H. Chang, and S. Huang. Nature, **10**, 336 (2009).
30. D.A. Charlebois, N. Abdennur, and M. Kaern. Phys. Rev. Lett. **107**, 218101 (2011). doi:10.1103/PhysRevLett.107.218101. PMID:22181928.

Chapter 3

An Accelerated Method for Simulating Population Dynamics

Daniel A. Charlebois, Mads Kærn. *Comm. Comput. Phys.* (2013).

Commun. Comput. Phys.
doi: 10.4208/cicp.130612.121012a

Vol. **14**, No. 2, pp. 461-476
August 2013

An Accelerated Method for Simulating Population Dynamics

Daniel A. Charlebois[1,2,*] and Mads Kærn[1,2,3]

[1] *Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, Ontario K1N 6N5, Canada.*
[2] *Ottawa Institute of Systems Biology, University of Ottawa, 451 Symth Road, Ottawa, Ontario K1H 8M5, Canada.*
[3] *Department of Cellular and Molecular Medicine, University of Ottawa, 451 Symth Road, Ottawa, Ontario K1H 8M5, Canada.*

Received 13 June 2012; Accepted (in revised version) 12 October 2012

Available online 4 January 2013

Abstract. We present an accelerated method for stochastically simulating the dynamics of heterogeneous cell populations. The algorithm combines a Monte Carlo approach for simulating the biochemical kinetics in single cells with a constant-number Monte Carlo method for simulating the reproductive fitness and the statistical characteristics of growing cell populations. To benchmark accuracy and performance, we compare simulation results with those generated from a previously validated population dynamics algorithm. The comparison demonstrates that the accelerated method accurately simulates population dynamics with significant reductions in runtime under commonly invoked steady-state and symmetric cell division assumptions. Considering the increasing complexity of cell population models, the method is an important addition to the arsenal of existing algorithms for simulating cellular and population dynamics that enables efficient, coarse-grained exploration of parameter space.

PACS: 87.10.Mn, 87.10.Rt, 87.16.Yc, 87.17.Ee

Key words: Accelerated stochastic simulation algorithm, constant-number Monte Carlo, gene expression, population dynamics and fitness.

1 Introduction

Cell populations are heterogeneous entities. Part of this heterogeneity arises from the stochasticity inherently present in the process of gene expression, which can result in

*Corresponding author. *Email addresses:* daniel.charlebois@uottawa.ca (D. A. Charlebois), mkaern@uottawa.ca (M. Kærn)

significant variability even among cells with identical genotypes in identical environments [7, 15, 16, 20, 26, 29, 35]. This variability can in turn have significant impact on the overall reproductive fitness of a cell population [1, 2, 5, 9, 41, 42].

In some cases it is possible to derive analytical solutions for the statistical characteristics of gene expression for simple models (e.g., [25, 27, 30, 31, 36]). However, for more biologically realistic models, these characteristics are available only through numerical simulations. To permit investigations, we previously developed an algorithm for the stochastic simulation of heterogeneous population dynamics at a single-cell resolution [4]. This Population Dynamics Algorithm (PDA) combines the Gillespie stochastic simulation algorithm (SSA) [10, 11] to simulate gene expression in individual cells and a constant-number Monte Carlo (MC) method [17, 21, 22, 28, 34] for simulating population dynamics.

To benchmark the performance and accuracy of the method, we compared simulation results from the PDA with steady-state and time-dependent analytical solutions for several scenarios, including steady-state and time-dependent gene expression, and the effects on population heterogeneity of cell growth, division, and DNA replication [4]. Additionally, we used the PDA to model gene expression dynamics within bet-hedging cell populations during their adaption to environmental stress. Later, in [5] the PDA and analytical solutions developed for determining the first-passage time dependent fitness of a cell population exposed to a drug over a single generation were found to be in agreement. We refer the reader to these papers for details on the analytical work. These comparisons demonstrated that the PDA accurately captures how complex biological features influence gene expression and population dynamics. However, simulation runtimes can be extensive when the biochemical reaction kinetics that take place within a large number of individual cells are simulated using conventional MC approaches.

To address this problem, we have developed an accelerated method for simulating population dynamics (AMSPD). We first demonstrate that the AMSPD algorithm is numerically accurate and provides a significant speedup compared to the PDA. We then use the AMSPD to perform a parameter scan of a simple model for the development of non-genetic drug resistance to illustrate that it can be advantageous to use the AMSPD and PDA in combination to find an optimal balance between efficiency and accuracy.

2 Algorithm

In this section we present the AMSPD algorithm. The stochastic simulation algorithm [10, 11] and the constant-number MC method [17, 21, 22, 28, 34] are also described for completeness.

2.1 Accelerated method for simulating population dynamics

The first step in the AMSPD algorithm is to generate a single stationary time series (such that the moments of the corresponding distribution are not changing) for each biochemical variable in the system using an appropriate simulation method (e.g. the SSA [10, 11]

D. A. Charlebois and M. Kærn / Commun. Comput. Phys., **14** (2013), pp. 461-476 463

– see Section 2.2) and store the values of the time series in an array of length N. Each row of the array corresponds to a separate biochemical variable. It is not uncommon in simulation studies to assume that one or more biochemical species are in a steady-state (e.g., [1, 19, 31, 37]). The AMSPD algorithm then employs this time series to simulate the gene expression and fitness dynamics of a population of cells (Fig. 1a). Specifically, at the start of the simulation each cell of the initial population is assigned a positive integer (randomly generated from a uniform distribution on the interval [1,N]), which corresponds to its column 'position' in the array. During a given sampling interval, each cell progress through the pre-generated time series values stored in the rows of the array. Each time a cell's internal clock is incremented by a pre-specified value time increment Δt, so is its column position in the array (note that the pre-generated time series values were obtained from sampling the SSA simulation using the same Δt). If a cell happens to

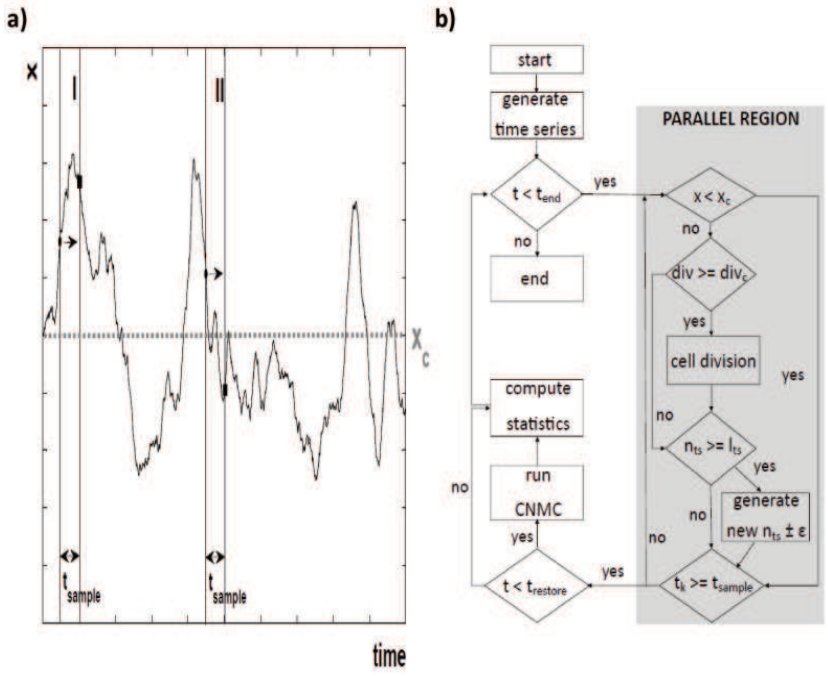


Figure 1: The accelerated method for simulating population dynamics (AMSPD) algorithm. (a) Schematic showing how individual cells are simulated by the AMPSD algorithm. After the cells are randomly assigned positions on the time series (dots), their positions are incremented until the end of the sampling interval t_{sample} is reached (squares). If a reproductive stress is not incorporated into the simulations, then mother cells simply reproduce at a specified rate. However, if the fitness of the cells depends on the level of a particular biochemical variable, then cells can only reproduce if this variable remains above a specified threshold. For instance, in region I, the gene expression value x remains above a critical threshold x_c during the sampling interval and therefore the cell is able to reproduce during the entire interval. In region II, the gene expression value of the cell falls below x_c and is therefore flagged and unable to reproduce after this point. (b) Flow diagram of the AMSPD algorithm presented in the main text (see Table 1 for AMSPD variable and parameter descriptions).

reach the last column before the end of the sampling interval, then the cell is randomly assigned a new position on the array within some error ϵ from the last value, for each biochemical variable in the system. There is a tradeoff between accuracy and efficiency as ϵ is varied (data not shown). For smaller values of ϵ, simulation runtimes are longer but the results are more accurate, and vice versa for larger values of ϵ. In this study we use an ϵ of 10 or lower. For simulations involving the presence of a stressor (e.g. a drug), a biochemical variable of interest (e.g. protein concentration) can be used to determine cellular fitness. For example, if the value of this variable falls below a critical threshold then the cell can be flagged and its biochemical variables no longer simulated nor the cell able to reproduce (Fig. 1a). More elaborate fitness functions than a simple step function can also be incorporated into the AMSPD algorithm. For example, a 'softer' fitness threshold can be modeled using a Hill function with low values of the Hill coefficient n (e.g., $n=2-4$).

Once the end of the sampling interval is reached for all the cells in the population, the constant-number MC method [17,21,22,28,34] is used to keep the number of cells in the population fixed (see Section 2.3). If a cell divides during the sampling interval the concentration of each variable is assumed to remain constant. This is equivalent to assuming that the cellular contents are equally partitioned into equal volumes or that the transient time to steady-state is negligible. This assumption has been used in several other studies (e.g., [3,5,6,18,31]). The daughter cell is then randomly assigned a position on the time series within some tolerance ϵ of each of the mother cell's biochemical variables at the moment of division.

The AMSPD algorithm can be expressed by the flow diagram (Fig. 1b) and the sub-

Table 1: AMSPD parameters and variables.

Parameter/Variable	Description
div and div_c	Division variable and corresponding threshold at which division occurs.
ϵ	Error term for assignment or re-assignment of position on the stationary time series.
l_{ts}	Length of the stationary time series.
$NC_{daughter}$	Number of daughter cells born in a given sampling interval.
$NC_{population}$	Total number of cells in the population.
n_{ts}	Position on the stationary time series.
t	Global simulation time.
t_{end}	Simulation end time.
t_k	Local or cell specific simulation time.
t_{sample}	Sampling interval for statistics.
$t_{restore}$	Interval between population size restores.
x and x_c	Biochemical variable of interest and the corresponding threshold below which cells are unable to reproduce.

D. A. Charlebois and M. Kærn / Commun. Comput. Phys., **14** (2013), pp. 461-476 465

sequent pseudocode (Algorithm 2.1). In the pseudocode the AMSPD parameters and variables are defined as follows: div is the division parameter (generally time or volume) and div_c the corresponding threshold (if applicable) at which division occurs, l_{ts} the length (number of points) of the time series, t the global simulation time, t_{end} the user specified simulation end time, t_k the local or cell specific simulation time, t_{sample} the sampling interval for statistics, $t_{restore}$ the interval between population size restores, $NC_{daughter}$ the number of daughter cells, $NC_{population}$ the total number of cells in the population, n_{ts} the position on the time series, x a biochemical variable of interest and x_c the corresponding threshold (if applicable) below which cells are unable to reproduce. The AMSPD parameters and variables for pseudocode the are summarized in Table 1.

Algorithm 2.1: AMSPD

```
Generate a stationary time series for each variable using the SSA (see Algorithm 2.2)
Randomly obtain an initial n_ts for each cell
while t < t_end do
   begin parallel region
   for all NC_population such that t_k < t_sample do
      Update t_k and div
      if x ≥ x_c then
         Update n_ts and x
         if n_ts ≥ l_ts then
            Randomly generate new n_ts (until x(n_ts) within ±ε of x(l_ts)) and update x
         end if
         if div ≥ div_c then
            Execute cell division
            Increment NC_daughter
         end if
      end if
   end for
   end parallel region
   Update t and t_sample
   Execute constant-number MC (see Algorithm 2.3)
   Compute statistics
end while
```

2.2 SSA

In the Direct Method Gillespie SSA [10, 11], M chemical reactions with rate constants $c_1,\cdots,c_M$ among N chemical species $X_1,\cdots,X_N$, are simulated one reaction event at a time. The next reaction to occur Ω and its timing Γ are determined by calculating M reaction propensities $a_1,\cdots,a_M$, given the current number of molecules of each of the N chemical species, to obtain an appropriately weighted probability for each reaction. It can be implemented via the following pseudocode:

Algorithm 2.2: SSA

1: **if** $t<t_{end}$ and $\alpha_0=\sum_{v=1}^{M}a_v\neq 0$ **then**
2: **for** $v=1,M$ **do**
3: Calculate α_v
4: **end for**
5: $\alpha_0=\sum_{v=1}^{M}a_v$
6: Generate uniformly distributed random numbers (r_1,r_2)
7: Determine when $(\Gamma=\ln(1/r_1)/\alpha_0)$ and which $(\min\{\Omega \mid \alpha_\Omega \geq r_2\alpha_0\})$ reaction will occur
8: Set $t=t+\Gamma$
9: Update $X_1,\cdots,X_N$
10: **end if**

2.3 Constant-number Monte Carlo method

The constant-number MC method [17,21,22,28,34] permits the statistically accurate simulation of a representative sample of an exponentially growing cell population. In this implementation of the method, all the daughter cells born since the last update $NC_{daughter}$ are stored and simulated using a separate array from the mother cells. To avoid simulating the daughters of daughter cells, the interval between population size updates $t_{restore}$ is chosen such that mother cells divide at most once, and daughter cells not at all, during a particular $t_{restore}$ interval. The constant-number MC method can be represented by the following pseudocode:

Algorithm 2.3: CNMC

1: **if** $t>t_{restore}$ and $NC_{daughter}\geq 1$ **then**
2: **for all** $NC_{daughter}$ **do**
3: Randomly select mother cell
4: Replace mother cell with oldest available daughter cell
5: **end for**
6: **end if**

3 Numerical results and discussion

To evaluate the accuracy and the speedup of the accelerated method, we compare simulation results obtained using the AMSPD algorithm to those obtained using the previously validated PDA [4].

For benchmarking we first examine a univariate model of protein production and decay (Section 3.1). Then we consider a multivariate model of gene expression where mRNA and protein production and decay are both incorporated (Section 3.2). To further

D. A. Charlebois and M. Kærn / Commun. Comput. Phys., **14** (2013), pp. 461-476 467

benchmark the algorithm when the reproductive fitness of the cell population in the presence of a drug is incorporated, we reproduce the results from [5]. In this work, we used an Ornstein-Uhlenbeck (OU) model to simulate gene expression (Section 3.3). Finally, in Section 3.4, we demonstrate that the accelerated method can enable efficient and numerically accurate coarse-grained exploration of the parameter space corresponding to a population model. Specifically, we use the AMSPD algorithm to perform a scan of the parameter space of the OU model of gene expression, and compare the resulting fitness landscape of the population with results obtained using the PDA.

Both algorithms were implemented in Fortran 90 and executed on an IBM with 2 quad-core processors (1.86GHz cores) and 2.0GB of RAM. All units unless indicated otherwise are arbitrary. Statistics were estimated from 10 realisations of populations consisting of 1000 cells unless otherwise indicated.

3.1 Univariate model

We consider gene expression as a birth-death process modeled by the following equations

$$\varnothing \xrightarrow{k_P} P, \tag{3.1}$$

$$P \xrightarrow{\delta_P} \varnothing, \tag{3.2}$$

where P is a protein produced in a single step at a rate k_P (Eq. (3.1)), and decays at a rate δ_P (Eq. (3.2)).

We first model cell division without incorporating cellular volume, that is each cell divides once its cellular 'clock', or time since last division div, reaches or exceeds a predefined cell division time div_c. In this case, excellent agreement is found between the AMSPD algorithm and PDA (Fig. 2a). The runtime of the AMSPD algorithm is shown in Fig. 2b. When the time to generate the time series for the AMSPD algorithm is incorporated into the runtime of the AMSPD algorithm, then the AMPSD's runtime increases linearly with k_P. However, if the time to generate the time series is not factored into AMSPD's runtime, then the runtime of the AMSPD algorithm does not vary with k_P since a time series of the same length is used in each of the simulations. This applies for instance if a time series that was previously generated can be used again, for example, if the simulation is to be repeated, or if a variable assumed not to affect gene expression (such as div_c) is changed. The AMSPD algorithm is found to be significantly faster as the rate of protein production is increased (Fig. 2c). For instance, when k_P is 1 the AMSPD algorithm is three times faster than the PDA. However, when k_P is increased to 100, the speedup is sixty times. If the time to generate the time series for the AMSPD algorithm is not factored into the speedup calculation, then speedups of several hundred times are observed. We attribute the speedup to the fact that the AMSPD algorithm does not simulate every reaction occurring inside each cell of the population as the PDA does. Rather, the AMSPD algorithm performs a random access lookup in an array containing the values of the time series.

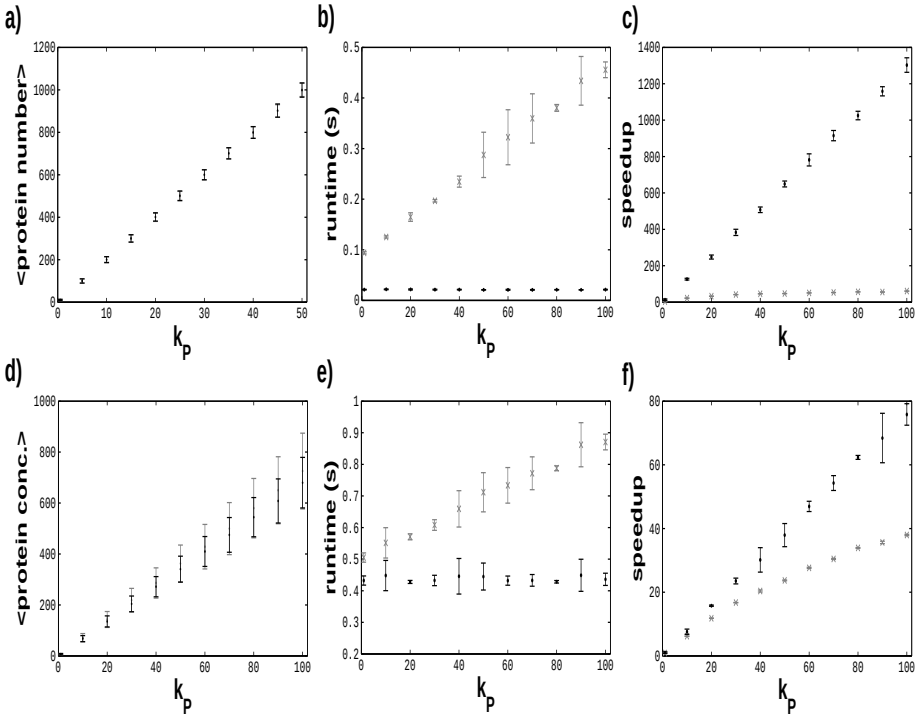


Figure 2: Comparison of accuracy and performance of the AMSPD algorithm and PDA [4] for a birth-death model of gene expression. Panels (a)-(c) correspond to simulation results for volume independent cell division and (d)-(f) volume dependent cell division. (a) and (d) show the average steady-state protein numbers and concentrations, respectively, as a function of the rate of protein production k_P for the AMSPD algorithm (gray) and the PDA (black). (b) and (e) show the runtime of the AMSPD simulation. (c) and (f) show the speedup of the AMSPD algorithm, when compared to the runtime of the PDA, as a function of the rate of protein production k_P. Gray x's in (b) and (e), and in (c) and (f), are the results obtained when the time to produce the gene expression time series is incorporated into the AMSPD's runtime and the speedup calculation, respectively. Black dots in (b) and (e), and in (c) and (f), are the results obtained when the time to produce the gene expression time series is not incorporated into the AMSPD's runtime and the speedup calculation, respectively. Simulations were started from steady-state ($p^s = k_P/\delta_P$), the initial time since last division div drawn from a uniform distribution [0,div], and the protein time series generated by the AMPSD algorithm contained 10^4 values. The parameters were set to $\delta_P = 0.01$, $\epsilon = 10$, $div_c = 100$, and $t_{end} = 1000$.

The incorporation of changing cellular volume throughout the cell cycle into simulations can be important when concentration dependent, rather than absolute number, effects are to be considered (e.g., [36, 40]). In a more complex model, we describe cell growth by an exponential growth law [4, 12, 13]

$$V_k(t_{div}) = V_0 2^{(t_{div}/\tau_0)}, \tag{3.3}$$

where V_0 is the cell volume at the time of its birth, and τ_0 is the interval between volume doubling. Cell division occurs when the cell volume reaches $2V_0$.

D. A. Charlebois and M. Kærn / Commun. Comput. Phys., **14** (2013), pp. 461-476 469

Again, there is excellent numerical agreement between the two simulation methods (Fig. 2d), and a significant speedup when using the AMSPD algorithm (Fig. 2f). For example, when k_P is 1 the speedup is thirteen times. However, when k_P is increased to 100, the AMSPD algorithm is about forty times faster than the PDA, and seventy five times faster when the time to generate the time series for the AMSPD algorithm is not included in the speedup calculation. As in the previous case (Fig. 2b), AMPSD's runtime either increases linearly with k_P or does not vary with k_P, depending on whether the time to generate the time series for the AMSPD algorithm is or is not incorporated into the runtime, respectively (Fig. 2e). The runtimes shown in Fig. 2e are longer than those in Fig. 2b due to the incorporation of cellular volume dynamics.

The results presented in this section indicate that the AMSPD algorithm can accurately simulate models that incorporate a univariate description of biochemical dynamics occurring inside of growing and dividing cells with a significant reduction in runtime when compared to the PDA.

3.2 Multivariate model

The previous section considered a univariate analysis. However, it is in the multiple variate scenario that the AMSPD algorithm is likely to be employed since any model incorporating a biologically realistic level of detail will require more than one variable. Due to the nonlinearity and dimensionality of the corresponding system of equations, a computational approach rather than an analytical one will generally be required to obtain solutions. However, as the dimensionality of the system increases so does the computation time along with the need for an accelerated simulation approach.

In order to benchmark the AMSPD algorithm in the multivariate case, we use a slightly more complex model where gene expression is simulated as a two-step process described by the following equations

$$D \xrightarrow{k_M} D + M, \tag{3.4}$$

$$M \xrightarrow{k_P} M + P, \tag{3.5}$$

$$M \xrightarrow{\delta_M} \varnothing, \tag{3.6}$$

$$P \xrightarrow{\delta_P} \varnothing, \tag{3.7}$$

where Eqs. (3.4)-(3.5) respectively describe the transcription and translation processes. The degradation of mRNA M and protein P are accounted for by Eqs. (3.6)-(3.7), respectively.

As in the univariate case, we consider volume independent (Fig. 3a-3d) and volume dependent cell division (Fig. 3e-3h). Excellent agreement is found between the algorithms for mRNA and protein steady-states (Fig. 3a and 3e, and Fig. 3b and 3f, respectively). AMPSD's runtime again either increases linearly with k_P or does not vary with k_P, depending on whether the time to generate the time series for the AMSPD algorithm

470 D. A. Charlebois and M. Kærn / Commun. Comput. Phys., **14** (2013), pp. 461-476

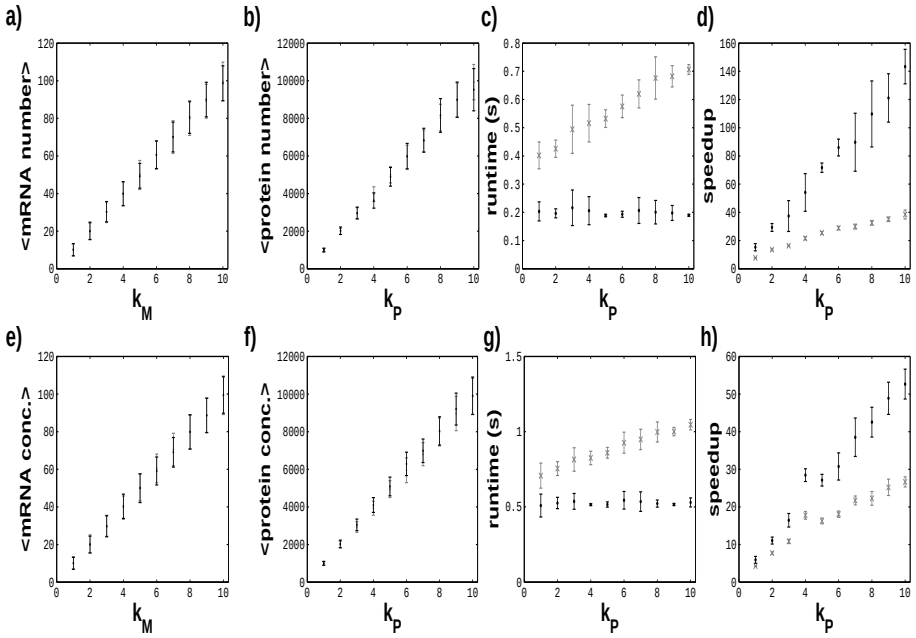


Figure 3: Comparison of accuracy and performance of the AMSPD algorithm and PDA [4] for a two-step model of gene expression. Panels (a)-(d) correspond to simulation results for volume independent cell division and (e)-(h) volume dependent cell division. (a) and (e) show the steady-state mRNA numbers and concentrations, respectively, as a function of the rate of mRNA production k_M for the AMSPD algorithm (gray) and the PDA (black). (b) and (f) show the average steady-state protein numbers and concentrations, respectively, as a function of the rate of protein production k_P for the AMSPD algorithm (gray) and the PDA (black). (c) and (g) show the runtime of the AMSPD simulation. (d) and (h) show the speedup of the AMSPD algorithm, when compared to the runtime of the PDA, as a function of the rate of protein production k_P. Gray x's in (c) and (g), and in (d) and (h), are the results obtained when the time to produce the gene expression time series is incorporated into the AMSPD's runtime and the speedup calculation, respectively. Black dots in (c) and (g), and in (d) and (h), are the results obtained when the time to produce the gene expression time series is not incorporated into the AMSPD's runtime and the speedup calculation, respectively. Simulations were started from steady-state ($M^s = k_M/\delta_M$ and $P^s = k_M k_P/\delta_M\delta_P$), the initial time since last division div drawn from a uniform distribution [0,div], and the protein time series generated by the AMPSD algorithm contained 10^4 values. The parameters were set to $k_M = 1$ (when k_P was varied), $k_P = 1$ (when k_M was varied), $\delta_M = 0.1$, $\delta_P = 0.01$, $\epsilon = 10$, $div_c = 100$, and $t_{end} = 1000$.

is or is not incorporated into the runtime, respectively (Fig. 3c and 3g). The AMSPD algorithm is significantly faster especially when the rate of protein production was high. For instance, considering volume independent division when k_P is 10, the AMSPD algorithm is roughly forty times faster than the PDA, and one hundred and forty times faster when the time to generate the time series for the AMSPD algorithm is not factored into the speedup calculation (Fig. 3d). When volume dependent division is incorporated and k_P is 10, the AMSPD algorithm is twenty five times faster than the PDA, and fifty five times faster when the time to generate the time series for the AMSPD algorithm is not included in the speedup calculation (Fig. 3h).

D. A. Charlebois and M. Kærn / Commun. Comput. Phys., **14** (2013), pp. 461-476 471

Together the results in this section demonstrate that the AMSPD algorithm can be extended to accurately and efficiently simulate multivariate biochemical networks when cell growth and division are incorporated into the model.

3.3 Environmental stress

In this section we use the AMSPD algorithm to reproduce the results obtained in [5] using the PDA to simulate the reproductive fitness of a cell population exposed to a drug. In that study, gene expression in individual cells was simulated as an OU process to capture the effect of fluctuations in gene expression x on the development of drug-resistant cell populations. It was found that if the fluctuation relaxation time scale in gene expression (non-genetic memory) was sufficiently long then drug resistant population could emerge independently of genetic mutations (genetic memory) [5]. The range of values for the non-genetic memory parameter for which drug resistance emerged independently of mutations was in agreement with 'mixing time' (defined as the lag where the autocorrelation function has decreased by 50%) results found experimentally in a human lung cancer cells [33].

The OU process can be described by the following Langevin equation

$$\frac{dx(t)}{dt} = \frac{1}{\tau}(\mu - x(t)) + c^{1/2}\xi_t, \tag{3.8}$$

where c and τ are the diffusion constant and the relaxation time, respectively, and ξ_t is Gaussian white noise ($\langle \xi_t \rangle = 0$, $\langle \xi_t \xi_{t'} \rangle = \delta(t-t')$) [38]. Without loss of generality, we set the mean μ equal to zero and use the fluctuation time-scale τ to model the time-scale of non-genetic memory.

As in [5], 'microfitness' $w(x)$ describes the effect of a drug on the reproductive fitness of individual cells with a given level of expression. For simplicity, in this model microfitness is described using a Heaviside step function, such that a cell is unable to reproduce if their expression level is below a critical value, $w(x<x_c)=0$, and unaffected by the drug otherwise, $w(x \geq x_c)=1$. For the OU process with a mean of zero, 50% of the cell population is instantaneously unable to reproduce when the drug is applied at generation zero. The 'macrofitness' W, or reproductive fitness of the cell population, is here calculated by dividing the number of cells that reproduced during a specified sampling interval by the total number of cells (held fixed by the constant-number MC method) in the population. Since we have set the cell division time such that each cell can only divide once during a given sampling interval, the maximum macrofitness that the cell population can attain is one.

Fig. 4 illustrates that as the number of generations increase, the cell population will reach a steady-state level of fitness. The level of drug resistance that the cell population develops depends on the degree of non-genetic memory. When the non-genetic memory is sufficiently low (i.e. $\tau <= 1$) the population completely succumbs to the drug, and

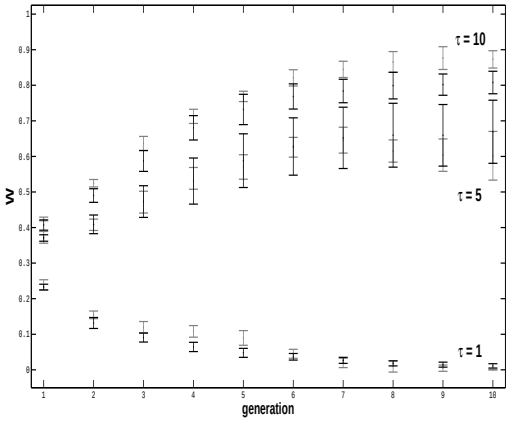


Figure 4: Comparison of accuracy of the AMSPD algorithm (gray) and PDA (black) [4] for a model capturing the effect of non-genetic memory τ on drug resistance at various timescales [5]. The reproductive fitness of the cell population (macrofitness) W as a function of generation is plotted for various values of τ. Simulations were started from the steady-state OU distribution (with mean $\mu=0$ and variance $\sigma^2=c\tau/2=1$), the initial time since last division div was drawn from a uniform distribution [0,div], and the protein time series generated by the AMPSD algorithm contained 10^6 values. The parameters were set to $\epsilon=1$ and $div_c=1$, and scaled by div_c. The threshold below which cells were unable to reproduce x_c was set to μ.

when non-genetic memory is sufficiently high ($\tau>1$) the macrofitness of the cell population increases (Fig. 4). This phenomenon occurs because higher values of τ have a higher probability of enabling individual cells to reside for sufficiently long times in advantageous regions of the fitness landscape, such that they can reproduce before succumbing to the effects of the drug. These results are in quantitative agreement with results previous obtained using the PDA algorithm [5] and demonstrate that the AMSPD algorithm can be used to simulate more biologically complex scenarios such as the effect of stress and noisy gene expression on the reproductive fitness of a cell population.

3.4 Parameter scans

In order to investigate the dynamics of a given population model, one can perform simulations across the corresponding parameter space. However, the use of a more accurate method, such as the PDA, to perform these simulations can prohibit a comprehensive parameter scan due to its computationally intensive nature. The use of an approximate method such as the AMSPD algorithm can enable an efficient preliminary exploration of the parameter space.

Using the OU model of gene expression and the framework presented in Section 3.3 to capture fitness dynamics, we simulate the reproductive fitness of a cell population after being exposed to a stress for 10 generations.

In Section 3.3 the variance of the OU distribution was fixed to 1 by varying the diffusion constant c as the relaxation time τ was increased. Here, τ and c are varied indepen-

D. A. Charlebois and M. Kærn / Commun. Comput. Phys., **14** (2013), pp. 461-476 473

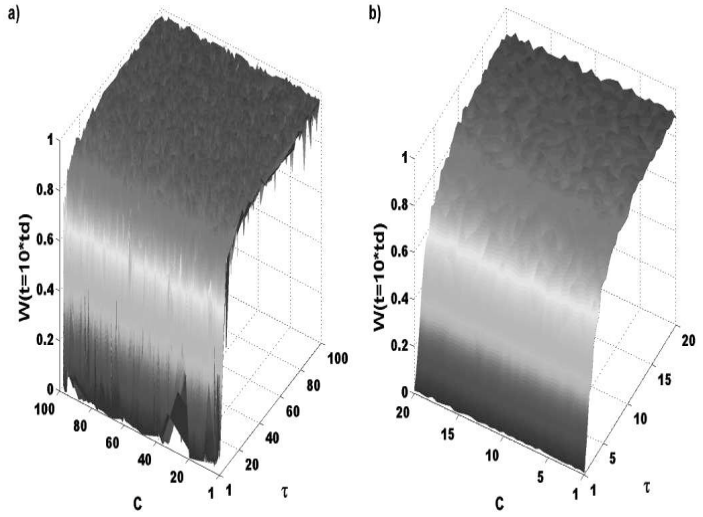


Figure 5: Parameter scans of an OU model of gene expression for the development of drug resistance. (a) Stochastic simulations carried out using the AMSPD algorithm. Here, 10^4 parameter combinations for the relaxation time τ and the diffusion constant c were generated using a Latin hypercube sampling method [23,24], in order to determine the reproductive fitness of the cell population (macrofitness) W after 10 generations. (b) A systematic scan of a region of the parameter space shown in (a) using the more accurate PDA [4]. Simulations were started from the steady-state OU distribution (with mean $\mu=0$ and variance $\sigma^2=c\tau/2$), the initial time since last division div was drawn from a uniform distribution [0,div], and the protein time series generated by the AMPSD algorithm contained 10^6 values. The parameters were set to $\epsilon=1$ and $div_c=1$, and scaled by div_c. The threshold below which cells were unable to reproduce x_c was set to μ.

dently to further examine the role that these parameters have on fitness. Using a Latin hypercube sampling method [23, 24], we generate 10^4 different parameter combinations and simulate the population dynamics using AMSPD (Fig. 5a). Based on the results of these simulations we then identify a region of parameter space of interest (reduced by a factor of 5 compared to the original parameter space), namely where the macrofitness of the cell population changes rapidly, and then perform the simulations using the PDA (Fig. 5b). The parameter scans show that in this model the diffusion constant does not affect population fitness independently of τ (Fig. 5a and 5b).

The fitness landscapes obtained using the two methods are qualitatively in agreement (Fig. 5a and 5b). This suggests that the AMSPD algorithm can be used to efficiently identify coarse parameter regimes, which can then be further refined by more accurate simulation using the PDA.

4 Conclusion

We have presented an accelerated method for simulating cellular population dynamics. The method generates and employs single representative time series to simulate the

gene expression and reproductive fitness dynamics of all the cells in the population. A constant-number MC method [17,21,22,28,34] is used by the AMPSD algorithm in order to simulate a statistically representative sample of an exponentially growing cell population. This approach allows for accurate simulations with a significant speedup compared to simulations obtained using a previously developed population dynamics algorithm [4]. The accelerated algorithm is a course-grained method designed for scenarios when all the variables of an intracellular biochemical reaction network can be assumed to be at steady-state and cells to divide symmetrically (e.g., [3,8,14,39]). In order to reduce the complexity of the model and simulation times, these assumptions are often invoked when simulating gene expression and cellular dynamics (e.g., [1,3,5,6,18,19,31,37]). Although these assumptions are not always biologically realistic, due to speed of the accelerated method, efficient scans over a large parameter space can be performed in order to identify regions of interest. Once the parameter space region of interest is identified, simulations can then be performed using a more accurate population simulation algorithm. Correspondingly, this method should prove useful for the simulation of gene expression and population models of ever increasing complexity. Furthermore, it is anticipated that the method will apply more generally to other scenarios, for example, to speedup simulations of biochemical reaction networks during periods when the rate parameters are not varying due to noise external to the system [32].

Acknowledgments

This work was supported financially by the National Science and Engineering Research Council of Canada (NSERC).

Author contributions

D.C. conceived and designed the research. D.C. developed the algorithm and performed the simulations. D.C. and M.K. wrote the manuscript. M.K. supervised the study.

References

[1] M. Acar, J. T. Mettetal and A. van Oudenaarden, Stochastic switching as a survival strategy in fluctuating environments, Nat. Genet., 40 (2008), 471-475.

[2] W. Blake, G. Balazsi, M. Kohanski, F. Isaacs, K. Murphy, Y. Kuang, C. Cantor, D. Walt and J. Collins, Phenotypic consequences of promoter-mediated transcriptional noise, Mol. Cell, 24 (2006), 853-865.

[3] B. M. Boman, M. S. Wicha, J. Z. Fields and O. A. Runquist, Symmetric division of cancer stem cells – a key mechanism in tumor growth that should be targeted in future therapeutic approaches, Clinical Pharmacology and Therapeutics, 81 (2007), 893-898.

[4] D. A Charlebois, J. Intosalmi, D. Fraser and M. Kaern, An algorithm for the stochastic simulation of gene expression and heterogeneous population dynamics, Commun. Comput. Phys., 9 (2011), 89-112.
[5] D. A Charlebois, N. Abdennur and M. Kaern, Gene expression noise facilitates adaptation and drug resistance independently of mutation, Phys. Rev. Lett., 107 (2011), doi: 10.1103/PhysRevLett.107.218101.
[6] A. S. Ribeiro, D. A. Charlebois and J. Lyold-Price, *CellLine*, a stochastic cell lineage simulator, Bioinformatics, 23 (2007), 3409-3411.
[7] A. Eldar and M. Elowitz, Functional roles for noise in genetic circuits, Nature, 467 (2010), 167-173.
[8] B. Feierbach and F. Chang, Roles of the fission yeast formin for3p in cell polarity, actin cable formation and symmetric cell division, Curr. Biol., 11 (2001), 1656-1665.
[9] D. Fraser and M. Kaern, A chance at survival: gene expression noise and phenotypic diversification strategies, Molec. Microbiol., 71 (2009), 1333-1340.
[10] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys., 22 (1976), 403-434.
[11] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem., 81 (1977), 2340-2361.
[12] T. Lu, D. Volfson, L. Tsimring and J. Hasty, Cellular growth and division in the Gillespie algorithm, Syst. Biol., 1 (2004), 121-128.
[13] D. Volfson,, J. Marciniak1, W. J. Blake, N. Ostroff1, L. S. Tsimring and J. Hasty, Origins of extrinsic variability in eukaryotic gene expression, Nature, 439 (2006), 861-864.
[14] W. B. Huttner and Y. Kosodo, Symmetric versus asymmetric cell division during neurogenesis in the developing vertebrate central nervous system, Curr. Opin. Cell. Biol., 17 (2005), 648-657.
[15] M. Kaern, T. C. Elston, W. J. Blake and J. J. Collins, Stochasticity in gene expression, Nat. Rev. Genet., 6 (2005), 451-464.
[16] B. B. Kaufmann and A. van Oudenaarden, Stochastic gene expression: from single molecules to the proteome, Curr. Opin. Genet. Dev., 17 (2007), 107-112.
[17] Y. Lin, K. Lee and T. Matsoukas, Solution of the population balance equation using constant-number Monte Carlo, Chem. Eng. Sci., 57 (2002), 2241-2252.
[18] T. Lu, D. Volfson, L. Tsimring and J. Hasty, Cellular growth and division in the Gillespie algorithm, Syst. Biol., 1 (2004), 121-128.
[19] D. Nevozhay, R. M. Adams, E. V. Itallie, M. R. Bennett and G. Balazsi, Mapping the environmental fitness landscape of a synthetic gene circuit, PLoS Comput. Biol., 8 (2012), doi:10.1371/journal.pcbi.1002480.
[20] N. Maheshri and E. K. O'Shea, Living with noisy genes: how cells function reliably with inherent variability in gene expression, Annu. Rev. Biophys. Biomol. Struct., 36 (2007), 413-434.
[21] N. V. Mantzaris, Stochastic and deterministic simulations of heterogeneous cell population dynamics, J. Theor. Biol., 241 (2006), 690-706.
[22] N. V. Mantzaris, From single-cell genetic architecture to cell population dynamics: Quantitatively decomposing the effects of different population heterogeneity sources for a genetic network with positive feedback architecture, Biophys. J., 92 (2007), 4271-4288.
[23] M. D. McKay, R. J. Beckman and W. J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21 (1979), 239-245.

[24] M. D. McKay, Sensitivity and uncertainty analysis using a statistical sample of input values, in: Y. Ronen (Ed.), Uncertainty Analysis, Ch. 4, pp. 145-186, CRC Press, Bcca Raton, Florida, 1988.
[25] R. Murugan, Multiple stochastic point processes in gene expression, J. Stat. Phys., 131 (2008), 153-165.
[26] J. Paulsson, Summing up the noise in gene networks, Nature, 427 (2004), 415-418.
[27] J. M. Raser and E. K. O'Shea, Control of stochasticity in eukaryotic gene expression, Science, 304 (2004), 1811-1814.
[28] D. Ramkrishna, The status of population balances, Rev. Chem. Engng., 3 (1985), 49-95.
[29] M. S. Samoilov, G. Price and A. P. Arkin, From fluctuations to phenotypes: The physiology of noise, Sci. STKE, 366 (2006), re17.
[30] M. Scott, B. Ingalls and M. Kaern, Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks, Chaos, 16 (2006), 026107.
[31] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, PNAS, 105 (2008), 17256-17261.
[32] V. Shahrezaei, J. Ollivier and P. Swain. Colored extrinsic fluctuations and stochastic gene expression, Mol. Syst. Biol., 4 (2008), 196.
[33] A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov and U. Alon, Variability and memory of protein levels in human cells, Nature, 444 (2006), 643-646.
[34] M. Smith and T. Matsoukas, Constant-number Monte Carlo simulation of population balances, Chem. Eng. Sci., 53 (1998), 1777-1786.
[35] J. L. Spudich and D. E. Koshland, Non-genetic individuality: chance in the single cell, Nature, 262 (1976), 467-471.
[36] P. S. Swain, M. B. Elowits and E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression, PNAS, 99 (2002), 12795-12800.
[37] M. Thattai and A. van Oudenaarden, Attenuation of noise in ultrasensitive signaling cascades, Biophys. J., 82 (2002), 2943-2950.
[38] G. Uhlenbeck and L. Ornstein, On the theory of Brownian motion, Phys. Rev., 36 (2008), 823-841.
[39] S. Woolner and N. Papalopulu, Spindle position in symmetric cell divisions during epiboly is controlled by opposing and dynamic apicobasal forces, Dev. Cell, 22 (2009), 775-787.
[40] R. Zadrag-Tecza, M. Kwolek-Mirek, G. Bartosz and T. Bilinski, Cell volume as a factor limiting the replicative lifespan of the yeast Saccharomyces cerevisiae, Biogerontology, 10 (2009), 481-488.
[41] Z. Zhang, W. Qian and J. Zhang, Positive selection for elevated gene expression noise in yeast, Mol. Syst. Biol., (2009), doi:10.1038/msb.2009.58.
[42] D. Zhuravel, D. Fraser, S. St-Pierre, L. Tepliakova, W. Pang, J. Hasty and M. Kaern, Phenotypic impact of regulatory noise in cellular stress-response pathways, Syst. Synth. Biol., 4 (2010), doi:10.1007/s11693-010-9055-2.

Chapter 4

Modeling and Simulating Replicative Aging and Cell Competition

The models and simulations presented in this chapter were performed to exemplify the utility of an object-oriented population simulation framework (a brief overview of the framework is presented in Section 4.1; for a more detailed description including pseudocode see Abdennur [2]). This work incorporates additional biologically relevant details, namely cell aging and competition, not considered elsewhere in this thesis.

4.1 Object-Oriented Framework for Simulating Heterogeneous Cell Populations

The AMSPD and PDA have many advantages including accurately simulating cellular population dynamics at a single-cell resolution. However, "real-time" cell communication cannot efficiently be incorporated into these algorithms. This is because in order

to permit parallelization of the AMSPD and PDA, each cell must be simulated independently and can only communicate with the other cells at specified synchronization barriers (time intervals at which the simulation briefly exits the parallel region). As the time between synchronization barriers is reduced to permit frequent communication events, the speedups gained from parallelization are lost. Likewise, the ability to couple cells to an external environment (e.g., to a growth-limiting resource in media such as glucose) is limited.

Abstracting cells as interacting objects with various associated attributes and behaviours is perhaps the most natural and versatile approach for modeling cellular and population dynamics. An object-oriented framework for simulating individual-based models of heterogeneous cell populations was presented in Abdennur [2].

The framework, inspired by Gillespie's reaction channel concept [5, 6], makes use of objects called "simulation channels" to handle the scheduling and firing of state-updating events on individual cells. This is analogous to Gillespie's algorithm, where the propensities of different reaction channels are used to determine when the next reaction will occur (scheduling) and how the numbers of molecules are changed when it occurs (firing). These state-updating events represent the outcome of arbitrary cell intrinsic processes, such as gene expression, that in some way, deterministically or stochastically, alter the continuous or discrete attributes of cells.

More generally, state-updating events that involve multiple cells are controlled by another set of objects, which are conceptually identical to single-cell simulation channels, but that act on global variables that impact the entire population or a smaller subset of cells. To increase simulation efficiency, dependency graphs are used to determine which simulation channels require rescheduling when another simulation channel modifies cell or global attributes. This was introduced by Gibson *et al.* [4] to improve the performance of the Gillespie algorithm, but can readily be generalized to

simulation channels.

Since each cell is simulated individually, complete cell lineages can be tracked until the size of the population becomes prohibitively large. To simulate population dynamics beyond this point, the framework allows a simulation to begin with as few as one cell and proceed until a pre-specified maximum population size is reached. Once the population size limit is reached, the CNMC method (Section 1.3.2) is used to keep a fixed sample population size with the appropriate composition.

In this framework, the dynamics of single cells and their interaction with the environment is facilitated by making each cell a unique "simulation engine" responsible for scheduling and firing its own simulation channels. Another engine controls the simulation channels that affect global attributes. These together constitute another entity called the "world". For example, a fluctuating environment can be simulated by defining a set of global environmental parameters and simulation channels that change these parameters. Channels associated with the world can fire channels associated with one or more cells and vice versa.

Cell-cell communication is possible when simulation channels are fired in a first-come, first-served basis by a global scheduler. This approach is referred to as the "synchronous method", since updates can be performed on some, or all, of the cells in the population every time a simulation channel is fired. Simulations can also be performed in this framework using an "asynchronous method" (i.e., the method implemented in the AMSPD and PDA where cells are simulated independently in between the synchronization barriers), which is ideal and parallelizable for non-interacting cells. The asynchronous method was used to simulate a model of replicative aging in yeast (Section 4.2) and the synchronous method is used to model resource competition between cell populations (Section 4.3).

4.2 Volume and Age-Dependent Growth in Yeast

4.2.1 Background

In many metazoan and yeast cells, there exists a limit on the number of cell divisions that a cell can achieve in its lifetime, known as the replicative lifespan of the cell. Cultured human cells have a limited capacity to divide (e.g., human fibroblasts), called the Hayflick limit [7,14], and therefore the population eventually goes extinct. These cells are non-immortal due to the accumulation of DNA damage resulting from the progressive shortening at each cell division of the telomere tips at the ends of chromosomes [14]. In contrast, there are also cell populations that can exist indefinitely (e.g., germline cells, stem cells, and cancer cells). Budding yeast cell populations are immortal as the older cells eventually stop replicating while the younger cells continue to replicate.

Unique to eukaryotes, the aptly named budding yeast cells divide asymmetrically by budding. The formation of the bud begins in S-phase and pinches off at the end of the cell cycle. A recent study suggested that the replicative lifespan of the cell is limited by the progressive increase in cell volume resulting from asymmetric division [19]. The authors of this study used three wild-type laboratory strains together with mutant versions of these strains with various defects in antioxidant enzymes. Each pair of wild-type and mutant strains demonstrated a cessation of budding after reaching the same final volume, but had different replicative lifespans. Based on these observations and results from a previous study that investigated the relationship between replicative age and cell volume, the authors proposed that a genetically determined critical cell volume may be the main factor limiting the ability of mother yeast cells to continue to divide. The authors suggested that progressive growth of cell volume with each cycle is an unavoidable consequence of the mechanism of reproduction,

because the mother cell experiences incremental growth during each G1 phase, before a new bud has formed. In the mutant cells, oxidative stress leading to DNA damage was presumed to cause delays at cell cycle checkpoints without hindering volume growth, so that mutants attain the limiting cell volume after a smaller number of cell divisions. A more recent study supports these findings [17]. The authors found that cells that underwent fewer replications accrued more volume per generation and vice-versa, reaching senescence at similar volumes.

4.2.2 Model

We implement a simple phenomenological model of this system in order to investigate single-cell and population-level dynamics when age-dependent effects are considered.

Cell volume growth was assumed to be linear at a given age (generation) and was modeled using the following equation

$$V(t) = V(t-1) + dv, \tag{4.1}$$

where dv was obtained (for fixed time step dt = 1) using

$$dv = \frac{kg}{1 + (g/k)^2} dt, \tag{4.2}$$

where k is the maximum growth rate and g is the generation (number of divisions since birth). The growth rate is a non-monotonic function of age, as it is faster when the cell is young, and slower when the cell is old. Here, cells divide asymmetrically, producing a larger mother and a smaller daughter (g = 0). At each division, the daughter cell receives a fixed fraction (β) of the total predivision mother cell volume, while the mother receives the remaining fraction (1-β). Cell division was set to occur every time a pre-specified (strain specific) volume increment ΔV was reached. When a critical volume is achieved after cell division (V_{crit}), the cell was flagged as senescent and no longer allowed to divide.

4.2.3 Results and Discussion

Figure 4.1a shows the growth pattern of a single-mother cell undergoing successive budding and division until reaching senescence after its cell volume crosses V_{crit} after division. The periodic volume decreases are a result of the budding off of daughter cells which are also successively larger after each division. Figure 4.1b presents a plot of the cell volume as a function of generation for a wild-type strain (BY4741) and an antioxidant defective mutant strain (Δprx). Different values of k and ΔV were selected to reproduce the experimental results of Zadrag-Tecza *et al.* [19] (Appendix - Fig. 4). Despite both strains having a similar initial volume at age $g = 0$, the mutant strain accumulates volume at a faster rate and displays a shorter replicative lifespan. The simulation data fit well to a linear trend as did the experimental data.

An important result is obtained if we compare the steady state volume distributions obtained from simulations of a set of cell chains (Fig. 4.1c) with a statistical ensemble of cells using the CNMC method (Fig. 4.1d). Namely, the second peak of cells at large volume in Fig. 4.1c is an artifact because senescent cells are never outcompeted by younger actively replicating cells. This highlights the importance of the CNMC method for simulating population dynamics.

It was shown recently by Levy *et al.* [90] that replicative age in budding yeast cells correlates with a stress resistance state. Specifically, older cells were found to be more likely to resist heat shock than younger cells due to the accumulation of a stress resistance protein. This phenomenon should be considered in future simulations of budding yeast populations under stress.

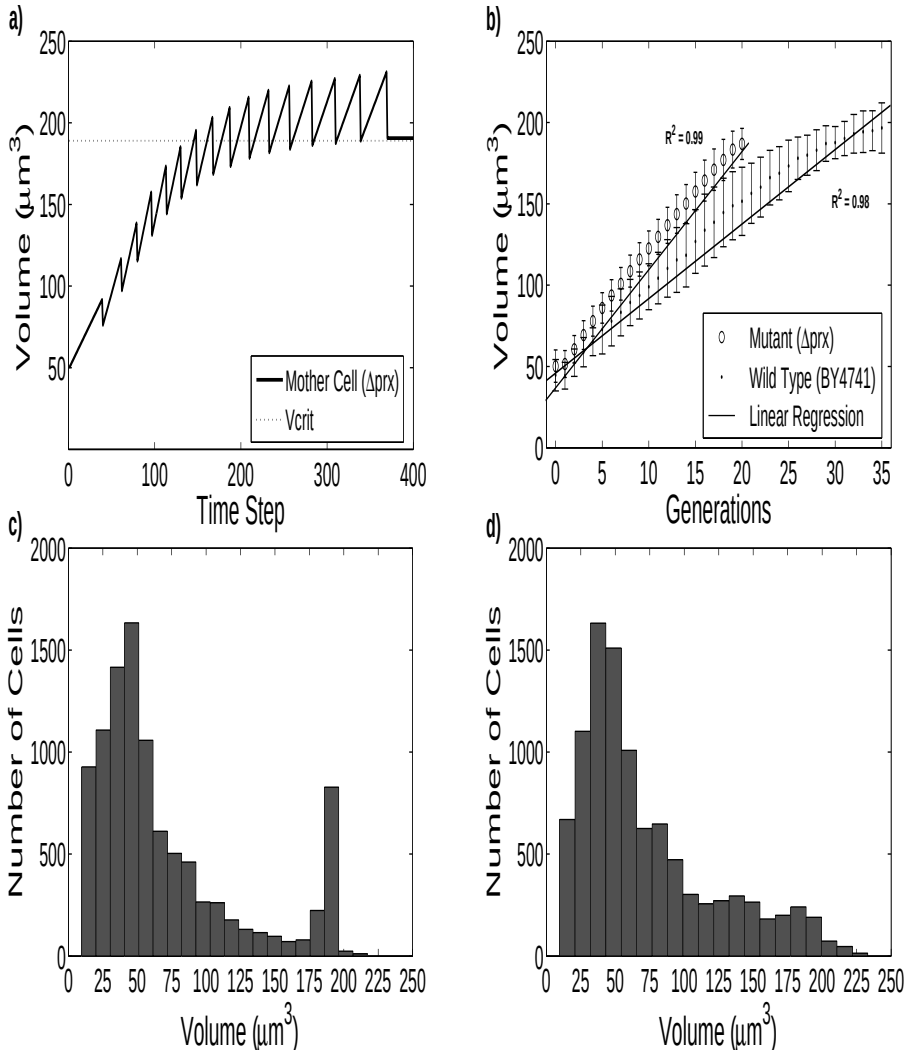


Figure 4.1 Age-dependent cell volume dynamics. a) History of a single mother cells volume over time. (b) Age distribution of strains of yeast with different k and ΔV. For "wild-type" BY4741, $k = 4.6$ and $\Delta V = 12\mu m^3$. For Δprx, $k = 1.15$ and $\Delta V = 42.3\mu m^3$. (c) Volume distribution obtained from of a collection of 10,000 Δprx cell chains. (d) Volume distribution obtained from a statistical ensemble of 10,000 Δprx cells. Figure used with permission from Abdennur [2].

4.3 Resource Competition

4.3.1 Background

Cell competition was discovered almost 40 years ago in the fruit fly *Drosophila melanogaster* [8]. This phenomenon was described as a situation in which slowly dividing, but otherwise viable, cells were eliminated from a population of more rapidly dividing cells [8,15,16]. In these studies, competition was between wild-type and mutant fruit fly cells, where the mutants were heterozygous for a deletion of a *Minute* gene. Heterozygous *Minute* mutations result in a smaller, less fertile fly with a longer cell division cycle (flies homozygous for a *Minute* mutation cannot be used in competition studies as they do not form viable populations). *Minute*-induced competition also appears to be conserved in mouse tissues [12]. Interestingly, especially in light of the CNMC method used in this thesis, it has been observed in fruit flies that competitor cells can proliferate by killing wild-type cells by inducing apoptosis, such that the total number of cells in the population does not change [3, 9, 10]. Consequently, clonal expansion of this form may not result in any morphological aberrations and it has been suggested that it may pose a challenge for the early detection of cancer [11].

4.3.2 Model

In this section a toy model of competition, for a shared resource between two cell populations is presented. Inspired by the *Minute* gene studies [8,15,16], the mutation negatively influences cellular growth and division. Each cell can uptake or excrete the resource across its cell membrane by passive diffusion. The individual cells can also consume the resource for use in their cellular processes. The two populations are identical except for the membrane permeability of the constitutive members. In this model, the mutation changes the permeability of the cell membrane. The wild-type

(wt) cells have an equal or higher membrane permeability than the mutant-type (mt) cells for an extracellular resource ($R(t)$) critical to growth. Resource diffusion and consumption is modeled in each cell as follows

$$r_i(t) = r_i(t - dt) + (V_i^{-1} k_{diff,j}(R - r_i) - k_{deg} r_i)dt, \tag{4.3}$$

where r_i is the intracellular concentration of the resource and V_i the cell volume of a given cell i, $k_{diff,j}$ is the membrane permeability (units of LT^{-1}) multiplied by the surface area of the cell in a given population j (wt or mt), and k_{deg} is the cellular consumption rate of the nutrient.

Each time the cell updates its resource level, the concentration of the global resource is updated by

$$R(t) = R(t - dt) + (-V^{-1} k_{diff}(R - r_i)(N_{total}/N_{CNMC}) + k_{dil}(R_{res} - R))dt, \tag{4.4}$$

where V is the volume of the environment (e.g., a test tube with fixed volume), k_{dil} is the dilution rate or steady flow rate of the media in and out of the environment, and R_{res} the constant nutrient concentration of a large resource reservoir. N_{CNMC} is the number of cells in the constant-number sample population while N_{total} is the estimated number of cells in the 'true' population that the sample represents. According to one interpretation of the CNMC algorithm [93, 100], each cell of the sample population represents N_{total}/N_{CNMC} cells of the true population. The N_{total} global attribute is incremented or decremented by the appropriate amount every time a birth or death event occurs in the sample population, respectively.

Cell division is modeled as a reaction whose propensity depends on the resource concentration inside the cell over time ($r_i(t)\lambda$, where division rate λ is a constant). Since the kinetic rate depends on time, the object-oriented framework uses an algorithm developed by Shahrezaei *et al.* [13] that extends the standard SSA to include extrinsic fluctuations to include rate constants that change continuously with time [2].

Population dynamics were accounted for using the synchronous method which enables real-time coupling of cell populations. Note that in this model there is no apoptosis. Cell competition is captured naturally by CNMC replacement (see Section 1.3.2), as the probability that a given mother cell will be replaced in the next generation by a daughter cell from the more fit population is higher than that of the mother cell being replaced by a daughter cell from the less fit population.

4.3.3 Results and Discussion

As expected, when the membrane permeability of the *wt* and *mt* populations are the same, we see an equal number of cells from each population in the total composite population (Fig. 4.2). However, when the rate of diffusion for the *wt* population is higher than the *mt* population, the *wt* population can acquire resources more rapidly and outcompetes the other population. For example, when the membrane permeability of the cell in the *wt* population is 1.2 times greater than the membrane permeability of the cells in the *mt* population, the total population is composed of roughly 80% of cells from the *wt* population after 10 generations (Fig. 4.2). The results of this model suggest that a relatively small phenotypic change in one of the subpopulations can have a significant effect on the composition of the population. The high degree of variability in the results stems from the small size of the CNMC sample compared to the true population size (Fig. 4.2). A small sample size was chosen to reduce simulation times because the simulation could not be parallelized. Another source of error is introduced because the dynamics of the cells in the sample depends on the number of cells in the true population, which itself is estimated from birth and death events in the sample.

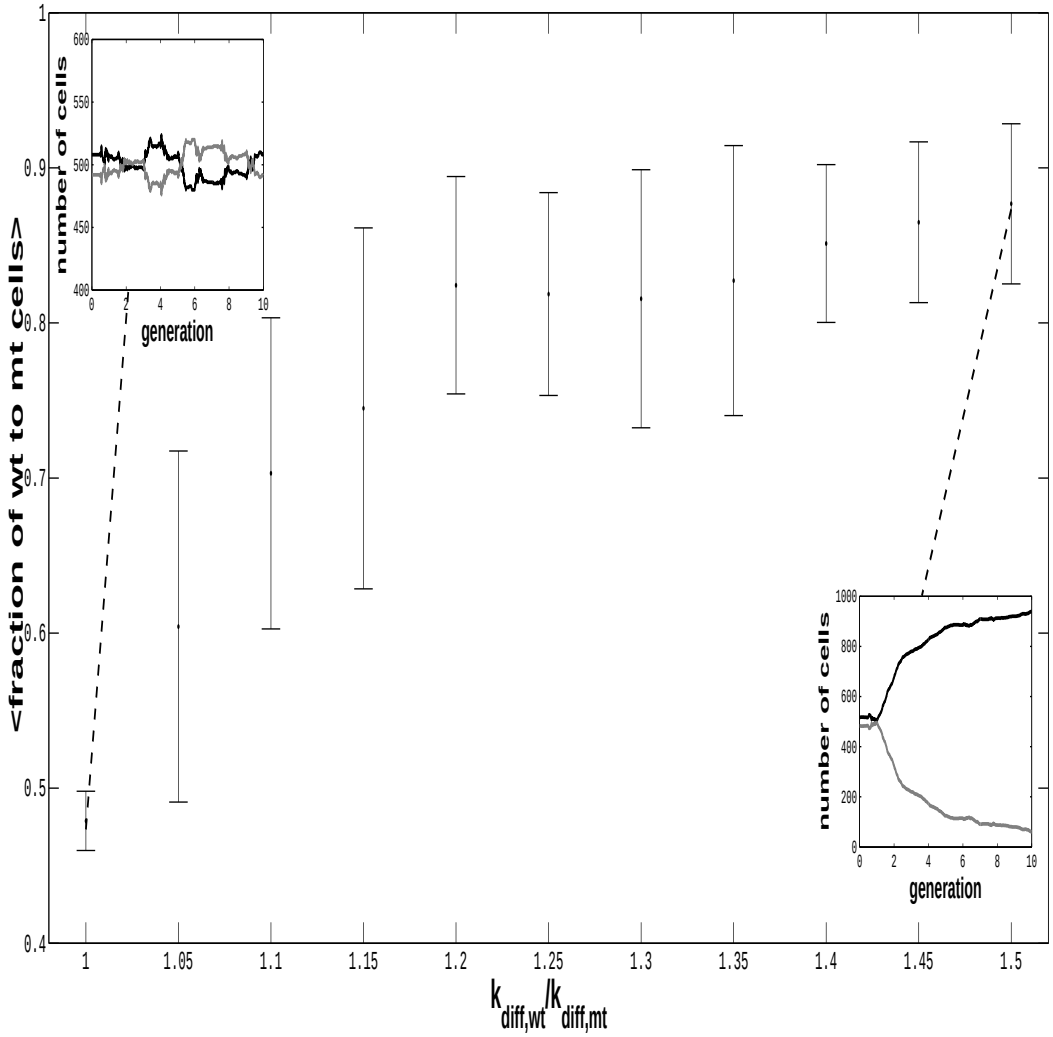


Figure 4.2 Resource competition between two cell populations. The fraction of cells from the wild-type *wt* population in the total composite population (*wt* and mutant-type *mt* cells) is plotted as a function of the ratio of the diffusion constants for each of the populations. The insets a single stochastic realization of the number of cells in each population for two different ratios of the diffusion constants (black line - *wt* population, gray line - *mt* population) over 10 generations. The following parameters were used: $N_{CNMC} = 1000$, $V = 1$, $V_i(t = 0) = 0.0001$, $R_{res} = 1$, $k_{diff,wt} = 0.01 - 0.015$, $k_{diff,mt} = 0.01$, $k_{deg} = 0.05$, and $k_{dil} = 0.01$. The cell division and death rates were set to 1 and 0.02, respectively.

4.4 References

[1] N. Abdennur, D. Charlebois, and M. Kærn. A framework for ensemble simulations of cell population heterogeneity. *RECOMB 2010: 3rd Annual Joint Conference on Systems Biology, Regulatory Genomics and Reverse Engineering Challenges*, 2010.

[2] N. Abdennur. A Framework for Individual-based Simulation of Heterogeneous Cell Populations. Master's thesis, University of Ottawa, 2012.

[3] C. de la Cova, M. Abril, P. Bellosta, P. Gallart, *et al.*. Drosophila myc regulates organ size by inducing cell competition. *Cell*, 117:107-116, 2004.

[4] M.A. Gibson and J. Bruck. Exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, 105:1876-1889, 2000.

[5] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403-434, 1976.

[6] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340-2361, 1977.

[7] L. Hayflick and P.S. Moorhead. The serial cultivation of human diploid cell strains. *Exp. Cell Res.*, 25:585-621, 1961.

[8] G. Morata and P. Ripoll. Minutes: mutants of *Drosophila* autonomously affecting cell division rate. *Dev. Biol.*, 42:211-221, 1975.

[9] E. Moreno, K. Basler, and G. Morata. Cells compete for decapentaplegic survival factor to prevent apoptosis in *Drosophila* wing development. *Nature*, 416:755-759, 2002.

[10] E. Moreno and K. Basler. dMyc transforms cells into super-competitors. *Cell*, 117:117129, 2004.

[11] E. Moreno. Is cell competition relevant to cancer?. *Nat. Rev. Cancer*, 8:141147, 2008.

[12] E.R. Oliver, T.L. Saunders, S.A. Tarle, and Glaser, T. Ribosomal protein L24 defect in belly spot and tail (Bst), a mouse Minute. *Development*, 131:39073920, 2004.

[13] V. Shahrezaei, J.F. Ollivier, and P. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, doi:10.1038/msb.2008.31, 2008.

[14] J.W. Shay and W.E. Wright. Hayflick, his limit, and cellular ageing. *Nat. Rev. Mol. Cell Biol.*, 1:72-76, 2000.

[15] P. Simpson. Parameters of cell competition in the compartments of the wing disc of *Drosophila*. *Dev. Biol.*, 69:182-193, 1979.

[16] P. Simpson and G. Morata. Differential mitotic rates and patterns of growth in compartments in the *Drosophila* wing. *Dev. Biol.*, 85:299308, 1981.

[17] J. Yang, H. Dungrawala, H. Hua, A. Manukyan, and *et al.*. Cell size and growth rate are major determinants of replicative lifespan. *Cell Cycle*, 10:144-155, 2011.

[18] R. Zadrag, M. Kwolek-Mirek, G. Bartosz, and T. Bilinski. Relationship between the replicative age and cell volume in *Saccharomyces cerevisiae*. *Acta Biochim. Pol.*, 53:747-751, 2006.

[19] R. Zadrag-Tecza, M. Kwolek-Mirek, G. Bartosz, and T. Bilinski. Cell volume as a factor limiting the replicative lifespan of the yeast *Saccharomyces cerevisiae*. *Biogerontology*, 10:481-488, 2009.

Chapter 5

Gene Expression Noise Facilitates Adaptation and Drug Resistance

Daniel A. Charlebois, Nezar Abdennur, Mads Kærn. *Phys. Rev. Lett.* (2011).

PRL 107, 218101 (2011) PHYSICAL REVIEW LETTERS week ending 18 NOVEMBER 2011

Gene Expression Noise Facilitates Adaptation and Drug Resistance Independently of Mutation

Daniel A. Charlebois,[1,2,*] Nezar Abdennur,[2,3] and Mads Kaern[1,2,3,†]

[1]*Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, Ontario, K1N 6N5, Canada*
[2]*Ottawa Institute of Systems Biology, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada*
[3]*Department of Cellular and Molecular Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada*

(Received 4 July 2011; published 14 November 2011)

We show that the effect of stress on the reproductive fitness of noisy cell populations can be modeled as a first-passage time problem, and demonstrate that even relatively short-lived fluctuations in gene expression can ensure the long-term survival of a drug-resistant population. We examine how this effect contributes to the development of drug-resistant cancer cells, and demonstrate that permanent immunity can arise independently of mutations.

DOI: 10.1103/PhysRevLett.107.218101 PACS numbers: 87.10.Mn, 87.10.Rt, 87.16.Yc, 87.23.Cc

Gene expression is a stochastic process that enables genetically identical cells in the same environment to exhibit phenotypic variation [1–3]. This noise-induced nongenetic (epigenetic) variability can be beneficial to cell populations experiencing acute stress by providing a temporary basis for natural selection [4–7].

Experimental observations suggest that gene expression is inherently associated with ''epigenetic memory,'' defined by the fluctuation relaxation time of a gene product within a cell lineage. In human lung cancer cells, this relaxation time can be as long as four generations [8].

Brock *et al.* [9] recently argued that epigenetic memory might accelerate tumor progression by contributing to the development of drug-resistant cancer cells. In this hypothesis, phenotypic variability from the noisy expression of gene X that confers resistance renders some cells (and their offspring) temporarily insensitive to the drug, thereby increasing the probability of acquiring a mutation conferring permanent immunity. In the present work, we develop a minimal model to study this phenomenon quantitatively.

To study how gene expression noise impacts the dynamics of isogenic cell populations under stress, we define the reproductive fitness (W) as the number of offspring produced in the presence of the stressor (i.e., a drug) relative to that produced in its absence. For simplicity, we assume that all cells produce offspring at the same rate in the absence of the drug, and define the generation time (t_D) as the time it takes for each cell to reproduce once. We set the generation time as unit time and report all time scales relative to t_D. We also assume that cells carry the gene X conferring drug resistance when its expression level x is sufficiently high, and that this gene is expressed stochastically in individual cells.

The effects of gene expression noise on populations under stress have previously been analyzed to explain why certain genes have high expression noise [5–7]. In these analyses, the dependency between gene expression and reproductive fitness was defined by the integral

$$W(t) = \int w(x) p_x(x, t) dx, \tag{1}$$

where $p_x(x, t)$ is the probability distribution function (PDF) describing the concentration (x) of the gene product across the population, and $w(x)$ is the microscopic fitness function describing the effect of the drug on the fitness of cells with a given expression level. The basic concept is illustrated in Fig. 1(a) using a model where $w(x)$ is described by the Heaviside step function, such that cells are unable to reproduce if their expression level is below a critical value, $w(x < x_c) = 0$, and unaffected by the drug otherwise, $w(x \geq x_c) = 1$. In this case, previous theoretical work [5–7] concluded that high gene expression noise is beneficial at high drug doses, since the fraction of cells expressing above a reproductive threshold x_c increases with the width of the initial expression distribution [Fig. 1(a)]. However, because $p_x(x, t)$ is assumed fixed at the time of drug treatment, this conclusion is valid only for instantaneous selection effects. The analysis of prolonged stress exposure necessitates an approach where selection, inheritance, and gene expression dynamics all contribute to the evolution of the population.

Population survival during prolonged drug exposure is a first-passage time problem. In the absence of mutations conferring permanent immunity, cells that survive the initial selection will eventually succumb to the drug since they cannot maintain high expression indefinitely. Consider a subpopulation of cells with the same level of x above x_c [Fig. 1(b)]. The time interval in which a given cell can reproduce is the first-passage (or sojourn) time $t_S(x)$, where the threshold x_c represents an absorbing barrier. Although cells are initially identical, the expression of the drug-resistance gene evolves differently in different cells, and the time to reach the reproductive threshold is a random variable described by the first-passage time distribution $p_S(x, t_S)$ [Fig. 1(b), Inset]. Since only cells with $t_S(x) > t_D$ reproduce, $w(x)$ in Eq. (1) is given by

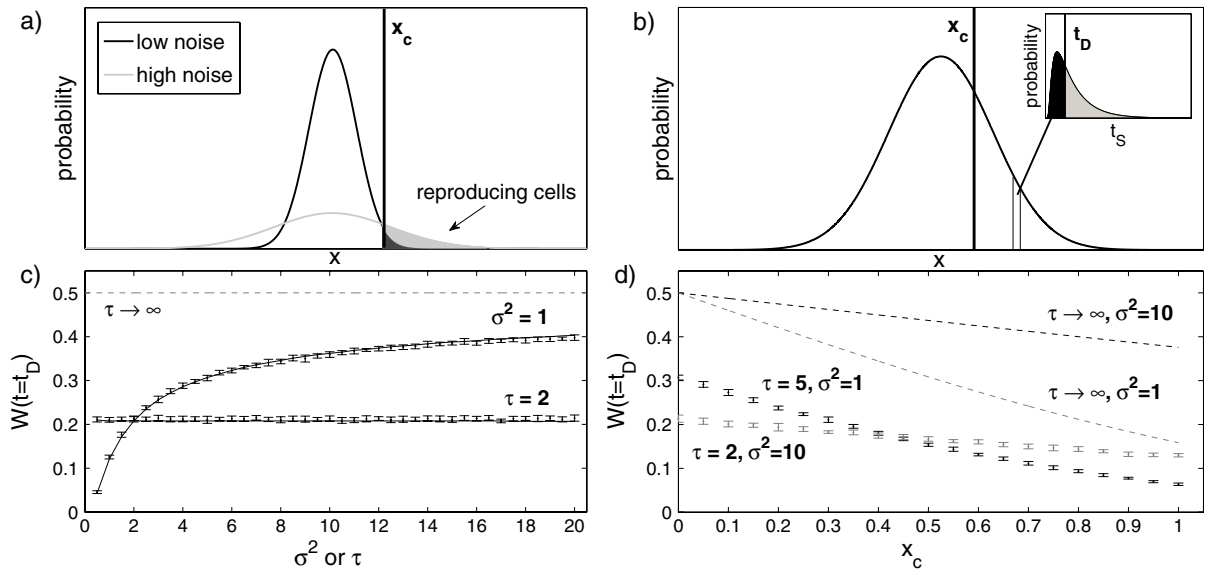


FIG. 1 (color online). Epigenetic effects on a cell population exposed to stress. (a) Schematic of instantaneous selection effects. (b) Schematic of generalized model. (c) Reproductive fitness at the time of first division $W(t = t_D)$ after the application of a stress (at $x_c = 0$) as a function of τ or σ^2 for fixed σ^2 or τ, respectively. Analytical curves (solid lines) were obtained via numerical solution of Eq. (3). (d) $W(t = t_D)$ as a function of x_c for high and low σ^2. τ and σ^2 are scaled by t_D. Dashed lines represent results obtained from Eq. (1), or equivalently Eq. (3) in the limit $\tau \to \infty$.

$$w(x) = \int_{t_D}^{\infty} p_S(x, t_S) dt'_S, \tag{2}$$

and the overall fitness of the population at time t can be written as

$$W(t) = \int_{x_c}^{\infty} \left(\int_{t_D}^{\infty} p_S(x, t_S) dt'_S \right) p_x(x, t) dx. \tag{3}$$

The population fitness in Eq. (3) has an explicit solution only in special cases. Previous analyses [5–7] circumvented this problem, in part, by focusing on initial selection effects ($t \to 0$). However, even in this limit, it is also necessary to assume that all cells above the threshold contribute to fitness [i.e., $w(x) = 1$ for $x > x_c$].

To investigate more general cases, we used the Ornstein-Uhlenbeck (OU) process to model the level of gene expression in individual cells [10]. This process can be described by the Langevin equation

$$\frac{dx(t)}{dt} = \frac{1}{\tau}[\mu - x(t)] + c^{1/2}\xi_t, \tag{4}$$

where c and τ are the diffusion constant and the relaxation time, respectively, and ξ_t is Gaussian white noise [$\langle \xi_t \rangle = 0$, $\langle \xi_t \xi_{t'} \rangle = \delta(t - t')$] [11]. The steady-state PDF of the OU process is a Gaussian distribution with mean μ and variance $\sigma^2 = c\tau/2$. Without loss of generality, we set $\mu = 0$ and use the fluctuation time scale τ to model the time scale of epigenetic memory.

The fluctuation time scale of gene expression has been determined experimentally in human lung cancer cells in terms of the "mixing time" τ_m, defined as the lag where the autocorrelation function has decreased by 50% [8]. The mixing time for the stationary OU process is $\tau_m = \tau \ln(2)$. The measured values of τ_m varied between 0.5 to 3.0 generations for different genes, corresponding to values of τ between 0.7 to 4.0 generations for the OU process.

First, we examined the effect of drug treatment on reproductive fitness after one generation time when the absorbing barrier is located at $x_c = 0$. In this case, the first-passage time PDF for $x > x_c$ is given by [12]

$$p_S(x, t_S) = \frac{x}{\sqrt{2\pi c}} \exp\left(\frac{-x^2 \exp(-t_S/\tau)}{2c\tau \sinh(t_S/\tau)} + \frac{t_S}{2\tau} \right) \times \left(\frac{1}{\tau \sinh(t_S/\tau)} \right)^{3/2}. \tag{5}$$

We evaluated the effects of varying the time scale of epigenetic memory and the noise amplitude by numerical integration of Eq. (3), using the steady-state OU distribution to describe the initial gene expression distribution. Figure 1(c) shows the results for fixed noise ($\sigma^2 = 1$) and variable τ, and fixed time scale ($\tau = 2$) and variable σ^2.

The time scale of epigenetic memory significantly affects "acute" reproductive fitness, even for very long fluctuation relaxation times. For example, when $\tau = 20$, W is reduced to 0.4, compared with the value of 0.5

obtained (irrespectively of the noise amplitude) in the permanent epigenetic memory limit $\tau \to \infty$ [Fig. 1(c)]. For $\tau = 2$, the reproductive fitness is approximately 0.2, and the majority of cells starting with $x > x_c$ are unable to maintain above-threshold gene expression long enough to reproduce. In this case, the acute reproductive fitness remains constant, presumably because changing the noise amplitude for $x_c = 0$ does not change the fraction of cells with $x > x_c$.

To examine cases where $x_c > 0$, it is necessary to use numerical simulations since a general closed-form solution of the first-passage time PDF is not available. For this purpose, we employed a population simulation algorithm [13] in which gene expression in each of N individual cells, $x_i(t)$ for $i = 1, \ldots, N$, is obtained by solving Eq. (4) numerically [14]. In these simulations (20 realizations of 10^4 cells unless indicated otherwise), cell division occurs when a deterministic cell cycle "clock," which is reset at each division, reaches t_D. Each cell keeps track of the time since its birth and can only advance its clock if they maintain gene expression above the threshold. Moreover, cells where $x_i(t) \leq x_c$ are assumed to be fixed and unable to change their expression level (i.e., $\tau = \infty$). Simulations were initiated by assigning, to each cell, random initial values of gene expression and the cell cycle clock from the steady-state distribution of the OU process and a uniform distribution $[0{:}t_D]$, respectively.

Numerical calculations of fitness for $x_c > 0$ identified τ as a critical determinant of population survival. Specifically, the fitness of a population with low gene expression noise can be greater than that of a population with high noise if the fluctuation relaxation time is sufficiently long. We observed this in simulations, shown in Fig. 1(d), with an increased threshold x_c for fixed time scales ($\tau = 2$ or $\tau = 5$) and two different fluctuation amplitudes ($\sigma^2 = 1$ or $\sigma^2 = 10$). When the two populations had the same finite value of τ, we observed that increased gene expression noise always provides a fitness benefit (data not shown). However, as expected from Fig. 1(c), incorporating stochastic gene expression dynamics (i.e., finite values of τ) generally yields a significant reduction in fitness compared to the asymptotic permanent memory limit. The magnitude of this reduction is sensitive to both the value of the threshold x_c and the value of τ. This is illustrated in Fig. 1(d) where the fitness of the high noise population is greater than the low noise population only when the value of x_c is sufficiently high.

In our second case, we analyzed the long-term effects of varying the time scale of epigenetic memory on population dynamics and reproductive fitness. For simplicity, we focus on the case where $x_c = 0$ and noise is fixed ($\sigma^2 = 1$). Figure 2(a) shows representative gene expression distributions obtained after 10 generation times for short- and long-term epigenetic memory. When the fluctuation time scale is short ($\tau = 0.5$, top panel), the number of cells that may reproduce (i.e., cells with $x_i(t) > x_c$) is reduced over time since, on average, cells reach the absorbing barrier faster than they reproduce. Correspondingly, given enough time, the population will go extinct. This is not the case when memory is long ($\tau = 10$, bottom panel) and the birth rate exceeds the rate of loss at the absorbing barrier. In addition, the mode of gene expression distribution shifts

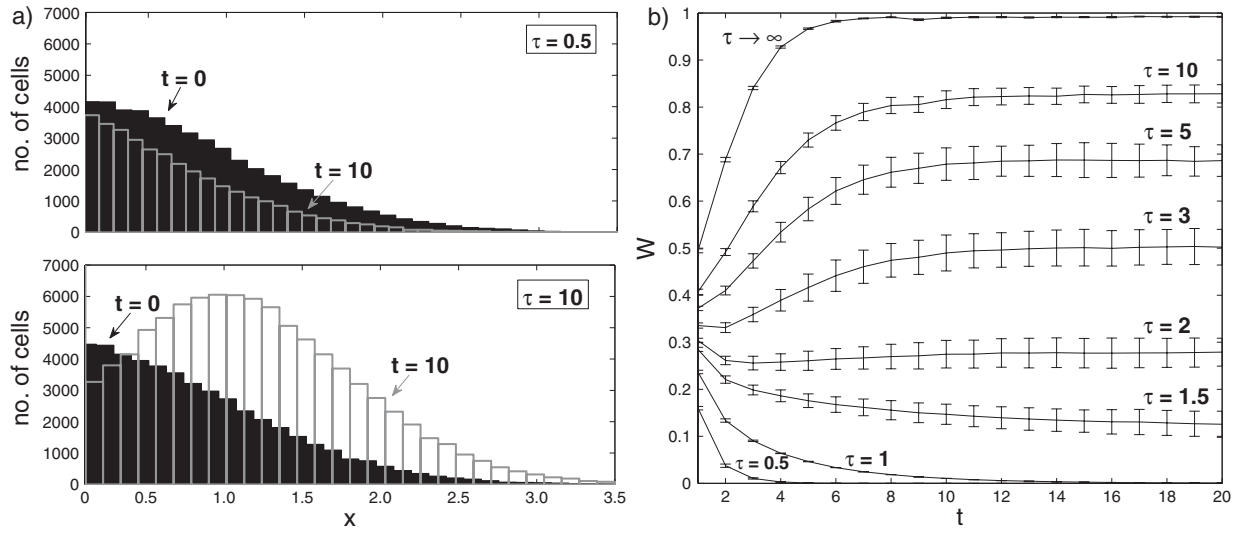


FIG. 2. Effect of epigenetic memory τ on drug resistance at various time scales. (a) Top and bottom plots show population distributions corresponding, respectively, to short ($\tau = 0.5$) and long ($\tau = 10$) epigenetic memory and show the fraction of drug-resistant cells (i.e., cells with $x > x_c$) after acute ($t = 0$) and prolonged ($t = 10$) drug exposures (single realization of 10^5 cells). (b) W as a function of t for various values of τ. t, τ, and σ^2 are scaled by t_D.

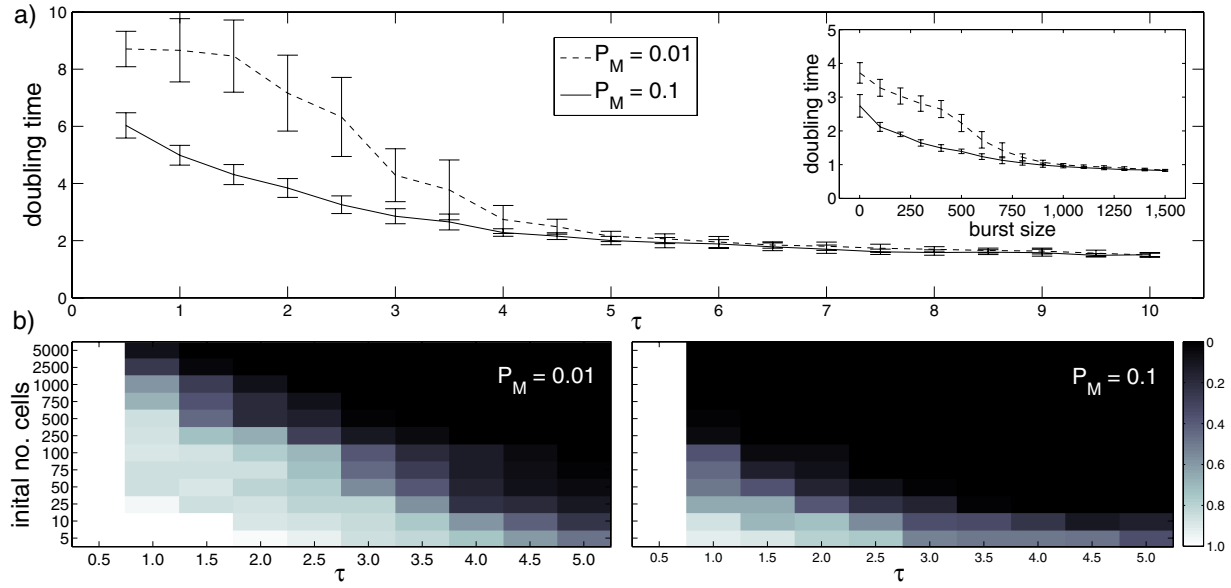


FIG. 3 (color online). Effect of τ and P_M on cancer cell populations undergoing prolonged drug treatment. (a) Effect of P_M on doubling time as a function of τ for an initial population of 1000 nonmutated cells with gene expression levels above a 95% drug threshold. Simulation results using a burst model of gene expression [15] shown in the inset. (b) Heat maps corresponding to (a) show probability of remission after 10 generations (100 realizations). P_M, τ, and σ^2 are scaled by t_D.

to higher values, in resemblance of experimental observations [7].

Relatively short-term epigenetic memory can result in permanent drug resistance even in the absence of mutations. This is illustrated in Fig. 2(b), which shows how the reproductive fitness of populations with different memory time scales evolves over time. In populations with long-term memory (e.g., $\tau = 5$, 10, or ∞), the number of cells that may reproduce increases steadily over time and settles in a steady state where more than half of them reproduce every generation time (i.e., $W(t) > 0.5$). Importantly, populations with memory at intermediate time scales (e.g., $\tau = 1.5$, 2, or 3) may retain long-term viability and finite rates of reproductive fitness. Because the simulations involve finite populations, the outcome of a given realization cannot always be predicted. For example, when $\tau = 1.5$, a viable population was observed to develop in 29% of the simulations while the population went extinct in the remaining 71% of simulations. While populations with short memory (e.g., $\tau = 0.5$ or 1) eventually go extinct, several cell cycles were needed for the drug to fully affect all cells.

In the third and final case, we investigated the added effect of genetic mutations on the development of drug-resistance. A central element of the Brock *et al.* hypothesis is that temporary drug resistance due to slow fluctuations in gene expression may contribute to tumor development by increasing the overall probability that some cells acquire a mutation conferring permanent immunity. To model this scenario, we allowed each cell with an expression level above x_c the chance to mutate once per generation time. We denote this probability P_M. If a cell acquired the mutation, it and its offspring were permanently resistant to the drug, and the survival of a continuously growing population inevitable.

We first investigated the added effect of mutations on the reemergence of a cancerous tumor under constant drug treatment. In these simulations, we chose x_c such that the drug instantaneously removed 95% of the population, and measured the time it took for the remaining cells to double in number. Figure 3(a) shows the dependency of this doubling time on τ when P_M is equal to 0.01 and 0.1. These mutation rates are unrealistically high biologically and were chosen to illustrate the effect of epigenetic memory in an extreme limit.

As expected, increasing the mutation probability significantly reduces the doubling time when the gene expression fluctuations are short-lived. Unexpected, however, the value of τ beyond which mutations do not have an additional effect is remarkably short despite the unrealistically high mutation rates. Specifically, the doubling time is more or less unaffected by P_M when τ is roughly above 4 generations, corresponding to the upper range of mixing times observed experimentally [8].

We confirmed our results using a semirealistic model of gene expression noise [15] where proteins are synthesized in irregular bursts at irregular intervals [Fig. 3(a), Inset]. We also tested the effect of replacing the fitness threshold with a more realistic sigmoidal fitness function and found

no qualitative difference (data not shown). In reality, gene expression dynamics may follow more complex kinetics than that of a simple mean-reverting process due, for example, to multistability and noise-driven switching [16,17]. Our simulation results demonstrate that such complexity is not required for gene expression noise to have a significant impact on population dynamics under prolonged stress.

We also determined how the probability of remission depends on the mutation rate, the initial number of cancer cells with above-threshold expression, and the time scale of gene expression noise. In these simulations, the cancer is in remission if no cells have above-threshold gene expression and have not acquired a mutation conferring permanent immunity within 10 generation times. As expected [Fig. 3(b)], the probability of remission is greatly decreased when the number of initial surviving cancer cells or the mutation rate is increased. Also, when τ is very short, remission is virtually guaranteed. However, the probability that a drug-resistant cell population will emerge can be quite substantial within the experimentally observed range of τ. Even with a relative low mutation rate ($P_M = 0.01$) and 10 surviving cells, the probability of remission is only 42% when $\tau = 4.0$.

In summary, we have analyzed the effect of gene expression noise on the reproductive fitness of isogenic cell populations under stress as a first-passage time problem. By explicitly incorporating the ''epigenetic memory'' of this noise (i.e., the fluctuation relaxation time), we have generalized previous theoretical work that explained the acute effects of noise amplitude but did not incorporate gene expression dynamics [5–7]. This generalization is important for two reasons. First, it has allowed us to demonstrate using a minimal model that gene expression noise with biologically realistic time scales has a significant effect on reproductive fitness under stress and is a critical determinant of population survival. Second, it enables theoretical and computational investigations of experimentally observed phenomena associated with prolonged stress exposure, including reversible shifts in gene expression distributions [7], and drug resistance. In this context, we have demonstrated that the time scale of epigenetic memory required to develop a drug-resistant cell population independently of mutations is comparable to that measured for certain genes in human cancer cells [8]. Correspondingly, long-term population survival may not require specialized memory-conferring mechanisms. It might, for example, be achieved without a significant fitness cost through bursty gene expression. An important next step is to confirm our findings using more realistic models of gene expression incorporating additional stochastic effects, such as partitioning errors [18], and correspondingly, to employ various analytical and numerical methods that may permit solution in these more complex cases (e.g., [19,20]). We anticipate that future analysis of such models will provide a deeper understanding of epigenetic interactions between genes, drugs, and population dynamics.

*daniel.charlebois@uottawa.ca
†mkaern@uottawa.ca

[1] M. Kaern, T. Elston, W. Blake, and J. Collins, Nat. Rev. Genet. **6**, 451 (2005).
[2] A. Eldar and M. Elowitz, Nature (London) **467**, 167 (2010).
[3] T. Jia and R. V. Kulkarni, Phys. Rev. Lett. **105**, 018101 (2010).
[4] W. Blake, G. Balazsi, M. Kohanski, F. Isaacs, K. Murphy, Y. Kuang, C. Cantor, D. Walt, and J. Collins, Mol. Cell **24**, 853 (2006).
[5] D. Fraser and M. Kaern, Mol. Microbiol. **71**, 1333 (2009).
[6] Z. Zhang, W. Qian, and J. Zhang, Mol. Syst. Biol. **5**, 299 (2009).
[7] D. Zhuravel, D. Fraser, S. St-Pierre, L. Tepliakova, W. Pang, J. Hasty, and M. Kaern, Syst. Synth. Biol. **4**, 105 2010.
[8] A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, and U. Alon, Nature (London) **444**, 643 (2006).
[9] A. Brock, H. Chang, and S. Huang, Nat. Rev. Genet. **10**, 336 (2009).
[10] V. Shahrezaei, J. Ollivier, and P. Swain, Mol. Syst. Biol. **4**, 196 (2008).
[11] G. Uhlenbeck and L. Ornstein, Phys. Rev. **36**, 823 (1930).
[12] M. Wang and G. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).
[13] D. Charlebois, J. Intosalmi, D. Fraser, and M. Kaern, Commun. Comput. Phys. **9**, 89 (2011).
[14] D. Gillespie, Phys. Rev. E **54**, 2084 (1996).
[15] T. Jia and R. V. Kulkarni, Phys. Rev. Lett. **106**, 058102 (2011).
[16] H. Chang, M. Hemberg, M. Barahona, D. Ingber, and S. Huang, Nature (London) **453**, 544 (2008).
[17] T. Kalmar, C. Lim, P. Hayward, S. Munoz-Descalzo, J. Nichols, J. Garcia-Ojalvo, and A. Arias, PLoS Biol. **7**, e1000149 (2009).
[18] N. Brenner and Y. Shokef, Phys. Rev. Lett. **99**, 138102 (2007).
[19] B. Munsky and M. Khammash, IET Syst. Biol. **2**, 323 (2008).
[20] M. Stamatakis and K. Zygourakis, J. Theor. Biol. **266**, 41 (2010).

Chapter 6

Coherent Feedforward Transcriptional Regulatory Motifs Enhance Drug Resistance

Daniel A. Charlebois, Gábor Balázsi, Mads Kærn. Submitted.

Coherent Feedforward Transcriptional Regulatory Motifs Enhance Drug Resistance

Daniel A. Charlebois[1,2],* Gábor Balázsi[3], and Mads Kærn[1,2,4†]
[1]*Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, Ontario, K1N 6N5, Canada.*
[2]*Ottawa Institute of Systems Biology, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada.*
[3]*Department of Systems Biology-Unit 950, University of Texas MD Anderson Cancer Center, 7435 Fannin Street, Houston, TX, 77054, United States.*
[4]*Department of Cellular and Molecular Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada.*

(Dated: November 28, 2013)

Fluctuations in gene expression give identical cells access to a spectrum of phenotypes that can serve as a transient, nongenetic basis for natural selection by temporarily increasing drug resistance. In this study, we demonstrate using a minimal mathematical model that certain gene regulatory network motifs, specifically the coherent feedforward and positive autoregulatory motifs, can facilitate the development of nongenetic resistance by increasing cell-to-cell variability and the time scale at which beneficial phenotypic states can be maintained. We also demonstrate that a regulatory network known to control the expression of genes conferring resistance to structurally and functionally unrelated drugs can facilitate persistent immunity to drug treatment. Our results highlight how regulatory network motifs enabling transient, nongenetic inheritance play an important role in defining reproductive fitness in adverse environments and provide a selective advantage subject to evolutionary pressure.

I. INTRODUCTION

Gene expression is a stochastic process that enables genetically identical cells to exhibit phenotypic variation [10, 11, 18, 20]. This noise-induced phenotypic variability can provide a fitness advantage in clonal cell populations experiencing the same drug environment [5, 13, 41, 42]. This phenomenon may contribute to limiting the efficacy of drug therapy, including those used to treat disease caused by uncontrolled proliferation in bacterial infections [38] and cancer [6].

It was recently argued by Brock *et al.* [6] that gene expression noise, independent of DNA mutation, may result in enduring and transiently heritable phenotypes that accelerate tumour progression by contributing to the development of drug-resistant cancer cells. In this hypothesis, phenotypic variability arising from noisy expression of a drug resistance gene allows some cells to develop a temporary insensitivity, which in turn could increase the probability that these cells acquire a mutation conferring permanent immunity to the drug.

In a previous study, we investigated the effect of gene expression noise on the reproductive fitness of isogenic cell populations under stress as a first-passage time problem [7]. This study generalized and expanded previous theoretical work that explained the acute effects of drug exposure [13, 41, 42], and considered the interplay between the fluctuation amplitude and the fluctuation frequency in defining the long-term impact of gene expression noise. To analyze the problem in general terms, we used the Ornstein-Uhlenbeck (OU) process [37] to model gene expression in individual cells. This analysis revealed not only that fluctuation frequency is a critical parameter in determining long-term survival, but also that gene expression noise with fluctuation time scales comparable to those measured for certain genes in human cancer cells [36] may allow for the development of permanent drug resistance independently of mutations [7]. However, it remains unclear to what extent these conclusions are supported by more biologically realistic models of gene expression dynamics.

To expand on our previous analysis, we investigate in the present study how the architecture of transcriptional regulatory networks can impact the development of drug resistance using more realistic models of gene expression. This analysis is inspired by the transcriptional regulation of a gene, PDR5, known to provide the budding yeast *Saccharomyces cerevisiae* with resistance to a broad range of drugs [21]. The PDR5 gene is a member of the highly conserved family of ATP binding cassette (ABC) transporters that are the cause of multidrug resistance in microbes [24], fungi [21, 33], and cancer cells [4, 29, 34]. These transporters represent the largest class of transmembrane pumps, and are responsible for

* daniel.charlebois@uottawa.ca
† mkaern@uottawa.ca

the transport of hundreds of substrates, including hormones, lipids, drugs and other toxins, across intracellular and extracellular membranes [34]. Notably, resistance to chemotherapeutic drugs has been correlated to the expression of ABC transporter genes in tumors [29, 34], and the pumps are the target of several anticancer drugs [23].

Budding yeast is an ideal organism for investigating how ABC transporters facilitates the development of drug resistance [29] because it carries a network of genes, the pleiotropic drug resistance (PDR) network, that confers a drug resistance phenotype similar to that of mammalian cells [21]. Among the 16 ABC transporter genes found in budding yeast, the PDR5 gene plays several particularly important role in cellular detoxification [3, 9, 16]. For example, in addition to removing externally added toxic compounds from the cell, the PDR5 protein also exports toxic metabolites that accumulate during growth, transports steroids, and translocates phospholipids across the plasma membrane (see [30] for a review). PDR5 is of particular interest with respect to fungal infections and cancers because the overexpression of drug efflux pumps belonging to the same ABC superfamily in *Candida* and cancerous tumours that are respectively known to confer drug resistance to antibiotics [17, 33] and chemotherapy [4].

The transcriptional regulatory network controlling the transcription of the PDR5 gene consists of a coherent feedforward loop (FFL) with a positive feedback loop (PFL) nested within it (Fig. 1). These motif arises from the transcriptional regulation of PDR5 by two homologue transcription factors (TFs) encoded by PDR1 and PDR3 [3], the latter of which has been shown to be autoregulated [9].

We hypothesize that the combination of a FFL and the PFL motif (FFL+PFL) facilitates the development of drug resistance by enhancing population heterogeneity in PDR5 expression and enabling nongenetic inheritance. This hypothesis is based on previous work demonstrating that the FFL and PFL motifs individually act to increase gene expression noise [1, 27, 35], and that they allow bacterial cells to maintain high gene expression levels following a transient stimulation [19].

To examine the possible contribution of the FFL and PFL motifs to the development of nongenetic drug resistance, we first use a minimal mathematical model to characterize and compare the deterministic and stochastic dynamics of individual PDR5 network components. This is done in Section II. Subsequently, in Section III, we investigate the development of drug resistance in a budding yeast cell population by stochastically simulating a model of the PDR network incorporating drug-dependent gene activation and diffusion across the cellular membrane. The minimal model facilitates a general and comprehensive characterization of each the components of the PDR network. The PDR network model allows us to investigate the results obtained for the minimal model using a more biologically realistic modeling strategy incorporating cellular and fitness dynamics. The results of our analyses demonstrate that the architecture of the PDR5 transcriptional regulatory network may contribute significantly to the resistance conferred by the PDR5 gene, and that certain gene regulatory network motifs may provide an evolutionary advantage by enhancing reproductive fitness under high stress conditions.

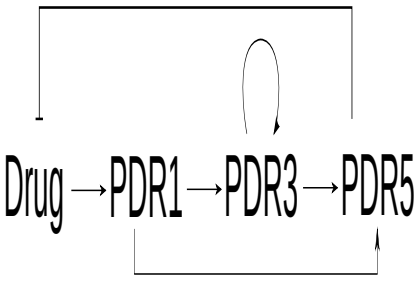


FIG. 1. The PDR5 transcriptional network. Regular arrows denote activation and flat-head arrow denotes repression.

II. MINIMAL MODEL

A. Modeling and Simulation

The PDR5 transcriptional regulatory network in Fig. 1 can be decomposed in to three elements: the direct activation (DA) of PDR5 transcription by PDR1, a FFL that combines DA with indirect activation through PDR3, and a PFL in which PDR3 activates its own expression. This decomposition defines the three distinct networks, DA, FFL, and FFL+PFL, illustrated in Fig. 2 where the three genes, PDR1, PDR3 and PDR5, are labeled X, Y

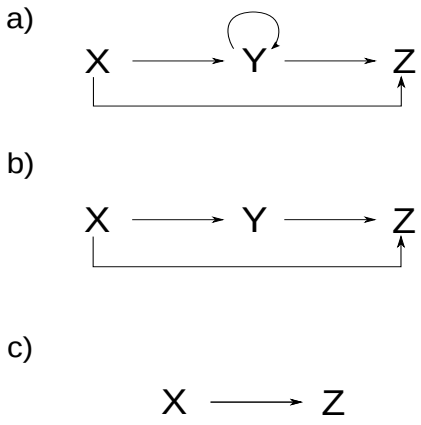


FIG. 2. PDR5 transcriptional network elements considered in the minimal model. (a) Coherent feedforward loop with positive feedback loop (FFL+PFL). (b) Coherent feedforward loop (FFL). (c) Direct activation (DA). X, Y, and Z represent respectively the PDR1, PDR3, and PDR5 genes. Arrows denote activation.

and Z, respectively.

Treating the activity of PDR1 as an adjustable, possibly time-dependent parameter x, the DA, FFL, and FFL+PFL networks have two variables whose dynamics can be described by the following system of coupled ordinary differential equations

$$\frac{dy}{dt} = \alpha_y \omega_1 f_y(x,y) - y \quad (1)$$

$$\frac{dz}{dt} = \alpha_z f_z(x,y) - z, \quad (2)$$

where y and z are respectively the protein concentrations associated with the expression of genes Y and Z, and α_y and α_z are the maximum level of activated protein production for Y and Z, respectively. The dilution and degradation rate of y and z are set to unity. The gene regulatory functions in Eq. (1) and Eq. (2) are defined by the Hill-type functions given by

$$f_y(x,y) = (x+\omega_2 y)^n/(K^n + (x+\omega_2 y)^n) \quad (3)$$

$$f_z(x,y) = (x+y)^n/(K^n + (x+y)^n), \quad (4)$$

where n and K are respectively the Hill coefficient and Hill constant, which for simplicity are here kept at equal values for the two genes. The activation of y by x, and the presence of positive feedback on y, are represented by the Boolean variables ω_1 and ω_2, respectively, and can be either ON ($\omega = 1$) or OFF ($\omega = 0$). Consequently, the model of the DA network is defined by $\omega_1 = \omega_2 = 0$, the FFL network by $\omega_1 = 1$ and $\omega_2 = 0$, and the FFL+PFL by $\omega_1 = 1$ and $\omega_2 = 1$. These equations were solved numerically using a medium order MATLAB intrinsic non-stiff differential equation solver (ode45).

Equations (1)-(2) can be translated into the following set of birth-death processes

$$\oslash \xrightarrow{k_y} y \quad (5)$$

$$\oslash \xrightarrow{k_z} z \quad (6)$$

$$y \xrightarrow{\delta_y} \oslash \quad (7)$$

$$z \xrightarrow{\delta_z} \oslash, \quad (8)$$

where Eq. (5) and Eq. (6) respectively describe the production of y and z. In Eq. (5), $k_y = \alpha\omega_1 f_y(x,y)$, and in Eq. (6), $k_z = \alpha f_z(x,y)$, where $f_y(x,y)$ is described in Eq. (3) and $f_z(x,y)$ in Eq. (4). The degradation of y and z are described by Eq. (7) and Eq. (8), respectively, where the degradation rates δ_y and δ_z are set to unity. Stochastic simulation of the chemical reactions were performed using the Gillespie algorithm [14].

To investigate the development of nongenetic drug resistance in clonal cell populations expressing one of these network topologies, we perform population-level simulations at single-cell resolution using the population dynamics algorithm (PDA) [8]. The PDA combines an exact method to simulate molecular-level fluctuations in single cells and a constant-number Monte Carlo approach to simulate the statistical characteristics of growing cell populations. In these simulations, gene expression in each of N individual cells is obtained by stochastically simulating Eqs. (5)-(8). Simulations were initiated by drawing the initial values of the cell cycle clock from a uniform distribution $[0, t_D]$, where t_D is the cell division time in absence of selection. Cell volume (v) was modeled using an exponential growth law

$$v(t_{div}) = v_0 2^{t_{div}/w(z)t_D}, \quad (9)$$

where v_0 is the initial volume and t_{div} is the time since last division. At cell division, t_{div} is reset to zero and the cell volume reset to v_0. We first model microscopic fitness (w), that is the reproductive fitness of an individual cell in the presence of a drug, using a step function. If z falls below a critical concentration (z_c) the cell is flagged and is subsequently unable to reproduce or change its protein levels. Then, we model microscopic fitness using a Hill function

$$w(z) = z^{n_w}/(K_w^{n_w} + z^{n_w}), \quad (10)$$

where n_w and K_w are respectively the Hill coefficient and the Hill constant used to set the fitness threshold. The macroscopic fitness (W) of the population is determined by the number of cell divisions that occur during a given generation divided by the fixed number of cells in the population.

All timescales in this study are reported with respect to t_D, which set to unit time. No qualitative difference in the results presented below was observed for up to a two-fold change in parameters.

B. Results and Discussion

1. *FFLs accelerate and prolong transcriptional responses*

To characterize the behaviour of the DA, FFL, and FFL+PFL networks following changes in an upstream activating signal x, the response time t_{ON} and the relaxation time t_{OFF} for different values x were obtained. t_{ON} was defined as the time for z to rise from zero to 50% of the steady-state value corresponding to the DA network when x is turned ON. t_{OFF} was similarly defined as the time, after x is turned OFF, for z to fall to 50% of the corresponding steady-state value when x is ON. The 50% DA steady-state value was chosen to ensure a controlled comparison of the three network topologies, such that the response and relaxation times for each network were determined for the same absolute change in z.

Both coherent feedforward networks decrease t_{ON} compared to the DA network [Fig. 3(a)]. There is a minimum t_{ON} at an x of about 0.5 for the FFL and FFL+PFL networks. The FFL has the quickest response for non-zero values of x less than about 8, when the FFL and FFL+PFL t_{ON} values begin to converge. This result is particularly interesting as positive autoregulation on its own generally increases response time [19, 25]. The t_{ON}

for the FFL and FFL+PFL approaches the t_{ON} of the DA network for higher levels of x. t_{ON} for the DA network is unaffected by varying x.

The FFL and FFL+PFL networks exhibit prolonged activation times relative to the DA network [Fig. 3(b)]. These results are in qualitative agreement with previous theoretical predictions [26] and experimental results [19]. Notably, the FFL+PFL network has the longest t_{OFF} for x values larger than 1, and the FFL network has the longest t_{OFF} for non-zero x values less than 1. The t_{OFF} for the DA network is unaffected by changing x values [Fig. 3(b)].

When considering the t_{ON} and t_{OFF} together for x values less than 1, the FFL has both a larger t_{ON} and t_{OFF} compared to the FFL+PFL and DA networks (Fig. 3). When x is increased by an order of magnitude, though the three networks have similar t_{ON} values, the FFL+PFL network has a larger t_{OFF}. These results suggest the if z confers drug resistance, the FFL network provides a fitness advantage when the activating signal (i.e., the drug dose) is low, and the FFL+PFL network a fitness advantage when the drug dose is high.

2. *FFLs provide stable high expression in fluctuating environments*

To investigate how the three networks respond to a fluctuating upstream activating signal, x was set to switch between ON and OFF at a frequency Ω_x.

When x fluctuates periodically, the mean concentration of z (μ_z) is higher in the FFL+PFL network compared to the other two networks [Fig. 4(a)], despite all three networks having the same mean when x is held constant (data not shown). μ_z is higher in the FFL network than the DA network, with both networks increasing μ_z with increasing values of Ω_x until they level off at around $\Omega_x = 2$.

The frequency-response plot for the three networks shows that the standard deviation of z (σ_z) in the FFL and FFL+PFL is lower than σ_z for the DA when Ω_x is less than about 4.5 [Fig. 4(b)]. As Ω_x is increased above 4.5, σ_z for the DA network falls below the σ_z values for the FFL and the FFL+PFL networks. The FFL+PFL network has lower values of σ_z than the FFL network until Ω_x is increased to about 2, then the σ_z value for the FFL falls below that of the FFL+PFL for higher values of Ω_x.

When x is set to fluctuate randomly, the FFL+PFL network provides stable high expression compared to the FFL and DA networks [Fig. 4(c)].

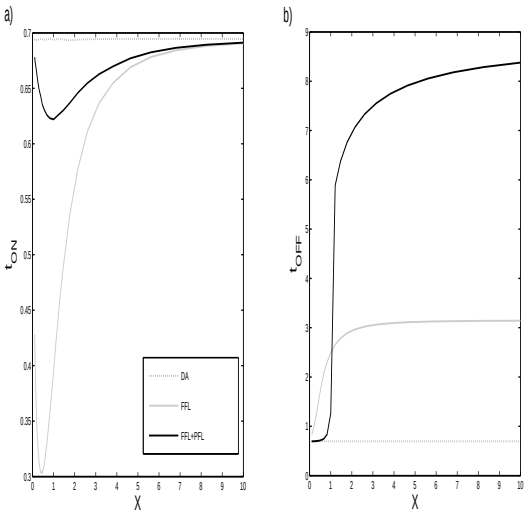


FIG. 3. Coherent feedforward networks enable fast and prolonged activation. (a) Response time t_{ON} (time for z to reach 50% of steady-state level) for the DA, FFL, and FFL+PFL networks as a function of an activating signal x. (b) Relaxation time t_{OFF} (time for z to fall to 50% of steady-state level) for the same network motifs considered in (a) as a function of x. Parameters were set to: $\alpha = 10$, $n = 2$, and $K = 1$.

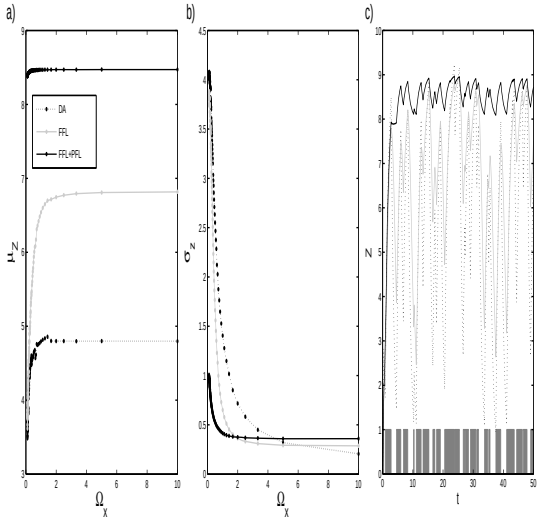


FIG. 4. Response of z to a fluctuating upstream activating signal. (a) The mean of z (μ_z) is shown for the DA, FFL, and FFL+PFL networks as a function of the ON-OFF switching frequency of x (Ω_x). (b) The standard deviation of z (σ_z) is shown for the same network motifs considered in (a) as a function of Ω_x. (c) z as a function of a randomly fluctuating Ω_x (dark gray rectangles denote presence of an activating signal) for the same network motifs considered in (a). Parameters were set to: $x_{ON} = 5$, $x_{OFF} = 0$, $\alpha = 10$, $n = 2$, $K = 1$.

3. FFLs increase population heterogeneity and mixing times

The set of chemical reactions [Eqs. (5)-(8)] corresponding to the deterministic model [Eqs. (1)-(2)] were simulated stochastically to compare the noise and mixing times of the three networks. The mixing time was defined previously by Sigal *et al.* [36] as the time for the autocorrelation function to decay by half.

The noise in network output z ($\eta_z = \sigma_z/\mu_z$) for a given μ_z is the highest for the FFL+PFL network [Fig. 5(a)]. The increase in noise due to positive feedback is expected as it amplifies fluctuations [2]. η_z for the DA and FFL are similar, with η_z for all three networks beginning to converge for μ_z around 5.

As x is varied, the relaxation times are highest for the FFL+PFL network, followed by the FFL network and then the DA network [Fig. 5(b)]. The longer relaxation time in the FFL+PFL network compared to the DA network is qualitatively in agreement with results found experimentally by Kalir *et al.* [19].

4. FFLs enhance drug resistance

Gene expression [Eqs. (5)-(8)] was coupled to population dynamics using the PDA [8] to investigate the effects of feedforward network motifs on drug resistance.

Prior to the application of the drug at generation 10 the fitness is 1, as all the cells in the population divide once per generation (Fig. 6). When we consider a threshold fitness function, all the cells in DA population are unable to reproduce immediately following application of the drug [Fig. 6(a)]. A significant fraction of cells in the FFL and FFL+PFL populations remain fit even after 40 generations of drug treatment.

Next we model microscopic fitness using Eq. (10). Interestingly, in this model, drug resistance develops in all three populations [Fig. 6(b)]. A much lower number of cells in the generation subsequent to the application of the drug reproduce as a result of low z level due to the transient time of z to the new steady-state. After a couple of oscillations, the fitness levels off such that a fraction of the pre-treatment population reproduces in each generation. The fraction of fit cells during drug treatment is roughly double for the FFL and FFL+PFL populations compared to the DA population.

These results suggest that in natural populations the FFL+FPL network architecture may provide cells with a fitness advantage in adverse environments. This hypothesis is further supported by the fact that the FFL+PFL network forms the topology of a network which regulates the expression of the PDR5 multidrug resistance conferring protein in yeast. It is plausible that the FFL+PFL architecture of the PDR5 transcriptional network inves-

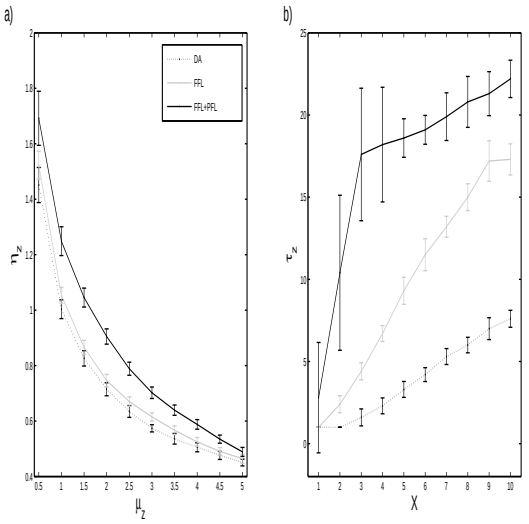


FIG. 5. Coherent feedforward networks increase noise and relaxation time. (a) The noise in z (η_z) for the DA, FFL, and FFL+PFL networks as a function of the mean level of z (μ_z). (b) The relaxation times for z (τ_z) for the same network motifs considered in (a) as a function of an activating signal x. Parameters were set to: $\alpha = 10$, $n = 2$, $K = 10$. 10 realization for 10^3 arbitrary time units were performed. Error bars show standard deviation.

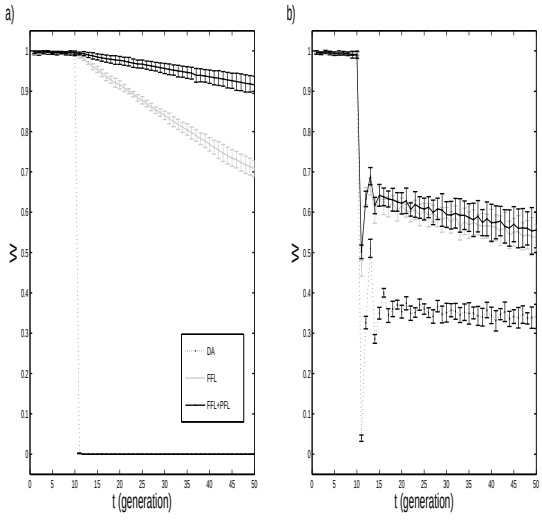


FIG. 6. Effect of network topology and fitness threshold on reproductive fitness (W). W for each of the DA, FFL, and FFL+PFL populations over 50 generations. Drug treatment is initiated at the tenth generation. (a) Population simulations using a step fitness function. (b) Population simulations performed using a Hill type fitness function. Parameters were set to: $x = 1$, $\alpha_y = 10$, $\alpha_z = 100$, $K = 1$, and $n = 2$. In (a) $z_c = 35$ and in (b) $K_w = 35$, and $n_w = 2$. 10 realizations of 1000 cells were performed. Error bars show standard deviation.

tigated in the next section provided yeast cells with an fitness advantage and evolved by means of natural selection.

III. PDR5 TRANSCRIPTIONAL NETWORK MODEL

A. Modeling and Simulation

In this section, we develop a mathematical model and simulate the dynamics of the PDR5 transcriptional regulatory network in order to investigate the development of drug resistance in the context of a more biologically realistic feedforward genetic network, incorporating passive and active diffusion of a drug across the cellular membrane. To the best of our knowledge, this section presents the first modeling of the PDR5 transcription network despite years of experimental investigation.

To examine if the conclusions from the analysis of minimal network models have bearing on PDR5-mediated drug resistance, we analyzed a model of the regulatory network in Fig. 1. This network differs from the FFL+PFL network in Section II by the presence of a negative feedback loop caused by the PDR5 efflux pump eliminating drugs and toxins from the cell. We note that the additivity of the gene regulatory functions in the minimal model [Eqs. (1)-(2)] are justified in the context of PDR5 transcriptional regulation because PDR1 and PDR3 are highly homologous and bind to the same elements in the PDR5 promoter [3]. Consequently, the PDR5 network can be modeled by extending the minimal model to include the upstream activating factor ($PDR1$) and the intercellular drug ($drug_{int}$) as dynamic variables. The resulting ordinary differential equations describing the network are given by

$$\frac{dPDR1}{dt} = \alpha_0 + \alpha_1 \frac{drug_{int}}{K_1 + drug_{int}} - \delta_1 PDR1 \quad (11)$$

$$\frac{dPDR3}{dt} = \alpha_3 \frac{(PDR1 + PDR3)^{n_3}}{(K_3^{n_3} + (PDR1 + PDR3)^{n_3})} - \delta_3 PDR3 \quad (12)$$

$$\frac{dPDR5}{dt} = \alpha_5 \frac{(PDR1 + PDR3)^{n_5}}{(K_5^{n_5} + (PDR1 + PDR3)^{n_5})} - \delta_5 PDR5 \quad (13)$$

$$\frac{d(drug_{int})}{dt} = k_{diff}(drug_{ext} - drug_{int}) - \frac{k_{pump} PDR5 drug_{int}}{k_{pump} + drug_{int}}, \quad (14)$$

when it is assumed that the drug enters and leaves the cell through a combination of passive and active transport, and that the activation of PDR1 by the drug can be captured by Michaelis-Menten kinetics. Eq. (11) describes the activation of $PDR1$ by $drug_{int}$, where α_0 is the basal rate of transcription and $\alpha_0 + \alpha_1$ the maximal activated rate of transcription. Eqs. (12) and (13) are the same as those presented in Eqs. (1)-(2). The last equation, Eq. (14), describes the passive diffusion of the drug across the cell membrane (1st term on the R.H.S.) as well as the pumping of the drug out of the cell via $PDR5$ (2nd term on the R.H.S.). In Eq. (14), $drug_{ext}$ is the extracellular drug concentration, k_{diff} the rate of passive diffusion across the cell membrane, and k_{pump} the rate of PDR5 mediated drug efflux.

The PDR network model [Eqs. (11)-(14)] can be translated into the corresponding birth-death processes

$$\oslash \xrightarrow{k_1} PDR1 \quad (15)$$

$$\oslash \xrightarrow{k_3} PDR3 \quad (16)$$

$$\oslash \xrightarrow{k_5} PDR5 \quad (17)$$

$$PDR1 \xrightarrow{\delta_1} \oslash \quad (18)$$

$$PDR3 \xrightarrow{\delta_3} \oslash \quad (19)$$

$$PDR5 \xrightarrow{\delta_5} \oslash \quad (20)$$

$$\oslash \xrightarrow{k_{drug_{int}}} drug_{int} \quad (21)$$

$$drug_{int} \xrightarrow{\delta_{drug_{int}}} \oslash, \quad (22)$$

where $k_1 = \alpha_0 + \alpha_1 drug_{int}^{n_1}/(K_1^{n_1} + drug_{int}^{n_1})$, $k_3 =$

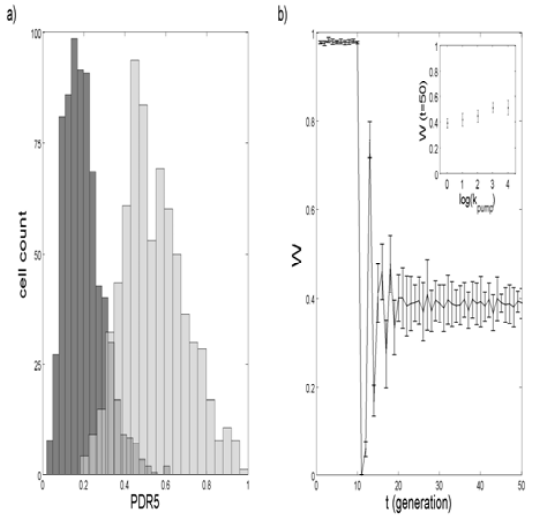


FIG. 7. Adaptation and fitness (W) facilitated by the PDR5 transcriptional network. (a) Number of cells in the population has the corresponding PDR5 expression level. Distributions are shown for the generation prior (dark gray) to drug application and 40 generations after (light gray) drug application. (b) W over 50 generations. Inset shows W at generation 50 for different rates of PDR5 mediated drug efflux (k_{pump}). Drug treatment is initiated at the tenth generation. Unless otherwise indicated, parameters were set to: $\alpha_0 = 1$, $\alpha_1 = 10$, $K_1 = 1$, $n_1 = 1$, $\alpha_3 = 10$, $K_3 = 1$, $n_3 = 2$, $\alpha_5 = 100$, $K_5 = 20$, $n_5 = 2$, $k_{diff} = 100$, $k_{pump} = 1$, and $drug_{ext} = 100$. 10 realizations of 1000 cells were performed.

$\alpha_3(PDR1 + PDR3)^{n_3}/(K_3^{n_3} + (PDR1 + PDR3)^{n_3})$, $k_5 = \alpha_5(PDR1+PDR3)^{n_5}/(K_5^{n_5}+(PDR1+PDR3)^{n_5})$, $k_{drug_{int}} = k_{diff}drug_{ext}$, and $\delta_{drug_{int}} = k_{diff} + k_{pump}PDR5/(k_{pump} + drug_{int})$. Eqs. (15)-(17) respectively describe the production of $PDR1$, $PDR3$, and $PDR5$. The degradation of $PDR1$, $PDR3$, and $PDR5$ is described by Eqs. (18)-(20), respectively. The passive diffusion of the drug into the cell is described by Eq. (21). The removal of the drug from the cell by both passive diffusion and pumping is described by Eq. (22). Here we model cell growth using Eq. (9) and cellular fitness in the presence of a drug as follows

$$w(z) = PDR5/(drug_{int} + PDR5). \quad (23)$$

This equation describes cellular fitness increasing with increasing PDR5 relative to the intracellular drug concentration, and it assumes that there is no fitness cost associated with maintaining a high level of PDR5.

B. Results and Discussion

In order to investigate if persistent nongenetic drug resistance would develop in the PDR5 transcriptional network model, we tracked cellular and fitness dynamics over 50 generations.

In Fig. 7a, the population PDR5 histogram prior to drug treatment is shown (generation 9) together with the population PDR5 histogram after 40 generations of drug treatment (generation 50). The corresponding mean PDR5 expression increases three fold upon application of the drug, in agreement with preliminary experimental data obtained in our laboratory for budding yeast populations after 24 hours of drug (Nocodazole) treatment (data not shown).

The resulting drug resistance dynamics in Fig. 7b are similar to those obtained using the minimal model with a Hill type fitness function (Fig. 6b). Namely, a stable fraction of reproductively viable cells develops after about 10 generations of drug treatment. When the rate of passive diffusion (k_{diff}) and extracellular drug concentration ($drug_{ext}$) are changed the level of fitness changes accordingly. For instance, when k_{diff} and $drug_{ext}$ are decreased 10 fold, the steady-state W increases to 0.9 (data not shown).

The main difference between the minimal model and the PDR5 model is the incorporation of negative feedback on the activating signal in the latter. In order to investigate the effects of the negative feedback on fitness, we varied k_{pump} over several orders of magnitude (inset Fig. 7b). Fitness after 40 generations of drug treatment increased from 0.39 when $k_{pump} = 1$ to 0.51 when $k_{pump} = 10^4$. As expected, when $k_{pump} = 0$, W is zero for all generations subsequent to drug application (data not shown).

These results demonstrate that the presence of a negative feedback in a drug-efflux pump network does not impede the development of persistent nongenetic drug resistance. Increasing the strength of the negative feedback had little effect on the fraction of drug resistant cells in the population. This is because although PDR5 functions to increase cellular fitness by actively pumping the drug out of the cell, it also reduces its own activation by indirectly reducing the activity of PDR1.

IV. CONCLUSION

This study demonstrates how certain transcriptional regulatory network motifs can facilitate the development of drug resistance by providing a broader spectrum of potentially advantageous fitness phenotypes upon which selection can act, and by enabling the their nongenetic inheritance to subsequent generations. This is important because while it is well established that genetic mutations can cause drug tolerance (e.g., [12, 15, 31, 32, 40]), less is known about how gene expression noise can influence drug resistance. Gaining an understanding of the genetic and nongenetic mechanisms underlying drug resistance, and in particular multidrug resistance, is critical for dealing with situations ranging from bacterial infections to cancer.

The results presented in our investigation suggest that the topology of the PDR5 network may have provided an evolutionary advantage for yeast cells. This network topology was found to decrease activation time, increase relaxation time, and increase noise. It was also able to buffer against a fluctuating drug environment. Fungal drug resistance is an especially important issue due to the limited number of antifungal compounds [22] and the increasing number of immunocompromised patients worldwide relying on these drugs [28, 39]. Our study provides novel foundational knowledge on the development of drug resistance and presents new opportunities to address this growing problem.

ACKNOWLEDGMENTS

The authors would like to thank Afnan Azizi for helpful discussions. D.C. was supported financially by a Queen Elizabeth II Graduate Scholarship in Science and Technology from the Government of Ontario and an Excellence Scholarship from the University of Ottawa. G.B. was supported by the NIH Directors New Innovator Award Program (Grant No: 1DP2 OD006481-01).

[1] M. Acar, A. Becskei, and A. van Oudenaarden. Enhancement of cellular memory by reducing stochastic transitions. *Nature*, 435:228–232, 2005.

[2] U. Alon. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, 8:450–461, 2007.

[3] E. Balzi and A. Goffeau. Yeast multidrug resistance: The pdr network. *J. Bioenerg. Biomembr.*, 27:71–76, 1995.

[4] B.E. Bauer and K. Kuchler H. Wolfger. Inventory and function of yeast abc proteins: about sex, stress, pleiotropic drug and heavy metal resistance. *Biochimica et Biophysica Acta*, 1461:217–236, 1999.

[5] W.J. Blake, G. Balazsi, M.A. Kohanski, and F.J. Issacs and *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Molec. Cell*, 24:853–865, 2006.

[6] A. Brock, H. Chang, and S. Huang. Non-genetic heterogeneity - a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.*, 10:336–342, 2009.

[7] D.A. Charlebois, N. Abdennur, and M. Kaern. Gene expression noise facilitates adaptation and drug resistance independently of mutation. *Phys. Rev. Lett.*, 107:218101, 2011.

[8] D.A. Charlebois, J. Intosalmi, D. Fraser, and M. Kaern. An algorithm for the stochastic simulation of gene expression and heterogeneous population dynamics. *Commun. Comput. Phys.*, 9:89–112, 2011.

[9] A. Delahodde, T. Delaveau, and C. Jacq. Positive autoregulation of the yeast transcription factor pdr3p, which is involved in control of drug resistance. *Mol. Cell. Biol.*, 15:4043–4051, 1995.

[10] A. Eldar and M.B. Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467:167–173, 2010.

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

[12] M.R. Farhat, B.J. Shapiro, K.J. Kieser, and R. Sultana and *et al.* Genomic analysis identifies targets of convergent positive selection in drug-resistant mycobacterium tuberculosis. *Nat. Genet.*, 2013.

[13] D. Fraser and M. Kaern. A chance at survival: gene expression noise and phenotypic diversification strategies. *Molec. Microbiol.*, 71:1333–1340, 2009.

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

[15] M.E. Gorre, M. Mohammed, K. Ellwood, and N. Hsu and *et al.* Clinical resistance to sti-571 cancer therapy caused by bcr-abl gene mutation or amplification. *Science*, 293:876–880, 2001.

[16] K. Gulshan and W. Scott Moye-Rowley. Multidrug resistance in fungi. *Eukaryot. Cell.*, 6:1933–1942, 2007.

[17] K. Izumikawa, H. Kakeya, H.F. Tsai, and B. Grimberg and *et al.* Function of candida glabrata abc transporter gene, pdh1. *Yeast*, 20:249–261, 2003.

[18] M. Kaern, T.C. Elston, W.J. Blake, and J.J. Collins. Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.*, 6:451–464, 2005.

[19] S. Kalir, S. Mangan, and U. Alon. A coherent feed-forward loop with a sum input function prolongs flagella expression in escherichia coli. *Mol. Sys. Biol.*, 2005.

[20] B.B. Kaufmann and A. van Oudenaarden. Stochastic gene expression: from single molecules to the proteome. *Curr. Opin. Genet. Dev.*, 17:107–112, 2007.

[21] M. Kolaczkowski, A Kolaczowska, J Luczynski, S Witek, and A Goffeau. In vivo characterization of the drug resistance profile of the major abc transporters and other components of the yeast pleiotropic drug resistance network. *Microb. Drug. Resist.*, 4:143–148, 1998.

[22] D.P. Kontoyiannis and R.E. Lewis. Antifungal drug resistance of pathogenic fungi. *Lancet*, 359:1135–1144, 2002.

[23] G.D. Leonard, O. Polgar, and S.E. Bates. Abc transporters and inhibitors: new targets, new agents. *Curr. Opin. Investig. Drugs*, 3:1652–1659, 2002.

[24] J. Lubelski, W.N. Konings, and A.J.M. Driessen. Distribution and physiology of abc-type transporters contributing to multidrug resistance in bacteria. *Microbiol. Mol. Biol. Rev.*, 71:463–476, 2007.

[25] Y.T. Maeda and M. Sano. Regulatory dynamics of synthetic gene networks with positive feedback. *J. Mol. Biol.*, 359:1107–1124, 2006.

[26] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *PNAS*, 100:11980–11985, 2003.

[27] D. Orrell and H. Bolouri. Control of internal and external noise in genetic regulatory networks. *J. Theor. Biol.*, 230:301–312, 2004.

[28] J. Perlroth, B.Choi, and B. Spellberg. Nosocomial fungal infections: epidemiology, diagnosis, and treatment. *Med. Mycol.*, 45:321–346, 2007.

[29] A. Persidis. Cancer multidrug resistance. *Nat. Biotechnol.*, 17:94–95, 1999.

[30] R. Prasad and A. Goffeau. Yeast atp-binding cassette transporters conferring multidrug resistance. *Annu. Rev. Microbiol.*, 66:39–63, 2012.

[31] C. Roche-Lestienne, J.L. Laï, S. Darré, and T. Facon *et al.* A mutation conferring resistance to imatinib at the time of diagnosis of chronic myelogenous leukemia. *N. Engl. J. Med.*, 348:2265–2266, 2003.

[32] H. Safi, S. Lingaraju, A. Amin, and S. Kim and *et al.* Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl--d-arabinose biosynthetic and utilization pathway genes. *Nat. Genet.*, 2013.

[33] D. Sanglard and F.C. Odds. Resistance of candida species to antifungal agents: molecular mechanisms and clinical consequences. *Lancet Infect. Dis.*, 2:73–85, 2002.

[34] K.W. Scotto. Transcriptional regulation of abc drug transporters. *Oncogene*, 22:74967511, 2003.

[35] V. Shahrezaei, J.F. Ollivier, and P. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, 4, 2008.

[36] A. Sigal, R. Milo, A. Cohen, and N. Geva-Zatorsky and *et al.* Variability and memory of protein levels in human cells. *Nature*, 444:643–646, 2008.

[37] G.E. Uhlenbeck and L.S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36:823–841, 1930.

[38] Y. Wakamoto, N. Dhar, R. Chait, and K. Schneider and *et al.* Dynamic persistence of antibiotic-stressed mycobacteria. *Science*, 339:91–95, 2013.

[39] D.W. Warnock. Fungal diseases: an evolving public health challenge. *Med. Mycol.*, 44:697–705, 2006.

[40] H. Zhang, D. Li, L. Zhao, and J. Fleming and *et al.* Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from china identifies genes and intergenic regions associated with drug resistance. *Nat. Genet.*, 2013.

[41] L.J. Zhang, S.W. Yan, and Y.Z. Zhuo. A dynamical model of dna-damage derived p53-mdm2 interaction. *Acta Physica Sinica*, 56:2442–2447, 2007.

[42] D. Zhuravel, D. Fraser, S. St-Pierre, and L. Tepliakova and *et al.* Phenotypic impact of regulatory noise in cellular stress-response pathways. *Syst. Synth. Biol.*, 4, 2010.

Chapter 7

Other contributions

In addition to the research presented in this thesis, at the beginning of my PhD I co-authored a book chapter (my contribution was based in part on my master's thesis research):

- **Daniel A. Charlebois**, Theodore J. Perkins, Mads Kærn. "Stochastic Gene Expression and the Processing and Propagation of Noisy Signals in Genetic Networks", in *Information Processing and Biological Systems*, A.S. Ribeiro and S. Niiranen (Eds.), Springer-Verlag, pg. 89-112, ISBN: 978-3-642-19620-1.

During the course of my doctoral studies I contributed to the master's research of Mr. Nezar Abdennur, an alumni of Dr. Mads Kærn's group. The follow manuscript based on this work is presently in preparation:

- "A general framework for simulating heterogeneous cell populations at single cell resolution". Nezar Abdennur, **Daniel A. Charlebois**, Andrei Anisenia, Mads Kærn.

In this study, I helped with benchmarking of the algorithm, model development, and simulations. Specifically, Mr. Abdennur and I developed models of resource

competition between two cell populations and replicative aging in budding yeast. I performed the simulations of these models (see Sections 4.2 and 4.3). I also wrote a draft of the manuscript.

I recently reviewed a book on numerical methods that is to be published in the near future:

- Review of "Number-Crunching: Taming Unruly Computational Problems from Mathematical Physics to Science Fiction", by Paul Nahin. **Daniel A. Charlebois**, *Physics in Canada*, in press.

Finally, my research was presented in poster and oral form at a number of scientific conferences listed below:

- 3rd Student/Postdoc Poster Day in Computational Biology and Biomedical Informatics, October 17, 2013, University of Ottawa, Ottawa, Canada. "Implications of Gene Network Architecture on Variations in Gene Expression" (poster presentation). Ian Roney, **Daniel A. Charlebois**, Mads Kærn.

- 3rd Student/Postdoc Poster Day in Computational Biology and Biomedical Informatics, October 17, 2013, University of Ottawa, Ottawa, Canada. "Feed-forward and feedback loop motifs conjoin in PDR5 regulatory network to enhance drug resistance in yeast" (poster presentation). Afnan Azizi, **Daniel A. Charlebois**, Mads Kærn.

- International Conference on Systems Biology, August 30-September 3, 2013, Copenhagen, Denmark. "Modeling and Experimental Investigation of the PDR5 Network Architecture and its Effects on Drug Resistance in Yeast" (poster presentation). Afnan Azizi, **Daniel A. Charlebois**, Mads Kærn.

- Institute for Systems Biology Symposium 2013, June 11-12, 2013, Mont Tremblant, Canada. "Modeling and Experimental Investigation of the PDR5 Network Architecture and its Effects on Drug Resistance in Yeast" (poster presentation). Afnan Azizi, **Daniel A. Charlebois**, Mads Kærn.

- Mathematical Tools for Evolutionary Systems Biology, May 26-31, 2013, BIRS, Banff, Canada. "Modeling & Simulation of Cellular Population Dynamics: The Case for Noise-Mediated Drug Resistance" (oral presentation). **Daniel A. Charlebois**, Mads Kærn.

- 4th IRCM Meeting on Systems Biology, Institut de recherches cliniques de Montréal, April 2-3, 2013, Montreal, Canada. "PDR5 network architecture and its effects on drug resistance in yeast" (poster presentation). Afnan Azizi, **Daniel A. Charlebois**, Mads Kærn.

- Gordon Research Conference on Stochastic Physics in Biology, January 13-18, 2013, Ventura, USA. "Coloured gene expression noise endows in silico cell populations with drug resistance independently of mutations" (poster presentation). **Daniel A. Charlebois**, Nezar Abdennur, Mads Kærn.

- International Conference on Stochastic Processes in Systems Biology, Genetics & Evolution, August 21-25, 2012, Rice University, Houston, USA. "Coloured gene expression noise endows in silico cell populations with drug resistance independently of mutations" (poster presentation). **Daniel A. Charlebois**, Nezar Abdennur, Mads Kærn.

- Cell Symposia: Epigenetics and the Inheritance of Acquired States, October 31-November 2, 2011, Boston, USA. "Coloured gene expression noise endows in silico cell populations with drug resistance independently of mutations" (poster

presentation). **Daniel A. Charlebois**, Nezar Abdennur, Mads Kærn.

- Student/Postdoc Poster Day in Computational Biology and Biomedical Informatics, October 24, 2011, University of Ottawa, Ottawa, Canada. "A framework for individual based simulation of heterogeneous cell populations" (poster presentation). Nezar Abdennur, Andrei Anisenia, **Daniel Charlebois**, Mads Kærn.

- RECOMB 2010: 3rd Annual Joint Conference on Systems Biology, Regulatory Genomics and Reverse Engineering Challenges, November 16-20, 2010, Columbia University, New York, USA. "A framework for ensemble simulations of cell population heterogeneity" (poster presentation). Nezar Abdennur, **Daniel Charlebois**, Mads Kærn.

Chapter 8

Conclusion

> In theory, theory and practice are the same. In practice, they are not.
>
> - Albert Einstein

My work over the past five years, both as a master's and a doctoral student under the supervision of Dr. Mads Kærn, can be divided into two components: 1) the development of algorithms to investigate gene expression and cellular population dynamics, and 2) the application of these tools to investigate the effects of noise and network topology on the fitness of cell populations under stress. This work has resulted in novel accurate and efficient cell population simulation frameworks (Chapters 3 and 4) and has furthered our understanding of the development of epigenetic drug resistance (Chapters 5 and 6).

8.1 Population Simulation Algorithms

Biological systems are not exempt from the laws of physics. However, biophysics is much more than doing physics with the names of the variables changed. A model must account for both the underlying physics and relevant biological features of the

system under investigation. An accurate mathematical model is a powerful tool that can lead to a deeper understanding of a system and can be used to perform *in silico* experiments when *in vivo* or *in vitro* experimentation is not possible due to technological or other practical constraints. These models can also be used to identify from the state space of possible wet laboratory experiments the ones most likely to yield significant results. This is particularly important as experiments in the biological sciences are often resource intensive. Furthermore, model development is an iterative process, where the model is refined based on the experimental results, which then leads to new experiments, and so forth.

The approach used to develop models in this thesis has been, as eloquently stated by Einstein, that "everything should be made as simple as possible, but not simpler". However, even with this in mind, it is astonishing how rapidly a model can become analytically intractable with the addition of a small amount of biological complexity. Thus, rather than trying to develop exact analytical solutions for a very limited subset of possible cases, I developed simulation algorithms that could model the entire cell populations at the single-cell level with an appropriate degree of biological complexity.

Chapter 3 presents an Accelerated Method for Simulating Population Dynamics. The AMSPD employs a single time series to simulate the dynamics of every cell in the population. Simulation results from this algorithm were found to be in good agreement with those produced using the Population Dynamics Algorithm (developed and published during my master's degree and used to perform the simulations in Chapters 5 and 6 of this thesis), which was previously benchmarked against analytical solutions [1]. The AMPSD algorithm was found to be significantly faster than the PDA. This is extremely useful as runtimes using the PDA can be extensive. However, the trade off for the gain in speed is that the AMPSD is valid only when steady-state

[1] D.A. Charlebois, J. Intosalmi, D. Fraser, M. Kærn. *Commun. Comput. Phys.*, 9:89-112, 2011.

and symmetric division assumptions hold. The AMSPD is an ideal tool for performing large parameter scans to identify parameter regimes of interest for more exact methods, such as the PDA or the object-oriented framework presented in Chapter 4. Thus, the AMPSD and PDA are tools that a researcher can use together to optimize the model development and simulation process.

Chapter 4 presents models incorporating the biologically relevant phenomena of cellular aging and competition that were originally developed as examples for an object-oriented simulation framework developed by a former master's student Mr. Nezar Abdennur [2]. This framework allows for cell communication, an important phenomenon that is difficult to incorporate efficiently in the AMSPD and PDA algorithms due to the parallel nature of their design. The models and simulations presented exemplified the utility of the object-oriented framework. Furthermore, the replicative aging example provides a novel mathematical model of this phenomenon in budding yeast. The numerical results were found to be in qualitative agreement with experimental observations. Additionally, results from the aging model provide a concrete example of why the constant-number Monte Carlo method should be used in population simulations over cell chains when fitness is incorporated into the model. The results of the cell competition model suggest that a relatively small phenotypic change in one of the subpopulations can have a significant effect on the composition of the population and, importantly, allow wild-type cells to outcompete mutant cells. The synchronous method proved to be particularly suited for simulating cell competition. A potential future application of this framework is to simulate coevolving populations, for example, the increased exchange of resistance genes between phage and bacteria in the gut during antibiotic treatment [3].

[2]N. Abdennur. Master's thesis, University of Ottawa, 2012.

[3]S.R. Modi, H.H. Lee, C.S. Spina, J.J. Collins. *Nature*, 499:219-223, 2013.

As a result of my graduate research, there are now several new algorithms available to investigate cell population dynamics. However, there is still a need for a cell population simulator that can be readily used by other scientists. The AMSPD and PDA were developed in Fortran and their use requires familiarity with the source code. Also, extensive modification can be required when a new model is to be simulated. The object-oriented framework presented in Chapter 4 was developed in part to address these concerns. This framework has now been implemented in the interpreted languages of Matlab and Python, but as a result is not sufficiently fast to perform complex and large scale population simulations. Though using Python does have certain advantages, namely, it is high-level dynamically typed language which renders program development and user modification relatively straightforward. We are presently exploring the possibility of using Cython to develop C extensions to speedup performance-critical parts. A C++ version of this framework was developed to decrease runtimes, but still requires extensive customization to handle general inputs. Since C++ is a low-level statically typed language, the development of a general, fast, and user friendly population simulator has proven challenging. This project is a work in progress and will likely involve the collaboration of several laboratories as well as professional programmers.

8.2 Epigenetic Drug Resistance

Drug resistance has been recognized as a threat to human health since the early 1940s, and the problem continues to grow and to evolve from one decade into the next [4]. In fact, there are more than 15 classes of antibiotics and none have escaped the development of a drug resistance mechanism. Not only are many pathogens drug

[4] S.B. Levy, B. Marshall. *Nat. Med.*, 10:S122-S129, 2004.

resistant, but they display resistance to multiple drugs. Multiple drug resistance was first observed among enteric bacteria, namely, *Escherichia coli*, *Shigella* and *Salmonella* in the late 1950s to early 1960s [5]. Since the 1980s, there has been a re-emergence of tuberculosis that is often multiple drug resistant [6]. Furthermore, the intensive use of antibiotics (which was estimated in 2002 to be 100,000-200,000 tonnes per annum worldwide [7] and which is, in total, well over 1 million tonnes since the 1940s [8]) has dramatically increased the frequency of resistance among human pathogens and threatens a loss of therapeutic options and a post-antibiotic era in which the medical advances to date are negated [9]. To successfully deal with drug resistance, we first need to understand all the underlying mechanisms, genetic and epigenetic, the latter of which is just beginning to come to light. This is the motivation behind the work presented in Chapters 5 and 6.

Chapter 5 presents a framework for modeling the effect of stress on the fitness of a noisy cell population as a first-passage problem. This study generalized previous work by incorporating the timescale of gene expression fluctuations. We demonstrated *in silico* that even relatively short-lived expression fluctuations can ensure the long-term survival of a drug resistant cell population. The range of values of the relaxation time parameter over which this survival occurred was in agreement with experimentally determined values in human lung cancer genes [10], which suggests that this novel nongenetic drug resistance mechanism could occur in natural systems. The goal is now to verify the model experimentally. However, there are challenges to investigating this in living cells. In particular, it is difficult to differentiate noise induced drug resistance

[5] S.B. Levy. *Clin. Infect. Dis.*, 33:S124-S129, 2001.

[6] B.R. Bloom, C.J.L. Murray. *Science*, 257:1055-1064, 1992.

[7] R. Wise. *J. Antimicrob. Chemother.*, 49:585-586, 2002.

[8] D.I. Andersson, D. Hughes. *Nat. Rev. Microbiol.*, 8:260-271, 2010.

[9] D.R. Guay. *Drugs*, 68:1169-1205, 2008.

[10] A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, *et al. Nature*, 444:643-646, 2008.

from a sense-response mechanism (in which a gene regulatory network responds by changing expression levels to a particular environmental stimulus). One possibility is to use single-cell imaging to distinguish between cells that survive a drug by activating a sense-response pathway, and highly expressing cells that have sufficient nongenetic memory and can stay above a critical protein threshold long enough to reproduce and contribute to the fitness of the cell population. Of course, it is also entirely possible that these two mechanisms act in combination. Future work for theoreticians includes deriving expressions for the probability and FPT distributions required for working analytically with our model. At present, these expressions are only available for the most simplistic cases.

Chapter 6 presents an investigation of the role that the topology of a gene regulatory network can play in the development of drug resistance. Here, we demonstrate that transcriptional regulation by feedforward and positive feedback loops can enhance drug resistance by increasing cell-to-cell variability in gene expression, and by enabling prolonged activation of gene expression in response to transient signals. These results further highlight how mechanisms enabling transient, nongenetic inheritance may play important roles in defining the effectiveness of drug treatment. Additionally, we performed the first mathematical modeling and simulation of the PDR5 transcriptional network, known to confer resistance to multiple drugs in *Saccharomyces cerevisiae*, despite decades of experimental research. The PDR5 gene is of particular significance because the overexpression of drug efflux pumps belonging to the same ABC transporter protein family in *Candida albicans* [11] and *Candida glabrata* [12], and cancers [13] are respectively known to confer drug resistance to antibiotics and chemotherapy. In agreement with what is known experimentally, up-

[11]K. Izumikawa, H. Kakeya, H.F. Tsai, B. Grimberg, *et al. Yeast*, 20:249-261, 2003.

[12]D. Sanglard, F.C. Odds. *Lancet Infect. Dis.*, 2:73-85, 2002.

[13]B.E. Bauer, K. Kuchler, H. Wolfger. *Biochimica et Biophysica Acta*, 1461:217-236, 1999.

regulation of PDR5 expression and drug resistance were observed in our simulations. The mathematical model should have other applications considering that the same network topology is also found in the human pathogenic fungi *Candida albicans* and *Candida glabrata*. I am presently collaborating with experimentalists in the laboratories of Dr. Mads Kærn and Dr. Gábor Bálazsi to use the modeling and simulation results to propose and direct experiments, and in turn obtain experimental data to refine and validate the models. This research will set the stage for studies in human cells, where the expression of ABC transporters is immensely complicated and currently not as amenable to the same level of quantitative analysis.

8.3 Final Thoughts

It has been known since as early as 1976 that bacterial cells grown in homogeneous conditions can display differences in phenotype that persist over their lifespans [14]. Our understanding of the origins and consequences of stochasticity in gene expression have advanced significantly since then. This advancement has been fueled by theoretical developments enabling biological hypothesis formulation using stochastic process and dynamical systems theory, as well as experimental breakthroughs in measurements of gene expression at the single cell level [15]. These advances have provided incontrovertible proof that there are important endogenous sources of stochasticity that drive biological processes [16].

Noise induced cellular heterogeneity has typically been ignored by molecular biologists [17]. In this thesis we have seen that noise can be more than just a nuisance and

[14] J.L. Spudich, D.E. Koshland. *Nature*, 262:467-471, 1976.

[15] M. Scott, B. Ingalls, M. Kærn. *Chaos*, 16:026107, 2006.

[16] M.S. Samoilov, G. Price, A.P. Arkin. *Sci. STKE*, 366:re17, 2006.

[17] S. Huang. *Development*, 136:3853-3862, 2009.

may be beneficial to cell populations under stress. There however is still much work to be done and many open questions: Could noise be used to fight disease, since after all, there is a fitness cost that a noisy diseased cell population must pay for the cells that deviate from the fitness optimum? What definition of fitness should we be using as reproductive fitness does not account for persister cells? Have we now come full circle since Darwin? That is, from heritable variation being required for evolution by natural selection, to an unwavering focus on mutations since the discovery of Watson and Crick, back to a more general definition of heritable variation that includes genetic and epigenetic mechanisms?

Stuart Kauffman told me when I started research in this field as an undergraduate seven years ago, that life exists at the boundary of order and chaos, that way it can function while still being able to adapt. The truth of this statement has never been more evident, as gene expression has been shown to be a stochastic, though far from a completely random process. This allows cells to meander around the fitness landscape, better positioning some of them for future environmental assaults, while others benefit the population by remaining close to the fitness optimum of the present environment.

Appendices

Appendix A: Enlarged Figures - Chapter 3

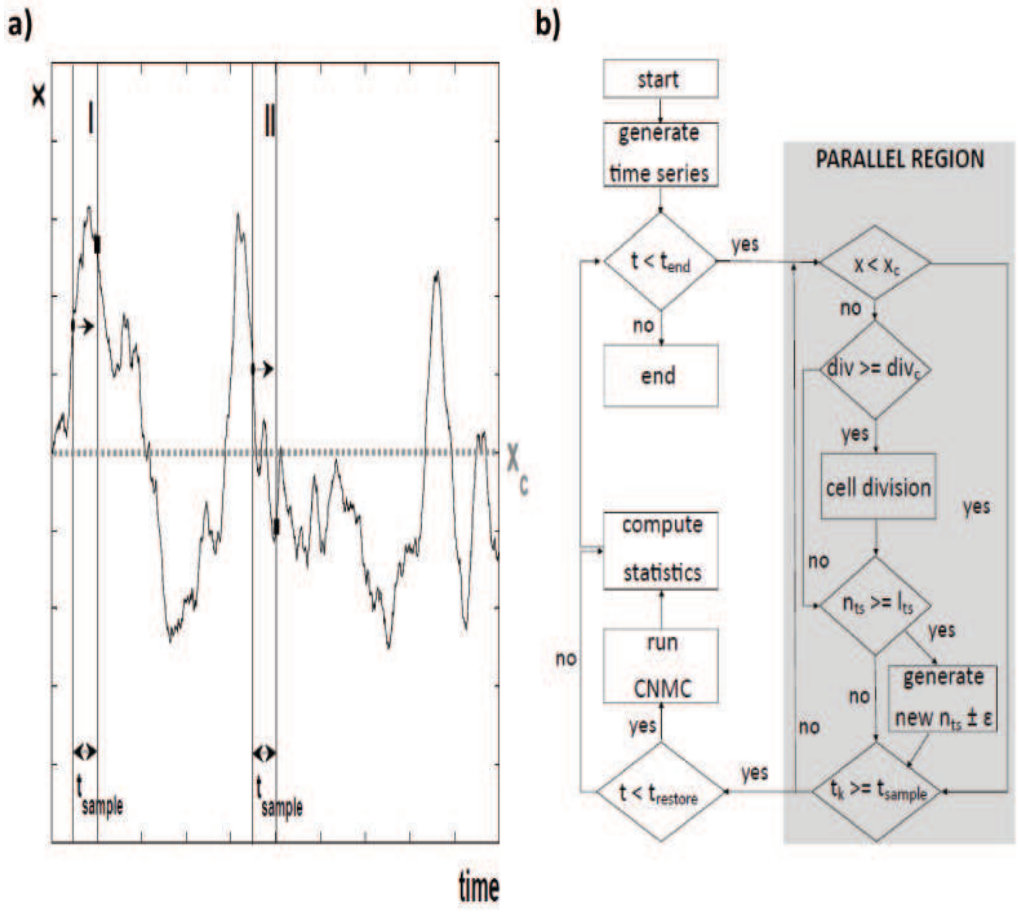


Figure 1 The accelerated method for simulating population dynamics (AMSPD) algorithm. (a) Schematic showing how individual cells are simulated by the AMPSD algorithm. After the cells are randomly assigned positions on the time series (black dot), their positions are incremented until the end of the sampling interval t_{sample} is reached (black square). If a reproductive stress is not incorporated into the simulations, then mother cells simply reproduce at a specified rate. However, if the fitness of the cells depends on the level of a particular biochemical variable, then cells can only reproduce if this variable remains above a specified threshold. For instance, in region I, the gene expression value x cell remains above a critical threshold x_c during the sampling interval and therefore is able to reproduce during the entire interval. In region II, the gene expression value of the cell falls below x_c and is therefore flagged and unable to reproduce after this point. (b) Flow diagram of the AMSPD algorithm presented in the main text for AMSPD variable and parameter descriptions).

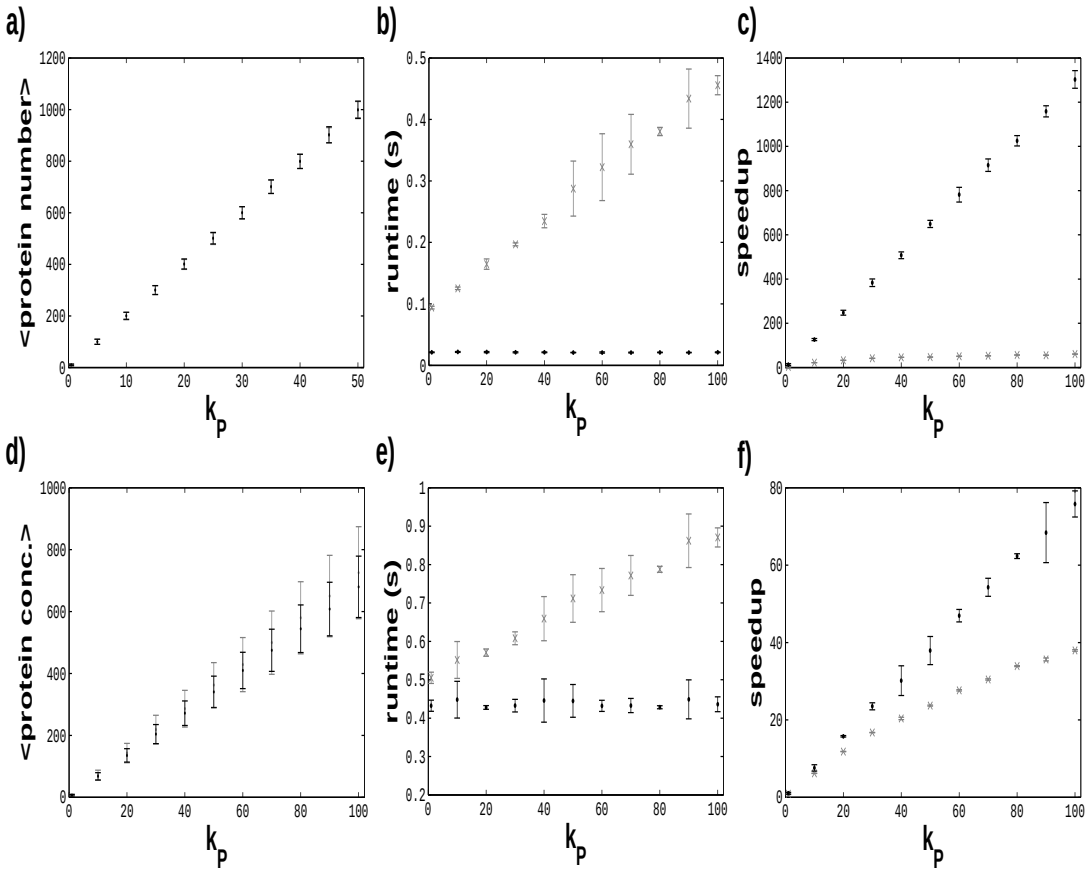


Figure 2 Comparison of accuracy and performance of the AMSPD algorithm and PDA (Charlebois *et al.*, Commun. Comput. Phys., 9, 1, 2011) for a birth-death model of gene expression. Panels (a) - (c) correspond to simulation results for volume independent cell division and (d) - (f) volume dependent cell division. (a) and (d) show the average steady-state protein numbers and concentrations, respectively, as a function of the rate of protein production k_P for the AMSPD algorithm (gray) and the PDA (black). (b) and (e) show the runtime of the AMSPD simulation. (c) and (f) show the speedup of the AMSPD algorithm, when compared to the runtime of the PDA, as a function of the rate of protein production k_P. Gray x's in (b) and (e), and in (c) and (f), are the results obtained when the time to produce the gene expression time series is incorporated into the AMSPD's runtime and the speedup calculation, respectively. Black dots in (b) and (e), and in (c) and (f), are the results obtained when the time to produce the gene expression time series is not incorporated into the AMSPD's runtime and the speedup calculation, respectively. Simulations were started from steady-state ($p^s = k_P/\delta_P$), the initial time since last division div drawn from a uniform distribution [0,div], and the protein time series generated by the AMPSD algorithm contained 10^4 values. The parameters were set to $\delta_P = 0.01$, $\epsilon = 10$, $div_c = 100$, and $t_{end} = 1000$.

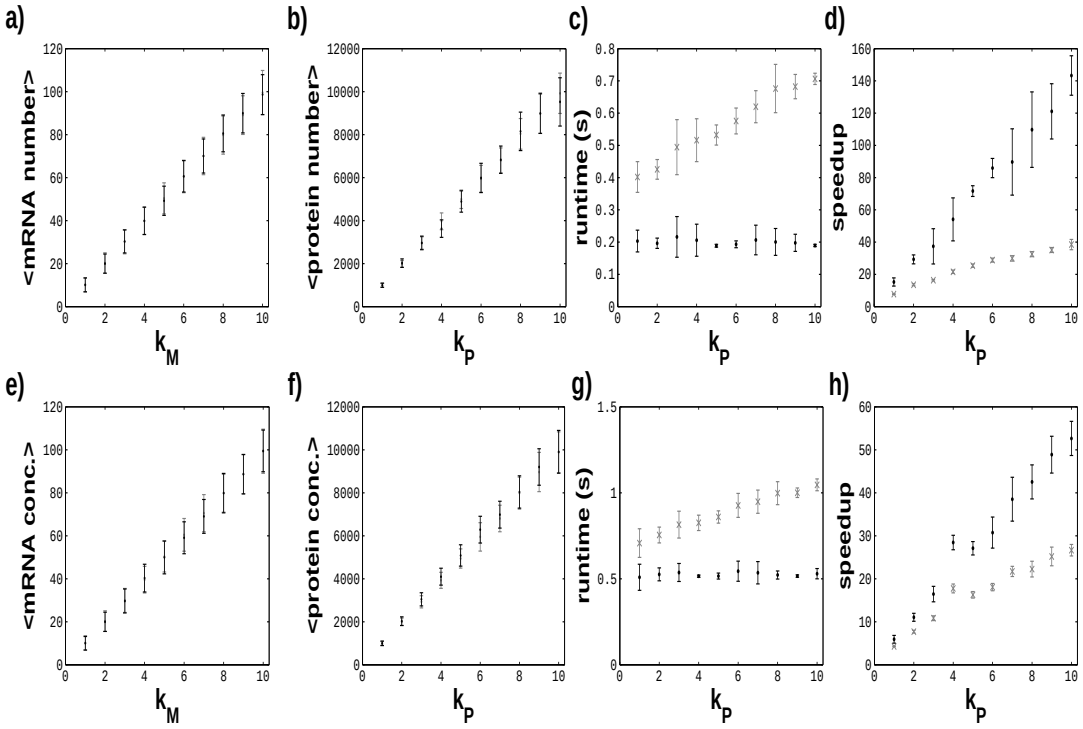


Figure 3 Comparison of accuracy and performance of the AMSPD algorithm and PDA (Charlebois *et al.*, Commun. Comput. Phys., 9, 1, 2011) for a two-step model of gene expression. Panels (a) - (d) correspond to simulation results for volume independent cell division and (e) - (h) volume dependent cell division. (a) and (e) show the steady-state mRNA numbers and concentrations, respectively, as a function of the rate of mRNA production k_M for the AMSPD algorithm (gray) and the PDA (black). (b) and (f) show the average steady-state protein numbers and concentrations, respectively, as a function of the rate of protein production k_P for the AMSPD algorithm (gray) and the PDA (black). (c) and (g) show the runtime of the AMSPD simulation. (d) and (h) show the speedup of the AMSPD algorithm, when compared to the runtime of the PDA, as a function of the rate of protein production k_P. Gray x's in (c) and (g), and in (d) and (h), are the results obtained when the time to produce the gene expression time series is incorporated into the AMSPD's runtime and the speedup calculation, respectively. Black dots in (c) and (g), and in (d) and (h), are the results obtained when the time to produce the gene expression time series is not incorporated into the AMSPD's runtime and the speedup calculation, respectively. Simulations were started from steady-state ($M^s = k_M/\delta_M$ and $P^s = k_M k_P/\delta_M \delta_P$), the initial time since last division div drawn from a uniform distribution [0,div], and the protein time series generated by the AMPSD algorithm contained 10^4 values. The parameters were set to $k_M = 1$ (when k_P was varied), $k_P = 1$ (when k_M was varied), $\delta_M = 0.1$, $\delta_P = 0.01$, $\epsilon = 10$, $div_c = 100$, and $t_{end} = 1000$.

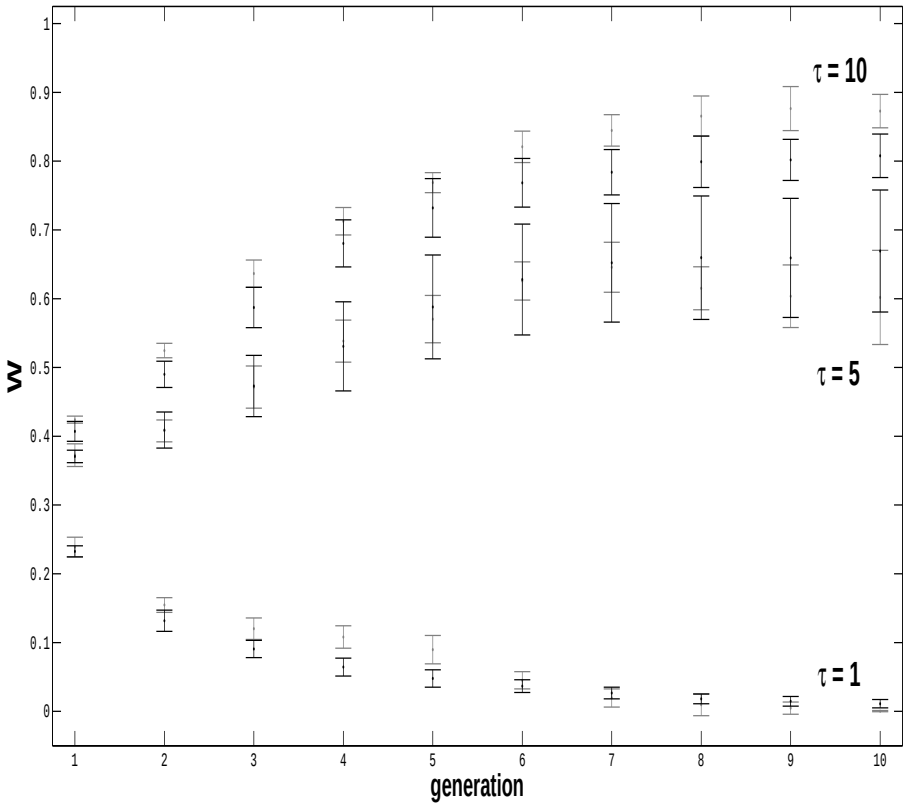


Figure 4 Comparison of accuracy of the AMSPD algorithm (gray) and PDA (black) (Charlebois *et al.*, Commun. Comput. Phys., 9, 1, 2011) for a model capturing the effect of non-genetic memory τ on drug resistance at various timescales (Charlebois *et al.*, Phys. Rev. Lett., 107, 218101, 2011). The reproductive fitness of the cell population (macrofitness) W as a function of generation is plotted for various values of τ. Simulations were started from the steady-state Ornstein-Uhlenbeck distribution (with mean $\mu = 0$ and variance $\sigma^2 = c\tau/2 = 1$), the initial time since last division div was drawn from a uniform distribution [0,div], and the protein time series generated by the AMPSD algorithm contained 10^6 values. The parameters were set to $\epsilon = 1$ and $div_c = 1$, and scaled by div_c. The threshold below which cells were unable to reproduce x_c was set to μ.

Appendix B: Supplementary Figure - Chapter 4

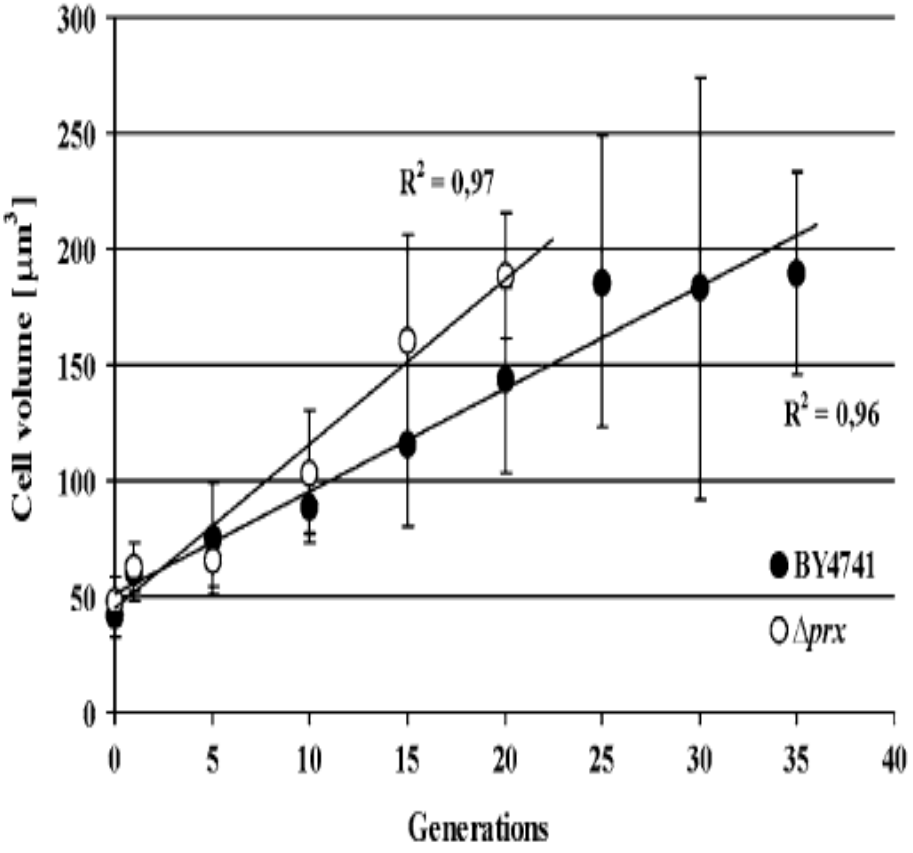


Figure 5 Experimental data demonstrating the dependence of cell volume on the number of cell divisions accomplished by mother yeast cells for a "wild-type" yeast strain (BY4741) and a mutant yeast strain (Δprx) deficient in an antioxidant defense protein. Error bars indicate standard deviation. Figure used by permission from Springer: Biogerontology 10: 481-488, copyright 2009.

Appendix C: Enlarged Figures - Chapter 6

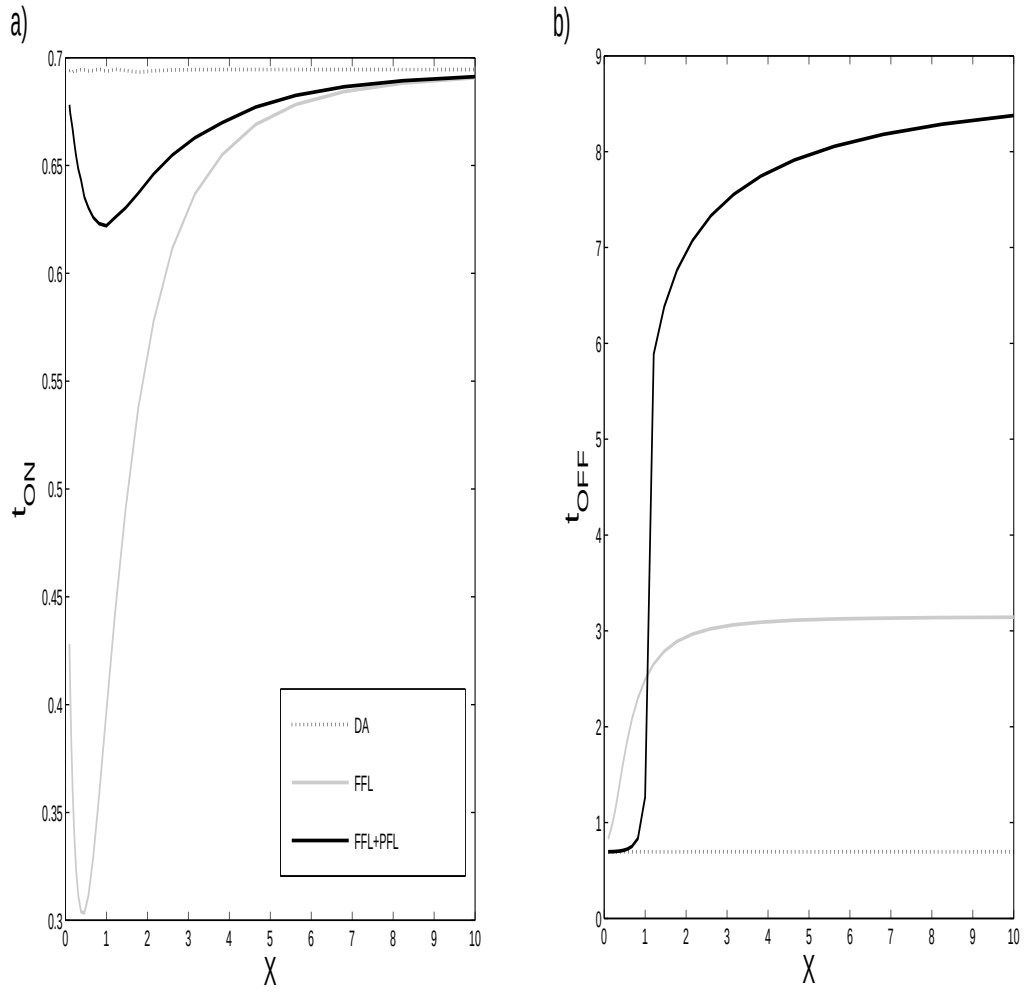


Figure 6 Coherent feedforward networks enable fast and prolonged activation. (a) Response time t_{ON} (time for z to reach 50% of steady-state level) for the DA, FFL, and FFL+PFL networks as a function of an activating signal x. (b) Relaxation time t_{OFF} (time for z to fall to 50% of steady-state level) for the same network motifs considered in (a) as a function of x. Parameters were set to: $\alpha = 10$, $n = 2$, and $K = 1$.

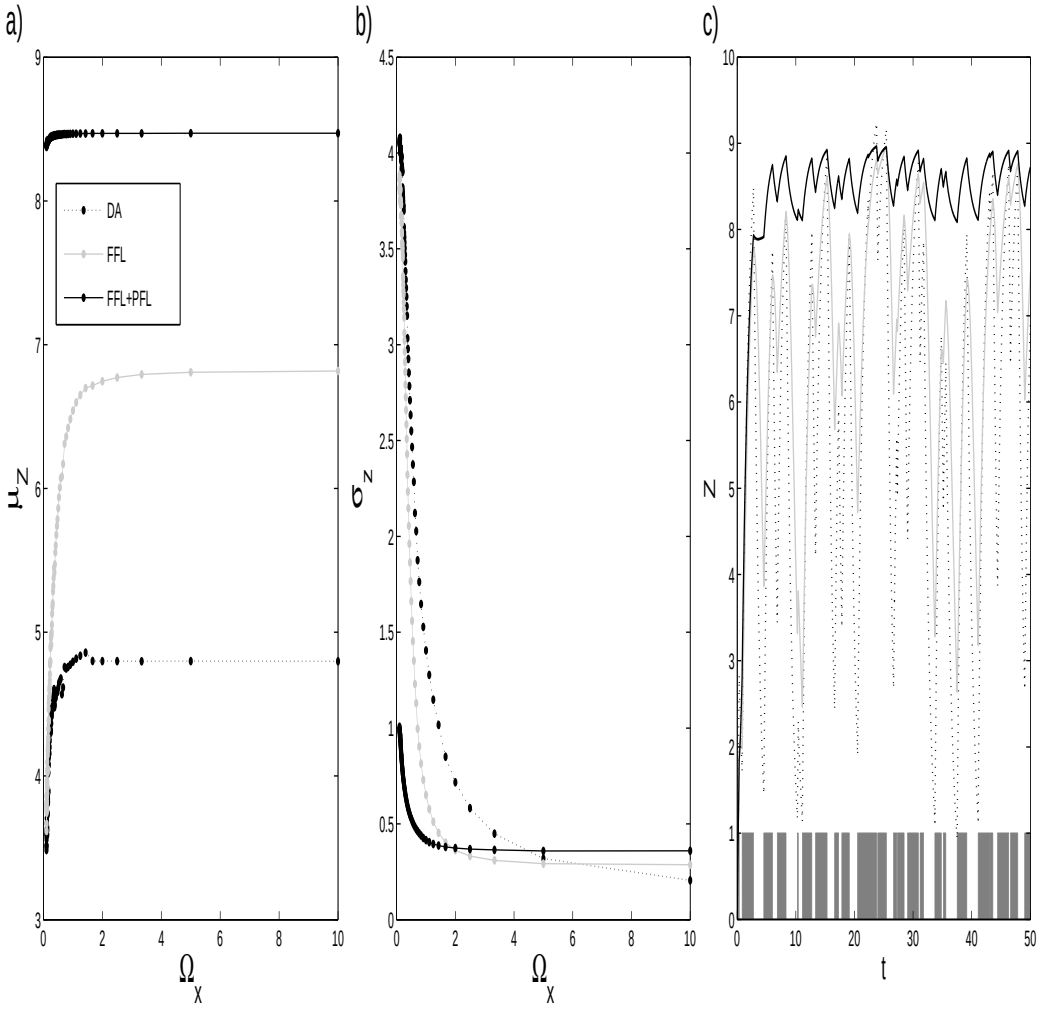


Figure 7 Response of z to a fluctuating upstream activating signal. (a) The mean of z (μ_z) is shown for the DA, FFL, and FFL+PFL networks as a function of the ON-OFF switching frequency of x (Ω_x). (b) The standard deviation of z (σ_z) is shown for the same network motifs considered in (a) as a function of Ω_x. (c) z as a function of a randomly fluctuating Ω_x (dark gray rectangles denote presence of an activating signal) for the same network motifs considered in (a). Parameters were set to: $x_{ON} = 5$, $x_{OFF} = 0$, $\alpha = 10$, $n = 2$, $K = 1$.

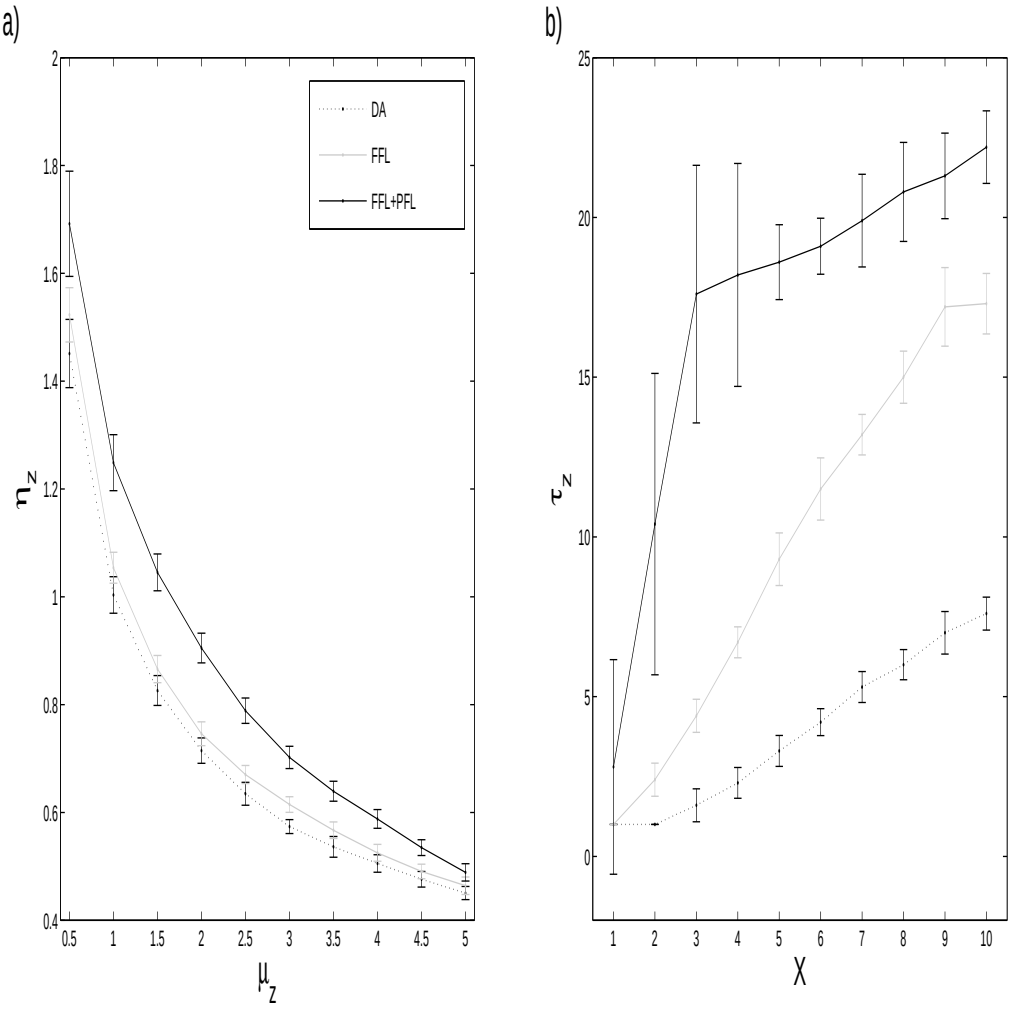


Figure 8 Coherent feedforward networks increase noise and relaxation time. (a) The noise in z (η_z) for the DA, FFL, and FFL+PFL networks as a function of the mean level of z (μ_z). (b) The relaxation times for z (τ_z) for the same network motifs considered in (a) as a function of an activating signal x. Parameters were set to: $\alpha = 10$, $n = 2$, $K = 10$. 10 realization for 10^3 arbitrary time units were performed. Error bars show standard deviation.

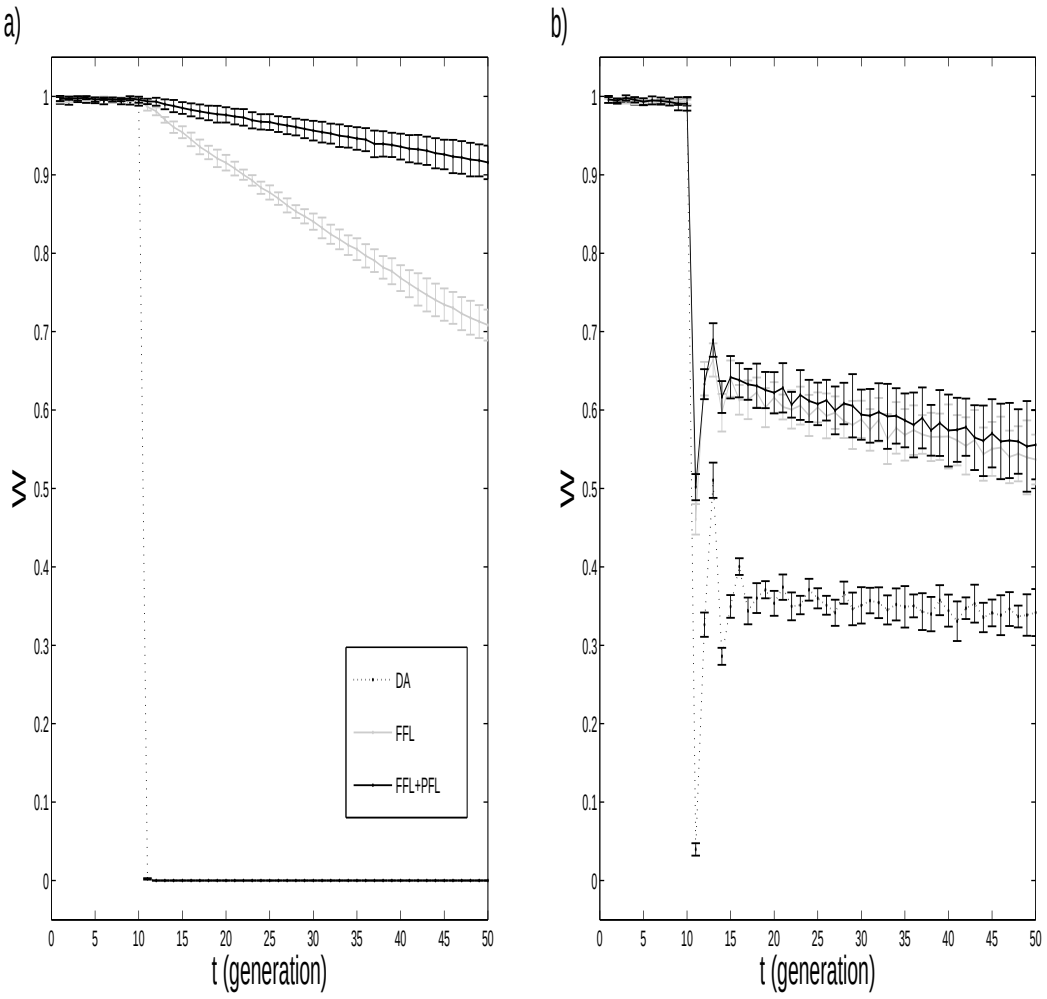


Figure 9 Effect of network topology and fitness threshold on reproductive fitness (W). W for each of the DA, FFL, and FFL+PFL populations over 50 generations. Drug treatment is initiated at the tenth generation. (a) Population simulations using a step fitness function. (b) Population simulations performed using a Hill type fitness function. Parameters were set to: $x = 1$, $\alpha_y = 10$, $\alpha_z = 100$, $K = 1$, and $n = 2$. In (a) $z_c = 35$ and in (b) $K_w = 35$, and $n_w = 2$. 10 realizations of 1000 cells were performed. Error bars show standard deviation.

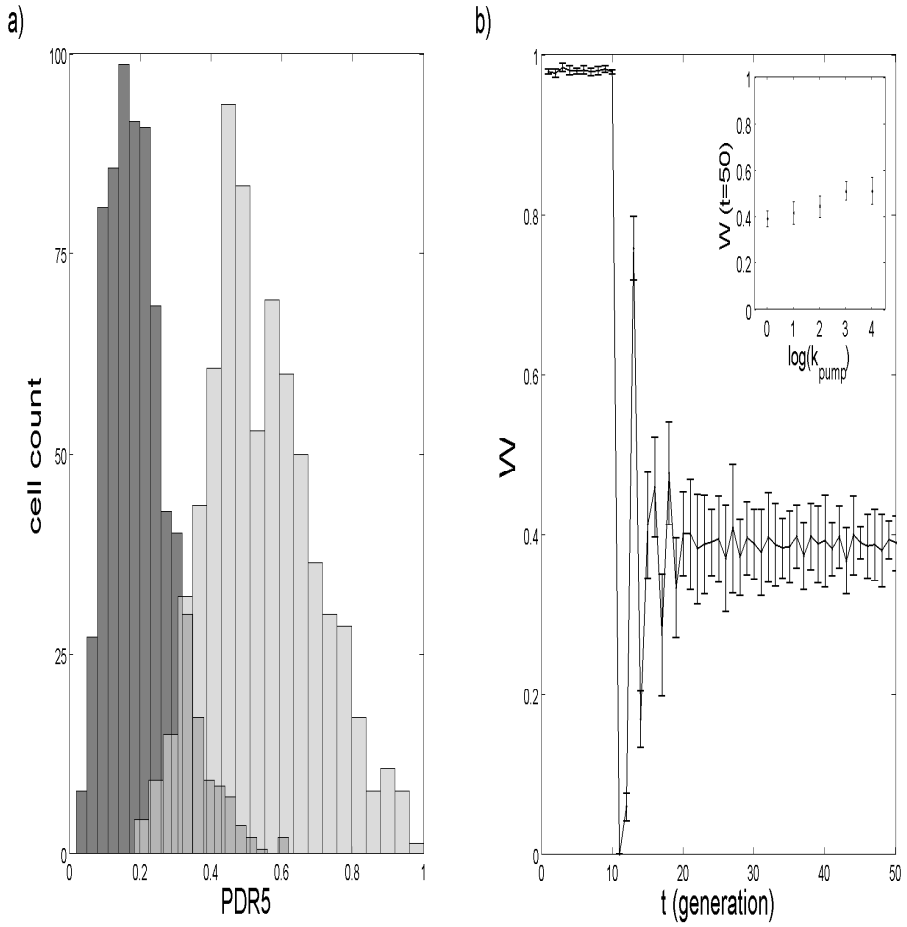


Figure 10 Adaptation and fitness (W) facilitated by the PDR5 transcriptional network. (a) Number of cells in the population has the corresponding PDR5 expression level. Distributions are shown for the generation prior (dark gray) to drug application and 40 generations after (light gray) drug application. (b) W over 50 generations. Inset shows W at generation 50 for different rates of PDR5 mediated drug efflux (k_{pump}). Drug treatment is initiated at the tenth generation. Unless otherwise indicated, parameters were set to: $\alpha_0 = 1$, $\alpha_1 = 10$, $K_1 = 1$, $n_1 = 1$, $\alpha_3 = 10$, $K_3 = 1$, $n_3 = 2$, $\alpha_5 = 100$, $K_5 = 20$, $n_5 = 2$, $k_{diff} = 100$, $k_{pump} = 1$, and $drug_{ext} = 100$. 10 realizations of 1000 cells were performed.