QUANTITATIVE ANALYSIS OF GENE REGULATORY NETWORKS: FROM SINGLE CELLS TO CELL COMMUNITIES

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

KONSTANTINOS BILIOURIS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

YIANNIS KAZNESSIS, ADVISOR
PRODROMOS DAOUTIDIS, CO-ADVISOR

June, 2013

Acknowledgements

One of the amusements of fulfillment is to look over the past and remember all the people who contributed to achieving your goal.

First and foremost, I would like to express my utmost gratitude to my Advisor, Yiannis Kaznessis, for his guidance, patience, enthusiasm and inspiration throughout my five years at the University of Minnesota. Yiannis was always there when I needed him and discussing with me not only scientific topics but also personal concerns; his contribution towards my personal and professional growth was instrumental and is immensely acknowledged. It goes without saying that the tremendous scientific and moral support of my Co-Advisor, Prodromos Daoutidis, also played a catalytic role in my studies and is sincerely appreciated. Prodromos has been remarkably supportive since I met him in 2007, when I decided to move to the US for graduate studies, and without his precious pieces of advice my career path would have been totally different. In retrospect, I feel that the experience of having two Greek Advisors is unparalleled and something I will never regret.

I would like to express the deepest appreciation to my family, as without their permanent support I would have never been able to complete my studies. Of course, all my friends also deserve my sincere appreciation as they do constitute my family in the US. In particular, I would like to express special thanks to Panos, George, Yiannis, Charalambos, Mary, Thanasis, Nikos, Dennis, Dimitris, Christos, Sophia and Vassilis (x3) for the fun time we had in Minnesota and other States. The joy of spending time with these people will truly be unforgettable. I am also indebted to Vassilis and Venia for their extraordinary hospitality and friendship. I would also like to acknowledge the big support I had from Dimitris and Vassilis during my arrival at Minnesota and the first years of my studies. I am particularly thankful to Amy who played a fundamental

role in my progress and life at the University of Minnesota. Amy was always supportive and sympathetic during the ups and downs of my studies and she deserves my profound appreciation.

Of high importance was also the support of my past and current group mates. I cannot find words to express my gratitude to Kat who has been supporting me unconditionally over the last five years. She is a great group mate and friend who has been through the good and bad times of my PhD life; my studies would have been lonely without her. Last but not least, many thanks go to Vassilis, Patrick, and Ben for the fun time and long scientific conversations we had over the last years.

My acknowledgements would not be complete without expressing my sincere appreciation to the members of my final exam committee: Prof. Ben Hackel, Prof. Claudia Schmidt-Dannert and Prof. David Odde.

Dedication

This thesis is dedicated to my parents, Anna and Michail, and my sister, Elisavet, for their endless support, encouragement, and patience.

Abstract

Although the great advances in experimental biology have fueled our ability to explore the behavior of natural and synthetic biological systems, key challenges still exist. A major shortcoming is that, unlike other research areas, biological systems are significantly non-linear with unknown molecular components. In addition, the inherent stochasticity of biological systems forces identical cells to behave dissimilarly even when exposed to the same environmental conditions. These challenges limit in-depth understanding of biological systems using solely experimental techniques.

The current research is focused on the joint frontier of mathematical modeling and experimental work in biology. Guided by experimental observations, quantitative modeling analysis of two natural and two synthetic biological systems was carried out. These systems are all gene regulatory networks and range from the single cell level to the population level. The objective of this research is three-fold: 1) The development of detailed mathematical models that capture the relevant biomolecular interactions of the systems of interest. Experimental data are used to inform and validate these models. 2) The use of the models as a means for understanding the complexity underlying biological systems. This allows for explaining the behavior of biological systems by quantifying the molecular interactions involved. 3) The simulation of the behavior of biological systems and the associated molecular parts. This helps to quickly and inexpensively predict the behavior of these systems under various conditions and motivates new sets of experiments.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Recent technological advances allow for the "reading and writing" of DNA remarkably fast and inexpensively. Since 2000, DNA genome sequencing has been accelerated by 50,000 times providing new understandings of the components and functionalities underlying biological systems. In addition, the chemical synthesis of DNA has become almost infallible while its cost has been slashed to less than $0.5 per base pair [1]. This has enabled scientists to cut and paste the DNA of different organisms removing, reinserting, and tweaking genes of interest rendering their function human-controlled. This tremendous technological achievement has culminated in a new scientific field, synthetic biology, which constitutes a discipline between engineering and biology. Synthetic biologists have been redesigning naturally occurring biological systems and constructing new synthetic biological entities primarily aiming to fully grasp their complexity [2, 3]. A desired function can be then imparted to these synthetically produced biological systems so that they meet specific performance criteria.

New understandings of gene and genome function enabled the construction of synthetic systems at the single-cell resolution with remarkable functions that range from oscillatory or bistable protein production [4, 5, 6, 7], to tight regulation of gene expression [8, 9] and logical gate-like behavior [10, 11, 12, 13]. Multicellular synthetic systems with intriguing functions that are chiefly based on cell-to-cell communication have been also developed [14, 15, 16, 17, 18]. Notable functions involve the synchronization, rescuing, and killing of cells within a population [19, 20, 21].

The application of synthetic biological systems has found use in various fields spanning from energy to biomedicine and biotechnology. Today, synthetic biology technologies facilitate the inexpensive production of fuels, drugs, and various bio-materials [22, 23, 24]. Moreover, synthetic biology has contributed towards detecting and healing infections, advancing cancer treatments, designing sophisticated vaccines, and improving cell therapy and regenerative medicine [25, 26].

Although the great advances in experimental biology have fueled our ability to explore the behavior of natural and synthetic biological systems, key challenges still exist [27]. For example, how to optimally construct and tune synthetic systems still remains enigmatic. The major shortcoming is that, unlike other areas, biological systems are significantly non-linear with unknown molecular components. In addition, the inherent stochasticity of biological systems forces identical cells to behave dissimilarly even when exposed to the same environmental conditions [28]. As such, an in-depth systems-level understanding of (synthetic) biological systems using pure experimental techniques is challenging. To overcome this hurdle, tools from mathematics are fruitfully applied in biology [29, 30, 31, 32]. In fact, M. Laubichler notes that *"In twenty-first-century biology, modeling has a similar role as the microscope had in earlier centuries; it is arguably the most important research tool for studying complex phenomena and processes in all areas of the life sciences, from molecular biology to ecosystems analysis"* [33]. It therefore comes as no surprise that the Global Agenda Council on Emerging Technologies listed mathematical modeling of biological systems in the top ten emerging technologies that were expected to have major social, economic and environmental impacts worldwide in 2012.

The current dissertation is focused on the joint frontier of mathematical modeling and experimental work in the field of biology. Quantitative analysis, guided by experimental observations from our lab or previously published studies, of two natural and two synthetic biological systems was carried out. These systems are gene regulatory networks that regulate the expression of genes of interest and range from the single-cell level to the population level. The objective of this research is three-fold: 1) The development of detailed mathematical models capturing the biomolecular interactions engaged in the biological systems of interest. Experimental data are used to inform and validate these models. 2) The use of the models as a means for understanding the

complexity underlying the biological systems. This allows for explaining the behavior of biological systems by quantifying the molecular interactions involved and fully characterizing them. 3) The simulation of the behavior of biological systems and the associated molecular parts. This helps to quickly and inexpensively predict the behavior of these systems under various conditions and drive new sets of experiments.

The ensuing analysis is organized as follows:

- Chapter 1 discusses the stochasticity underlying biological systems and introduces the approaches used to model and simulate their behavior.

- Chapter 2 presents the experimental and computational analysis of sugar consumption pathways in the model bacterium *Escherichia coli* (*E. coli*). Bacteria can subsist on diverse nutrients present in the environment. Many of these nutrients induce dedicated utilization pathways responsible for nutrient uptake and catabolism. Among the myriad of utilization pathways in bacteria, only two have been studied at the single-cell level. In the absence of nutrient catabolism these pathways have been shown to exhibit an "all-or-none" response. In the presence of catabolism, these pathways generate diverse responses. Although mathematical models have predicted that natural pathways should generally exhibit graded responses, it remains to be established how the plethora of utilization pathways in bacteria respond to nutrients at the single-cell level. In this chapter, we pair mathematical modeling with experimentation to argue that various utilization pathways in *E. coli* exhibit a bimodal, and not a graded, response that can be described as attenuated bistability. In-depth characterization of the L-arabinose utilization pathway revealed that the latter elicits bimodal responses with reduced hysteresis. Our findings demonstrate that nutrient utilization pathways in individual bacterial cells naturally exhibit divergent responses with limited memory of previous conditions, suggesting complex, single-cell interactions between microorganisms and their environments.

- In Chapter 3 the dynamic behavior of the tetracycline operon is explored. The tetracycline operon is a self-regulated system and occuring naturally in bacteria where it confers resistance to antibiotic tetracycline [34]. Because of the performance of the molecular elements of the tetracycline operon, these elements are

widely used as parts of synthetic gene networks, where the protein production can be efficiently turned on (off) in response to the presence (absence) of tetracycline. In this chapter, we investigate the dynamics of the tetracycline operon through a mathematical model that was informed by experimental observations. Computer simulations of the tetracycline operon afford augmented insight into the interplay between its molecular components and provide useful explanations of how the components and their interactions have evolved to best serve bacteria carrying this operon. Such simulations may assist in designing novel gene network architectures consisting of tetracycline operon components.

- Unlike the previous two chapters that interrogate the behavior of natural systems, Chapter 4 revolves around two synthetic genetic switches that were designed, modeled, built, and characterized in our lab [8]. These switches turn on and off the expression of a protein in a robust and sophisticated fashion. As these biological switches are artificial, we have optimized them to tightly regulate protein production in a user-defined manner. These biological switches switch on the production of a target protein in bacteria in response to the presence of anhydrotetracycline. These synthetic systems were experimentally built in bacteria as well as computationally modeled. The simulation results were used to quantify the strength of the biomolecular interactions governing the behavior of these switches. These new, well characterized artificial devices are becoming useful additions to the molecular toolboxes used by biologists to control bacterial gene expression.

- In contrast to the previous three chapters that are focused on biological systems at the single-cell resolution, Chapter 5 focuses on the design of a synthetic cell community [35]. It has been demonstrated that for cells to achieve a complex task, communication between them may be required. To date, all the synthetic cell communities reported in the literature refer to communication between cells of the same species, for example bacteria-bacteria or yeast-yeast. However, for cell communities to have a broader range of applications, heterogeneous cell populations may be required. Therefore, guided by previously experimentally developed systems, we designed *in silico* a synthetic heterogeneous ecosystem composed of bacteria and yeast. This system was found to exhibit interesting dynamics that

are common in natural ecosystems, such as obligatory and facultative mutualism, extinction, commensalism and predator-prey like dynamics. A potential industrial application of this system could be in the area of bio-fuels, where the use of cell cultures would require both yeast and bacteria to completely break down the raw material (cellulosic biomass) for maximum bio-fuel production efficiency.

- Chapter 6 provides a brief summary of the conclusions and main contributions of the present dissertation and proposes potential avenues of future research.

Chapter 2

Background information on synthetic biology and computational modeling

2.1 Synthetic biology

Inspired by naturally occurring mechanisms that orchestrate cell behavior, scientists develop non-natural cell functions, behaviors, and their controls for both research and industrial applications [22, 23, 24]. To date, numerous non-natural, or synthetic, systems have been designed and built experimentally. The function of these systems exploits the feedback mechanisms underlying biological systems and relies on coupling an extracellular signal (or input) to a unique intracellular target (or output) [36]. Significant efforts have been focused on building systems to study or address previously unmet needs in gene expression regulation at the level of transcription [8, 12], mRNA stability [37], translation [38], or protein stability [39, 40]. While all four of these approaches are very powerful, the following discussion concentrates on the development of synthetic systems that regulate gene transcription.

Two pioneer systems constructed in *E. coli* to achieve novel gene expression control include the toggle bistable switch and the repressillator [7, 4]. The former constitutes a gene regulatory network composed of two mutually repressive promoters and the latter is

a gene network that consists of three genes arranged in a way that each gene represses the expression of another: the first gene represses the expression of the second, the second gene represses the expression of the third gene, and the third represses back the expression of the first. The bistable switch allows for switching between producing and not producing a protein while the repressilator yields the production of a protein in an oscillatory manner. The development of these two systems exemplified a novel genetic engineering approach and paved the way for an emerging scientific field, synthetic biology.

Another class of synthetic biological systems includes sophisticated "genetic switches" which turn on and off the expression of a target gene in a user-defined fashion. Two notable examples include the TetOn and TetOff system [41, 42, 43, 44]. These systems were designed in eukaryotes to tightly control the expression of a target gene in a tetracycline-dependent fashion. Because of their unprecedented function, TetOn and TetOff have been widely utilized in applications where gene expression control is required. Similar switches have been built in prokaryotes and dubbed prokaryotic-TetOn (proTeOn) and prokaryotic-TetOff (proTeOff) [8]. An extensive description of the function of proteOn and proTeOff is presented in Chapter 5. A gamut of other synthetic biological switches and oscillators have been developed and characterized [45, 46, 6, 47].

The design and characterization of the aforementioned synthetic systems pushed the development of more complicated systems with multiple layers of regulation. A family of such systems involves biological constructs that behave as logic-gates. The design of these systems is based on computer science principles, as they implement a digital-like Boolean function by coupling an output to one or more inputs. In the vast majority of these constructs, the output corresponds to gene expression, which can be on (1) or off (0), and the input refers to the presence (1) or absence (0) of a particular molecule. For example, in a synthetic system behaving like an AND gate, gene expression switches on if and only if two particular molecules are present. In a similar vein, systems behaving like an OR, NOT, NAND, NOR, and other logic-gates have been built experimentally [12, 13, 10, 11, 48, 49, 50]. With the development of this type of systems, the field of synthetic biology took one step forward.

Using as a foundation the synthetic constructs designed in single-cells, synthetic

biologists have recently built synthetic systems whose function leans on cell-to-cell communication. The communication among cells within a population is primarily achieved by quorum sensing [17, 51]. The use of multicellular systems in different applications may be beneficial as, unlike the monocultures, multicultures allow the different species to share the different steps of a particular task. This function sharing not only mediates the burden in the metabolism of the cells but also may meliorate the fine-tuning of the artificial function [17]. The function of multi-cellular synthetic systems are various and span from the formation of different patterns of differentiation, such as rings and clovers [52], to the regulation of cell metabolism by neighbor cells [53]. A more detailed discussion of multi-cellular systems is provided in Chapter 6.

2.2 Computational modeling of gene regulatory networks

According to Monod's postulate, the biological complexity of biological systems can be reduced to networks of biomolecular interactions [54]. These biomolecular interactions are diverse and include but are not limited to protein:protein (un)binding, protein:DNA (un)binding, Ribosome:mRNA (un)binding. These interactions are traditionally captured by biochemical reactions and give rise to interesting cellular dynamics that are described by mass action kinetics (ordinary differential equations). For example, the dynamics of a protein A found in a cell and participating in a dimerization process can be captured by reaction 2.1 and its evolution can be described by equation 2.2:

$$2A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A_2 \tag{2.1}$$

$$\frac{d[A]}{dt} = -2 \cdot k_1 \cdot [A]^2 + 2 \cdot k_2 \cdot [A_2] \tag{2.2}$$

where [A] is the concentration of A. The deterministic approach in equation 2.2 provides only the average concentration of A. In other words, it neglects the dissimilarities, if any, in the behavior of individual A molecules and solely calculates the bulk behavior. This approach is accurate when A is at the thermodynamic limit, i.e. the number of molecules of A is significantly high. However, when the amount of A is low, the behavior of individual A molecules, or the fluctuations around the mean behavior, become

important and cannot be neglected. The significance of fluctuations can be visualized through the following toy example:

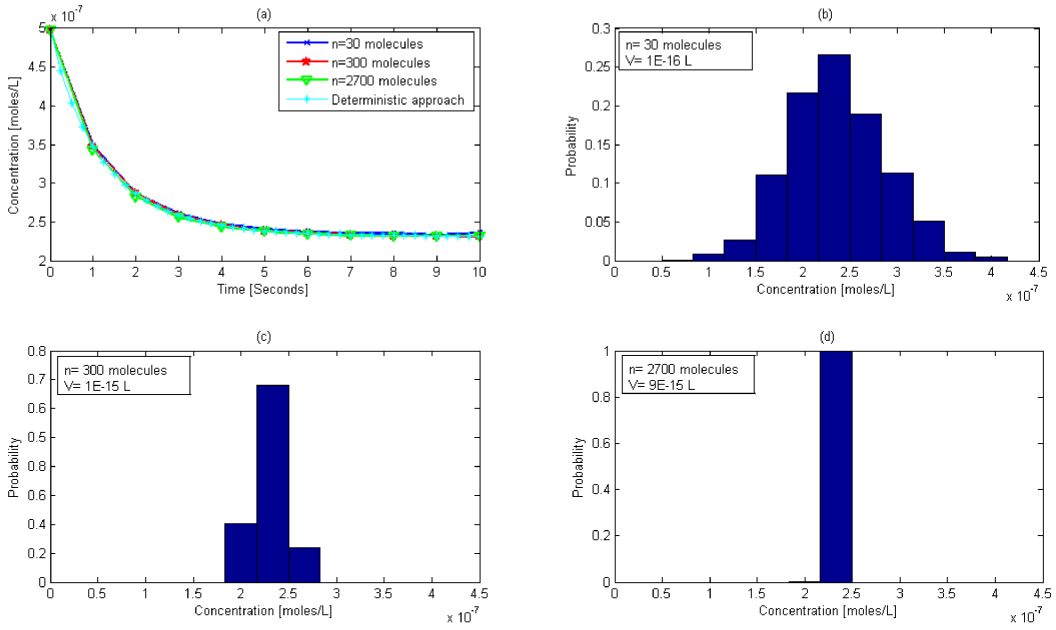

Figure 2.1: (a): Average concentration of A over 1,000 trials calculated using a stochastic approach and concentration of A calculated using a deterministic approach. (b),(c),(d): Probability distribution of the concentration of A at steady state for different number of molecules and volumes calculated using a stochastic approach.

The dynamics of A in the reversible reaction 2.1 was simulated implementing a deterministic (equation 2.2) and a stochastic approach [55](parameters adopted from [56]). Figure 2.1(a) shows how the concentration of A evolves according to the deterministic approach. It also shows the average (over 1,000 trials) concentration of A, for three different cases, calculated using a stochastic approach. In the latter cases, both the initial number of A molecules and the system's volume were modified so that the concentration of A was kept constant. As observed, the average values of the stochastic model match the solution of the deterministic model for all the three cases. However, in this case the distribution of the concentration of A is normal. In more complex systems where the underlying probability distribution is not normal (e.g. bistable systems), the average value of the stochastic approach could differ significantly from the deterministic solution [57].

Figures 2.1(b), 2.1(c) and 2.1(d) depict the probability distribution of the concentration of A at steady state (10 seconds) as calculated by the stochastic model. The difference between these three plots is the number of molecules and the volume of the system. These plots demonstrate how the fluctuations around the mean value decrease as the number of molecules increases. When the number of molecules of A is small (30), the solution fluctuates significantly around the mean value and thus the use of a deterministic approach would be invalid, as it would only provide the average value. According to the stochastic model, there is only a probability approximately equal to 0.27 that the result provided by the deterministic approach is accurate. When the number of molecules of A increases (300), the fluctuations are decreased dramatically, but even in this case the behavior of the system cannot be described accurately by the deterministic approach. Interestingly, only when the number of molecules of A is relatively high (2,700), and therefore close to the thermodynamic limit, fluctuations vanish and the result from the deterministic approach matches the result from the stochastic approach. It should be kept in mind that these numbers strongly depend on the number of bins of the histogram and therefore they only refer to this particular bin number.

As shown in this example, the pronounced effects of fluctuations in systems far from the thermodynamic limit cannot be ignored as they render the behavior of these systems stochastic. It has been extensively demonstrated that stochasticity is ubiquitous in biological systems as they are composed of constituents, such as promoters and operators, that participate in the biological processes in low copy numbers [58, 28, 59, 60, 61]. This stochasticity necessitates the use of stochastic approaches which, unlike the deterministic approaches, take into consideration the inherent stochasticity of biological systems and account for the different behaviors exhibited by individual molecules within the same population. The following section discusses current stochastic approaches that are used to simulate the behavior of biological systems.

2.2.1 Chemical Master Equation

A sophisticated way to account for the stochasticity underlying a system, or each possible "state" (number of molecules of each species) that a system can be found at, is to treat its dynamics as a Markov chain. A Markov chain is a stochastic process which satisfies the Markov property: the future state of the system depends exclusively on its

current state and not on its entire history of states. The ensuing discussion focuses on how such a system can be treated as a Markov chain [62].

In a well-mixed biological system (e.g. a cell) with N species the state vector corresponds to the number of molecules of each species at time t, $\underline{X}(t) = X_1(t), ..., X_N(t)$. Assuming that these species interact through M biochemical reactions, there is an $M \times N$ matrix, $\underline{\underline{\nu}}$, that incorporates the stoichiometric coefficients of these species. In this respect, ν_{ij} represents the change in the number of molecules of i^{th} species due to the occurrence of the j^{th} reaction. In contrast to the deterministic reaction rate, the reaction rate of each biochemical reaction, $\underline{a}_j(\underline{X}(t))$, is now probabilistic and depends on the volume of the system (see Table 2.1). As such, $\underline{\alpha}_j(\underline{X}(t))dt$ is the probability that the j^{th} reaction will occur in the time interval $[t, dt]$. Starting from a particular state, $\underline{X}_o(t)$, at time t_o, the probability for the system to be at the state $\underline{X}(t)$ at time $t_o + dt$ is calculated according to equation 2.3 [63]:

$$P\left(\underline{X}; t_o + dt\right) = P\left(\underline{X}; t_o\right) \cdot \left[1 - \sum_{j=1}^{M} \alpha_j\left(\underline{X}\right) dt\right] + \\ + \sum_{j=1}^{M} \left[P\left(\underline{X} - \underline{\nu}_j; t\right) \cdot \alpha_j\left(\underline{X} - \underline{\nu}_j\right) dt\right] \quad (2.3)$$

Table 2.1: Propensities of common reaction types. V represents the volume of the system and N_A is the Avogadro Number.

Reaction	Propensity [molecules/sec]
$\emptyset \xrightarrow{k} A$	$a = kVN_A$
$A \xrightarrow{k} B$	$a = kA$
$A + B \xrightarrow{k} C$	$a = \left(\frac{k}{VN_A}\right) AB$
$A + B + C \xrightarrow{k} D$	$a = \left(\frac{k}{(VN_A)^2}\right) ABC$
$2A \xrightarrow{k} B$	$a = \left(\frac{2k}{VN_A}\right) A(A-1)/2$
$3A \xrightarrow{k} B$	$a = \left(\frac{6k}{(VN_A)^2}\right) A(A-1)(A-2)/6$

Equation 2.3 is called Chemical Master Equation (CME) [63]. The probability for

a system to be at state $\underline{X}(t)$ at time $t_o + dt$ is equivalent to the following sum of probabilities: 1) the probability that the system is already found at state $\underline{X}(t)$ at time t_o and no reaction occurs during $[t_o + dt]$, 2) the probability that the system is found at state $\underline{X} - \underline{\nu}_j$ at time t_o and one exactly reaction occurs at $[t_o + dt]$ which brings the system to the state $\underline{X}(t)$. This definition assumes that dt is particularly small so that only one reaction can take place during this time window. Upon straightforward algebraic rearrangements and taking the limit $dt \to 0$, equation 2.3 yields:

$$\frac{\partial P\left(\underline{X};t\right)}{\partial t} = \sum_{j=1}^{M} \Big[\alpha_j \left(\underline{X} - \underline{\nu}_j\right) P\left(\underline{X} - \underline{\nu}_j; t\right) - \alpha_j \left(\underline{X}\right) P\left(\underline{X};t\right) \Big] \tag{2.4}$$

As an example, the CME for the dimerization process (reaction 2.1) is provided in equation 2.5:

$$\frac{\partial P\left(A, A_2; t\right)}{\partial t} = \left(\frac{k_1}{V N_A}\right) \cdot \Big[\left(A+2\right)\left(A+1\right) P\left(A+2, A_2-1; t\right) - A\left(A-1\right) P\left(A, A_2; t\right) \Big] + \\ + k_2 \cdot \Big[\left(A_2+1\right) P\left(A-2, A_2+1; t\right) - A_2 P\left(A, A_2; t\right) \Big] \tag{2.5}$$

The CME takes into consideration all the possible states that a system can be found at a particular time point. For example, a system with x species and y copies per species would feature y^x possible states. Bearing in mind that a eukaryotic cell has roughly 10^4 different proteins, and approximately 10^6 copies per protein, the possible states that the cell can end up being are about $10^{60,000}$. In addition, different proteins bind to each other generating more species and hence additional possible states. It is therefore clear that the state space is vast rendering the analytical solution of CME impossible for all but the simplest cases [64]. In addition, the computational cost of solving the CME numerically is prohibitive. To address this, Monte Carlo techniques have been developed approximating the probability distribution of the state space which would otherwise be calculated by the CME [65]. The first attempt was made by Gillespie who developed the so-called Stochastic Simulation Algorithm (SSA) [66, 67]. SSA and other recently developed algorithms which accelerated SSA are discussed in the following section.

2.2.2 Stochastic Simulation Algorithm and its Derivatives

Since the CME cannot be solved for the vast majority of systems, SSA was developed in 1976 [67] as a means to generate numerical realizations of $X(t)$ over time. The principle behind the theory of this algorithm is the following probability density function:

$P(\tau, \mu) \equiv$probability at time t that the next reaction in volume V will occur in the differential time interval $(t+\tau, t+\tau+dt)$ and will be an R_μ reaction.

$P(\tau, \mu)$ is nothing but a joint probability density function on the space of time, $t(0 \leq \tau < \infty)$, and $\mu(\mu = 1, 2, ..., M)$. This joint probability constitutes the mathematical basis of SSA and is calculated as follows (for a detailed description see [67]):

$$P(\tau, \mu) = \alpha_\mu \cdot exp\left(-\sum_{j=1}^{M} \alpha_j \cdot \tau\right) \tag{2.6}$$

where α_i is the propensity of reaction i, τ represents an exponential random variable with mean $\left(\sum_{j=1}^{M} \alpha_j\right)^{-1}$, and j are statistically independent integer random variables with probability $\alpha_j \Big/ \sum_{j=1}^{M} \alpha_j$. The first version of SSA, "direct method", implements the standard inversion method to propagate a pair of τ, j (time at which the next reaction happens and reaction index) according to the aforementioned distributions. More specifically, the algorithm draws two random numbers, r_1 and r_2, from a uniform distribution in the interval (0,1), and picks:

$$\tau = \frac{1}{\sum_{j=1}^{M} \alpha_j} \ln\left(\frac{1}{r_1}\right),$$

$$\mu = \text{the smallest integer satisfying: } \sum_{\mu=1}^{j} \alpha_j > r_2 \sum_{j=1}^{M} \alpha_j$$

In each step, SSA determines the reaction μ that is going to fire next and the time τ at which this reaction fires. This requires the simulation of each reaction event

separately rendering SSA significantly slow, especially when the number of simulated reactions is high. To address this challenge, several algorithms have been developed.

One derivative of the "direct method" was the "next reaction" method. Here, the algorithm calculates the time that each single reaction will occur next and picks the reaction with the smallest time as the reaction happening next. Even though this accelerated the "direct method", it is less efficient than the latter when the number of the reactions arising in the system is large.

Among others (for an excellent review see [65]), a remarkable improvement to the "next reaction" method was proposed by Gibson and Bruck [68]. In their method, the random numbers that are drawn but not involved in the next reaction (so they are not taken into account in the simulation of this reaction) are saved and reused. This sophisticated improvement resulted in a remarkable decrease of the required computational cost. It is important to underline that both the "next reaction" method and the method proposed by Gibson and Bruck accelerated SSA without introducing any approximation. Further improvements to SSA required the speed-accuracy trade off and several methods were developed aiming to accelerate SSA while sacrificing accuracy. These accelerated approaches fall into two categories: 1) the tau-leaping and 2) the system-partitiong.

The underlying assumption of the tau-leaping method is that the reaction propensities remain approximately constant during a time step and therefore the number of times a reaction fires is a Poisson random variable. An important method included in this family of approaches is the Chemical Langevin Equation (CLE). Besides the main assumption of the tau-leaping methods, this approach also postulates that in each time step each reaction fires many more times than once. Under these hypotheses, the amount of molecules changes continuously according to the following formula:

$$\frac{dX(t)}{dt} = \sum_{j=1}^{M} \nu_j \alpha_j \left(X(t)\right) + \sum_{j=1}^{M} \nu_j \sqrt{\alpha_j \left(X(t)\right)} \Gamma_j(t) \tag{2.7}$$

where $\Gamma_j(t)$ represents statistically independent Gaussian white-noise processes and equals to $\sqrt{\tau}$ times a statistically independent normal random variable with mean equal to zero and variance equal to one. The RHS of equation 2.7 is the sum of a deterministic drift term and a stochastic diffusion term which is proportional to Gaussian white

noise. As the system approaches the thermodynamic limit, the fluctuations become insignificant, since the stochastic diffusion term vanishes, and, in turn, the CLE becomes equivalent to the mass action kinetics. Overall, even though the tau-leaping methods advanced SSA, they are inaccurate in cases where the reaction propensities change significantly within a time step. In the present dissertation, the CLE was implemented in Chapter 6 to describe the *in silico* design of a synthetic bacteria-yeast ecosystem.

An example of implementing the CLE in the process of dimerization (2.1) to monitor the evolution of the monomer and dimer molecules is shown in 2.8 and 2.9 respectively.

$$dA = \left(2k_2A_2 - \left(\frac{k_1}{VN_A}\right)A(A-1)\right)dt - \sqrt{\left(\frac{k_1}{VN_A}\right)A(A-1)}dw_1 + 2\sqrt{k_2A_2}dw_2 \tag{2.8}$$

$$dA_2 = \left(\left(\frac{k_1}{VN_A}\right)A(A-1) - k_2A_2\right)dt + \sqrt{\left(\frac{k_1}{VN_A}\right)A(A-1)}dw_1 - \sqrt{k_2A_2}dw_2 \tag{2.9}$$

The second category, system-partitioning, separates the system into a slow and a fast subset. The reactions included in the slow subset are usually treated with SSA, or a similar approach, whereas the fast reactions are implemented using an approximation that eventually speeds up the algorithm. These "hybrid methods" encapsulate several approaches such as stochastic, deterministic, discrete, continuous, and combinations of those. These methods have been shown to not only enhance the computationally efficiency but also retain their accuracy to a large extend. Among the numerous hybrid methods described in the literature [69, 70, 56], here we focus on a method developed by Kaznessis group [55].

It is noteworthy that software packages that implement the above algorithms and feature a Graphical User Interface (GUI) have been also launched. These computational tools have significantly simplified the modeling and simulation of not only natural but also synthetic biological systems. Remarkable examples include COPASI [71], CellDesigner [72], SynBioSS [73, 74], and others [75, 76].

2.2.3 Hybrid Stochastic-Discrete, Stochastic-Continuous Algorithm

The main concept behind the hybrid, stochastic-discrete and stochastic-continuous, algorithm is the partition of the simulated system into a slow and a fast subset. In order for a reaction to belong to the slow subset, two criteria must be satisfied. These criteria are associated with the number of molecules participating in this reaction as well as the times that this reaction occurs within a time step. The mathematical formulation of these criteria is the following:

$$1)\ a_j(X(t)) \geq \lambda \gg 1,$$
$$2)\ X_i(t) > \epsilon|\nu_{ji}|$$

where i is the reactant or product of the j reaction, $a_j(X(t))dt$ is the number of times that the reaction happening within dt, and ϵ is the number of molecules of each reactant. The default values of λ and ϵ have been set equal to 10 and 100 respectively [55]. In other words, a reaction is considered slow if and only if it fires at most 10 times during dt and the number of each involved species is less than 100. If either of those criteria is not met, the reaction is then fast. In every time step, each reaction is characterized as slow or fast and therefore the reactions can transition multiple times between the fast and slow regime throughout the simulation depending upon the two aforementioned criteria.

The evolution of the fast reactions is calculated using the CLE described in 2.2.2. On the other hand, the slow reactions are simulated using a method inspired by the "next reaction" method. In particular, the algorithm uses a sophisticated way to monitor the occurrence of a reaction. This is achieved through equation 2.10 (for the derivation see [55]):

$$\frac{dR_j}{dt} = a_j(t)dt,\ IC : R_j = \log(r_j),\ j = 1, ..., M^{slow} \tag{2.10}$$

where R_j is the residual of the j reaction, and r_j is a uniform random number with mean equal to 0 and variance equal to 1. Equation 2.10 is a stochastic differential equation as it is a function of the state of the system which is indeed a stochastic process.

As observed in equation 2.10, a negative residual, R_j, is initially assigned to each reaction. This residual changes dynamically according to the propensity of the reaction. When a residual crosses zero, i.e. becomes positive, the corresponding reaction has just occurred. This reaction is then recorded, the state of the system is updated, and the algorithm proceeds. In this algorithm, the user can either allow for one reaction per time increment or can speed up the algorithm by allowing two or more reactions happening within the time increment. In the latter case, the slow reactions are executed in the order of occurrence.

This hybrid algorithm has been tested and found to outperform other comparable algorithms [55]. It features a GUI, Hybrid Stochastic Simulations for Supercomputers (HY3S), and is publicly available at http://hysss.sourceforge.net/ [74] and at http://synbioss.org/ [73].

In the current thesis, this algorithm has been used in three studies to capture the dynamics of both natural and synthetic systems. In particular, in Chapter 3, this algorithm elucidated the dynamics of a natural system which confers resistance to bacteria and widely used in synthetic biology applications. Similarly, in Chapter 4, the algorithm was used to describe the response of bacteria to different types of sugars. Finally, in Chapter 5, this algorithm was implemented to model and characterize two synthetic genetic switches built in bacteria. In each of those studies, a reaction network capturing the biomolecular interactions engaged in the system of interest was developed. The algorithm was then utilized to calculate the evolution of each species.

Chapter 3

Stochastic simulations of the tetracycline operon

Recent advances in our ability to mathematically investigate the dynamic complexity of biomolecular systems, have created inroads into these systems. Examples include natural systems, such as the lactose [77, 78] and tryptophan [79] operon, and synthetic systems such as the oscillator [4, 5], logic AND gates [12] and toggle switch [7]. In the present Chapter, we examine the dynamic behavior of the tetracycline (*tet*) operon. Although some studies have examined the interactions of different parts of the *tet* operon [80, 81], to our knowledge there is no mathematical model that describes all the biomolecular interactions of this intriguing system. The *tet* operon is found naturally in bacteria where it confers resistance to antibiotic tetracycline (Tc).

Tc used to be one of the most common antibiotics for treating bacterial infections. It functions by binding the bacterial ribosome (Rib), thereby impeding the process of translation (protein biosynthesis) and causing cell death [82]. Due to its low cost, Tc was used excessively. Because of the excessive use, bacteria developed resistance to it.

Already in 1964, there was evidence that *E. coli* bacteria had developed resistance to Tc, but the exact resistance mechanism was not clear [83]. To date, four resistance mechanisms have been identified [84]. These mechanisms are associated with the a) active efflux of Tc out of the cell, b) Rib protection from Tc, c) rRNA mutation and d) Tc inactivation. In the present work, we investigate the mechanism of active transport

of Tc out of the cell, whereby bacteria under attack by Tc, quickly produce a membrane protein that pumps Tc out of the cell. This resistance mechanism relies on the *tet* operon [85]. Several Tc resistance determinants have evolved [84, 86]. This work refers to class B (or Tn10-type) Tc resistance determinant. A qualitative model of the *tet* operon is shown in Figure 3.1 [87].

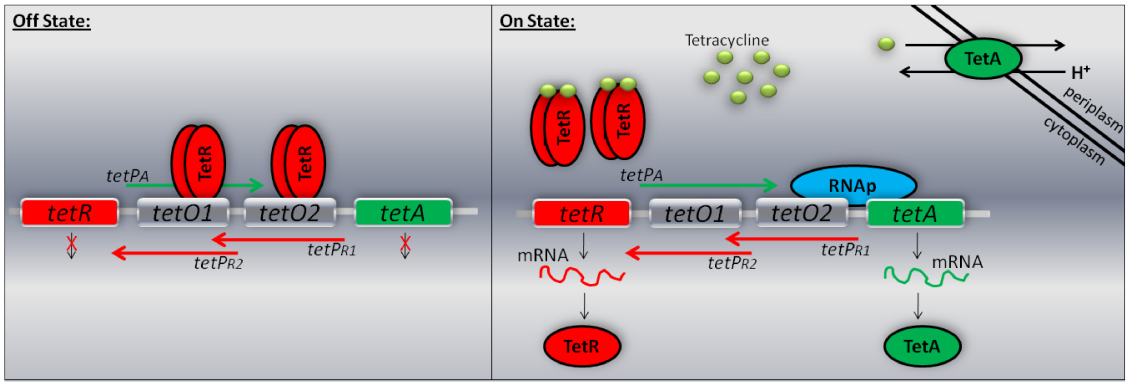

Figure 3.1: Schematic representation of the *tet* operon without (off state) and with (on state) Tc. When Tc is absent, the cells save energy by repressing the expression of *tetA* and *tetR*. When Tc is present, TetA is produced and pumps Tc out of the cell. More TetR is also produced to shut down expression again when it is no longer necessary.

The *tet* operon comprises of two genes, *tetA* and *tetR*. The first encodes TetA, a transporter protein which removes Tc from the cell. The *tetR* gene encodes TetR which binds to the operators, acts as a repressor and inhibits the expression of both of genes in the absence of Tc. Moreover, this operon consists of three promoters and two operators. One promoter belongs to the *tetA* gene while the other two belong to the *tetR* gene. The two operators control the expression of the two genes. The promoters and the operators overlap, significantly increasing the complexity of this biological switch.

In the absence of Tc, TetR binds as a dimer (TetR2) to both of the operators. TetR2 binding to the operators results in the repression of gene expression. When Tc diffuses into the cell, it binds to Rib preventing protein synthesis. Importantly, it also binds to the TetR2 repressor, with high affinity, causing a conformational change in the DNA binding region. This results in TetR2 unbinding from the operator sites. Once TetR2 has dissociated from the operator sites, gene expression is turned on producing

TetA and TetR. Subsequently, Tc is removed from the cell through the active transport mechanism mediated by TetA. After the expulsion of Tc from the cell, TetR2 protein binds the operator sites again, turning off the expression of the two genes [85, 87, 88, 89, 90, 91, 82, 92].

It is interesting to note that expression of genes is activated only in the presence of the antibiotic Tc. This is a remarkable cost-effective mechanism, a feature that has made the molecular components (promoters, operators, repressors) of this system attractive for the design and development of a gamut of synthetic gene networks [4, 12, 80, 93, 42, 94]. The molecular elements of this operon are used as parts of biological switches, taking advantage of the fast and robust switching. More specifically, the expression of desired proteins is placed under the transcriptional control of the *tet* components, where protein production can be efficiently turned on and off in response to the presence or the absence of Tc.

Herein, we formulate a novel mathematical model of the naturally occuring *tet* operon based on experimental findings. Our model incorporates all the biomolecular interactions of this system including those involved in dimerization, transcription, translation, degradation, repression and induction. The model represents each elementary biomolecular interaction with biochemical reactions.

We use a hybrid, stochastic-discrete and stochastic-continuous algorithm to simulate the behavior of the *tet* operon in order to understand how it responds to various disturbances [74, 55]. The use of the stochastic algorithm allows us to consider single cell time lines as well as variability across different cells. Lastly, a sensitivity analysis of the binding strength of Tc for TetR2 and of TetR2 for the operator sites is performed, providing mechanistic insight into the working of this interesting system.

3.1 Methods

3.1.1 Model description

Our model consists of 40 species and 61 biochemical reactions which capture the biomolecular interactions occuring in the *tet* operon. The entire reaction network is supplied in Table 3.1.

The definition of the species that participate in the *tet* operon reaction network and

the kinetic constants of the reactions are found in Tables 3.2 and 3.3 respectively. In what follows, we provide a description of the salient features of this operon. We explain also the reactions that we use to describe the biomolecular processes taking place in this system.

As illustrated in Figure 3.1, the *tet* operon is composed of 2 genes, *tetR* and *tetA*. These give rise to two proteins, TetR and TetA. Let us start with the interactions of TetR. TetR2, the dimer form of TetR, is involved in inhibiting gene expression in the absence of Tc [95]. We model the dimerization of the TetR as a second order reversible reaction (reaction 1).

TetR2 protein represses the system by binding to the operator sites of the *tet* operon. We designate these two operators as *tetO1* and *tetO2*. The binding of the repressor protein to the *tetO1* shuts down the expression of both the two genes. On the other hand, the binding to the *tetO2* impedes the expression of the *tetA* gene and only down regulates the expression of the *tetR* gene [96]. There is no indication of cooperativity between these two operator sites [97]. We represent the binding of TetR2 to the operator sites, *tetO1* and *tetO2*, as a second order bimolecular reaction (reactions 2,3).

TetR2 can bind to non-specific DNA sites (nsDNA) as well [97]. Even though the affinity of the repressor protein for the non-specific sites is low, there are about 4 million non-specific sites in the *E.coli* genome thereby significantly increasing the probability of TetR2 being bound to non-specific DNA. These non-specific binding interactions are captured by reaction 4.

It is noteworthy that the affinity of the repressor for *tetO2* is about five times higher than for *tetO1* [98]. It has been theorized that this mechanism is used by *E. coli* to avoid fortuitous expression of the *tetA* gene. An unexpected decrease in the number of TetR protein would favor the repression of the *tetA* gene first, until more TetR proteins are produced and expression of the two genes is inhibited again. In any case, the repressor shuts down the expression of the *tetA* gene first and soonafter the expression of the *tetR* gene [87].

The diffusion of the Tc into the cell is represented by a first order reaction capturing the effective rate at which Tc enters the cell (reaction 5). We should note that although Tc is certain to bind other targets in the cell, in our model we only regard an effective amount of Tc that interacts only with the *tet* operon. Once Tc diffuses into the cell,

it binds to a metal ion, M, yielding the complex M-Tc [91]. For simplicity, in our model, we neglect the formation of this complex which occurs almost instantaneously. Subsequently, Tc binds to the repressor (TetR2 bound to the operators or free) with high affinity. Upon binding, TetR2 undergoes a conformational change that reduces dramatically its affinity for the operators sites [99, 100]. The binding of Tc to TetR2 is coded as a second order bimolecular reaction. Each TetR2 protein harbors two binding sites. The two Tc molecules bind in succession, and without cooperativity, to yield TetR2:Tc2 [91, 101]. These interactions are described by reactions 6-13.

When the operator sites are free from TetR2, transcription will occur. Transcription can be separated into three stages: initiation, elongation and termination. In order for transcription initiation to occur, the holoenzyme, which is composed of sigma factor and RNA polymerase (RNAp), is recruited to the promoter sites. There, it forms a closed complex [102]. This entire process is effectively modeled as a second order bimolecular reaction (reaction 14). After the formation of the closed complex, the DNA double helix is unwound forming an open complex (reaction 15). Once the open complex has been formed, RNAp starts to transcribe each single nucleotide and the sigma factor is released from the holoenzyme. The transcription of the first nucleotide is considered a first order reaction (reaction 16). Subsequently, RNAp moves along the DNA transcribing each nucleotide independently of the others [102]. This process could be modeled as a series of reactions, each of them representing the transcription of each single nucleotide. Considering the reaction times as exponentially distributed events, we integrate all these reactions in a single reaction that represents a gamma distributed event (reaction 17) [68]. The scale parameter is equal to the rate of the transcription of each single nucleotide while the shape parameter equals the number of nucleotides that are transcribed.

As shown in Figure 3.1, there are three promoter binding sites in the *tet* operon where the holoenzyme is recruited. We designate these binding sites as $tetP_{R1}$, $tetP_{R2}$ and $tetP_A$. $tetP_{R1}$ and $tetP_{R2}$ regulate the *tetR* gene whereas $tetP_A$ controls the *tetA* gene. If an RNAp is recruited to one of the three sites, it prohibits another RNAp from being recruited to one of the other two [89]. Furthermore, the affinity of RNAp for each of the three sites is different [90, 103]. Thus, we include three different cases (one case for each single promoter) to capture all the different possible ways that transcription

initiation can occur. We model the first case (site $tetP_{R1}$) using the reactions 14-17.

Reactions 14-17 depict the transcription of the *tetR* gene when an RNAp is recruited to the $tetP_{R1}$ site. The $tetP_{R1}$ site accounts for approximately 5% of the total mRNA transcripts of the *tetR* gene [103]. The brackets indicate the promoter site from which transcription initiation occurs. In reaction 14, *Complex* refers to the DNA molecule when all of the promoter sites and all of the operators are free. When all the sites are free, they form a single, contiguous complex that RNAp can bind to and start transcription. The formation of this complex is described in reaction 41. It should be stressed that this reaction has no physical meaning and it serves as an algorithmic trick. To this end, reactions 42-45 are the same as forward reactions 2 and 3 and backward reactions 10 and 13. Similarly to reactions 14-17, reactions 18-21 represent the transcription of *tetR* gene when RNAp binds to the promoter site $tetP_{R2}$ (second case).

Tet operon is a tightly regulated system. One of its striking features is that even if the repressor (TetR2 protein) is bound to the operator *tetO2*, RNAp can still be recruited to the promoter $tetP_{R2}$ and start transcription if *tetO1* is free [96]. The affinity of RNAp for the $tetP_{R2}$ is slightly lower in this case than in the case when both the promoters, *tetO1* and *tetO2*, are free. This mechanism is modeled by reactions 22-24.

Similarly to the process of transcription, the process of translation also progresses in three steps: initiation, elongation, and termination. The initiation stage includes the association of mRNA with the ribosomal units (50S and 30S) and the initiator tRNA to form the initiation complex [102]. This entire process is represented as a first order irreversible reaction (reaction 25). After the initiation complex forms, Rib translates the mRNA molecule into its protein products [102]. The movement of the Rib from the ribosome binding site to the coding region along with the release of the mRNA molecule are described in reaction 26. The elongation process is integrated, as in the transcription process, in one gamma distributed reaction event(reaction 27) [68]. In this case the scale parameter is equal to the rate of the translation of each single codon and the shape parameter equals the number of amino acids that are produced. The species utilized in transcription and in translation (Rib, RNAp, promoters, operators) are finally freed to participate in these processes again. The translation mechanisms are

shown in reactions 25-27.

We use the same pattern of reactions to model the expression of the *tetA* gene as we did to model *tetR* expression. In this case, there is only one promoter site, namely $tetP_A$, where transcription initiation can take place (third case). Using a reaction formalism to capture transcription and translation we come up with reactions 28-34.

The proteins as well as the mRNAs are degraded during the cell life. mRNAs are degraded by ribonuclease enzymes. The protein degradation is catalyzed by proteases. The degradation of both the mRNA and proteins is modeled as simple first order reactions (reactions 35-38). We also consider Tc degradation (reaction 39).

Reactions 35 and 36 refer to the degradation of TetR2 protein. TetR2 can be degraded when it is free (reaction 35) or when it is bound by Tc (reaction 36). To our knowledge, there is no study that demonstrates TetA degradation. Therefore, we assume that TetA is not degraded, a plausible assumption for any membrane protein. However, in our model TetA gets diluted due to cell division.

Finally, reaction 40 captures the removal of Tc from the cell. As mentioned above, the exclusion of Tc from the cell occurs through a process controlled by the TetA protein. This process is energy dependent and is driven by a pH gradient across the membrane [92]. The removal of one molecule of Tc is coupled by the influx of one proton in the cell [88]. Eventually, TetA remains in the membrane of the cell while Tc is pumped out of the cell. There is evidence that TetA protein can exist as a dimer or even as a multimer in the cell, but this is still unclear [104]. For the sake of simplicity, in our formulation we consider TetA to exist as a monomer.

3.1.2 Model parameters

In Table 3.3, we present the values of the parameter for the model. The tetracycline operon is one of the best studied bacterial gene networks. As such, there is an atypical wealth of parameters on the strength of biomolecular interactions. Therefore, most of the equilibrium constants that we used in our model have been obtained through experimental procedures. The relevant literature references are also presented in Table 3.3.

A concern may be legitimately posed that in most cases strengths of biomolecular interactions are measured *in vitro*, that is outside of the pertinent biological context.

Indeed, this is a quintessential challenge faced by quantitative biology. This is the reason we chose to study the sensitivity of the operon behavior to changes in the values of important parameters.

In addition, in most cases of reversible reactions, only the equilibrium constant is actually available. Then the kinetic constant of the forward reaction is assumed and the reverse reaction kinetic constant is subsequently calculated through the equilibrium constant.

The reaction rate of reaction 41 is set equal to 10^8 aiming to make it a very fast reaction. Having in mind that the TetR2 dimerization equilibrium constant is greater than $10^7 M^{-1}$ [95], we consider it equal to $10^8 M^{-1}$ (equal to the one observed for the lactose repressor, LacI) [80].

It should be mentioned that the binding of Tc to TetR was modeled based on [101]. More specifically, the kinetic constant of the binding of the first Tc molecule to TetR2 (either unbound or bound to operators) is twice the kinetic constant of the binding of the second Tc molecule to TetR2. Consequently, the kinetic constant of the unbinding of the second Tc molecule from TetR2 is twice the kinetic constant of the unbinding of the first Tc molecule.

Another choice we made was on the degradation rate of TetR2 protein, which is chosen with 5 hours half time. Furthermore, the rate at which Tc is degraded is assumed to be equal to the degradation rate of Tc in distilled water [105]. In addition, the mRNA degradation rate was adjusted such that 20 protein molecules per mRNA transcript are produced [80].

Regarding the affinity of RNAp for the $tetP_A$ promoter, it has been shown to be approximately nine times higher than the combined $tetP_R$ promoters. To facilitate this, the total affinity of the RNAp for the two $tetP_R$ promoters is set equal to 1/9 of the affinity of RNAp for $tetP_A$. Yet, the open complex (between RNAp and DNA) formation kinetic parameter is assumed to be equal for all the three promoters. Additionally, $tetP_{R2}$ promoter accounts for 95 % of the total mRNA while the remaining 5% imputes to $tetP_{R1}$. Therefore, the binding strength of RNAp to $tetP_{R2}$ is 19 times higher than the binding strength of RNAp to $tetP_{R1}$. Moreover, when TetR2 is bound to *tetO2*, the expression of $tetP_{R2}$ is decreased by 16% and thus, the corresponding kinetic constant is equal to 84% of the case where *tetO2* is free [96].

In general, both the transcription and the translation rates vary significantly. In our formulation, we postulate that the transcription rate is 30 whereas the translation rate is 100 nucleotides per second [106].

3.1.3 Stochastic simulations of systems with chemical reactions

As described in Chapter 2, biological systems are generally not at the thermodynamic limit. They are composed of molecules, such as promoters and operators, that participate in the biological processes in low copy numbers [58, 28, 59]. The result is that randomness becomes important. Deterministic approaches cannot take into account this randomness in the behavior of biological systems. Stochastic approaches are therefore needed [107, 66]. For the purposes of this study, we used HY3S to simulate the behavior of the *tet* operon. A description of HY3S is provided in Chapter 2.2.3.

3.1.4 Underlying assumptions and conditions

In this section, we delineate the assumptions and conditions of our model. The cell is considered to be an isolated, well stirred, homogeneous reactor. The cell volume is considered equal to 10^{-15} L. It is also considered that the cell volume increases exponentially. Cell division is assumed to occur every 30 ± 4 minutes. The exact time for each cell division is randomly chosen from a Gaussian distribution whose mean is at 30 minutes. Moreover, we hypothesize that during the cell division, the number of molecules of the proteins, mRNAs, and Tc is halved. In order to simplify our model, we do not take into consideration the Tc binding to the Rib assuming that the intracellular Tc amount that is used here is the effective amount that the *tet* operon has to remove from the cell. Furthermore, the diffusion of intracellular Tc out of the cell is considered negligible compared to the Tc removal from the cell by TetA. Finally, the conditions of the cell, such as pH and temperature, are considered to be constant during the simulations.

For each single simulation, 1,000 trials (which correspond to 1,000 cells) were carried out. In what follows, we present not only the average behavior of the 1,000 cells, but also the behavior of single cells. It is important to stress that the use of stochastic algorithms allows for exploring single cell behaviors and variances across the 1,000 cells. This would

not be possible with the use of deterministic approaches. The initial conditions that were used in the simulations are 300 Rib, 180 RNAp, 3 promoters ($tetP_{R1}$, $tetP_{R2}$, $tetP_A$), 2 operators (*tetO1*, *tetO2*), and $4 \cdot 10^6$ nsDNA sites. Even though the actual number of Rib and RNAp that exist in the *E. coli* is higher [108], we consider that only a small fraction of the total number is available to this operon.

3.2 Results and Discussion

3.2.1 Simulation of the *tet* operon behavior

The first set of simulations explores the steady state of the system. According to experimental observations, at steady state and in the absence of Tc, there is no intracellular TetA and the amount of intracellular TetR is small. This is a convenient attribute of the *tet* operon. The few TetR proteins require only a small number of Tc molecules to induce the expression of the genes [109]. Moreover, TetA should not be produced in the absence of Tc because large amounts of TetA can lead to cell death [110]. TetA is a membrane protein that pumps Tc out of the cell while pumping protons from the periplasm into the cytoplasm. If TetA is present, even in the absence of Tc, transport of protons through the membrane may result in the collapse of the membrane potential and cell death.

In order to investigate the steady state of the system, we conducted a set of simulations in which neither TetA nor TetR protein initially exist in the cell. Therefore, gene expression is initially allowed to take place producing TetR which inhibits gene expression thus bringing the system to the steady state. The results for both the average behavior of the system and the behavior of a single cell are shown in Figure 3.2. For TetA (Figure 3.2e), we observe an average initial production of 113 molecules, whereas after 4 hours there is virtually no TetA left in any cell due to dilution. The initial pulse is the result of no initial repression by TetR. Concerning the average behavior of TetR and TetR2 (Figures 3.2a,3.2c), they reach a maximum of about 5 and 2 molecules per cell respectively and approximately after 3 hours reach steady state. At steady state, there are approximately 2 TetR and 1 TetR2 molecule in the cell.

It is important to notice that there is a difference between the average amount of TetA and TetR protein. This is because the promoter of the *tetA* gene is approximately

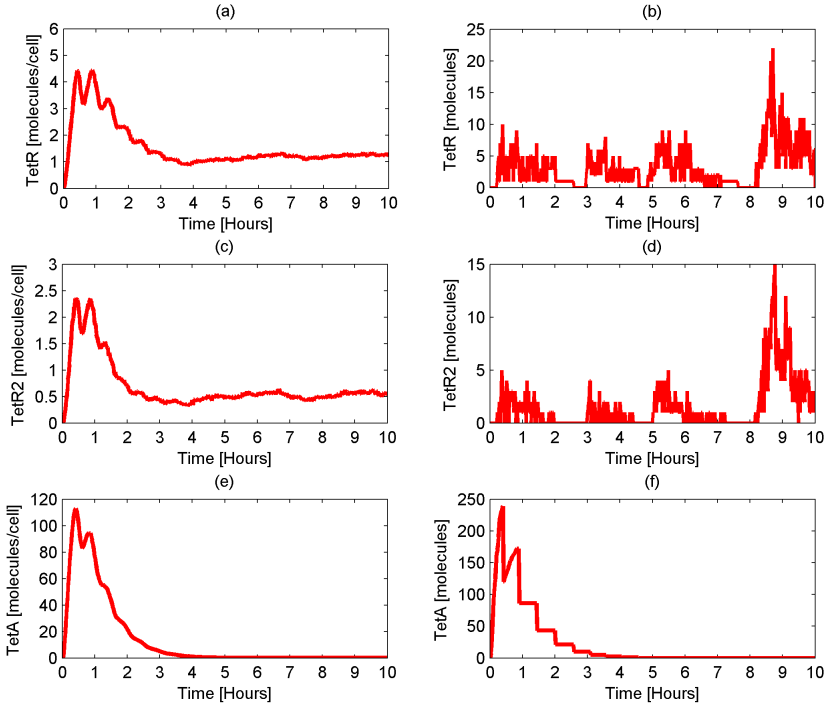

Figure 3.2: Average (Figures 3.2a,3.2c,3.2e) and single cell (Figures 3.2b,3.2d,3.2f) number of molecules of TetR, TetR2, and TetA protein at steady state. Initially, gene expression is allowed to take place. As a result, the amount of the two proteins (TetR and TetA) in the cell increases. Once TetR molecules have been produced, they repress gene expression allowing the system to reach a steady state.

nine times stronger than the combined *tetR* promoters [90]. This implies that the expression of the *tetA* gene is very high compared to the *tetR* gene when both of the genes are expressed. Moreover, it should be kept in mind that some of the TetR proteins are bound to non-specific sites as well as to Tc, making the amount of free TetR smaller.

Interestingly, looking at the single cell behavior (Figures 3.2b,3.2d) we observe large fluctuations in the number of TetR and TetR2 molecules. According to Figures 3.2b and 3.2d, the fluctuations of TetR and TetR2 reach a maximum of 22 and 15 molecules respectively. It should be kept in mind that the dynamics of those two Figures refer to a single cell. At steady state, the maximum TetR and TetR2 amount observed across the 1,000 cells is 31 and 38 molecules respectively (data not shown). It is worth mentioning that even though the average number of TetR and TetR2 is small (2 and 1 respectively),

there are cells that incorporate TetR2 molecules whose amount fluctuates around high values. However, these cells are only a few and therefore, the average amount of TetR and TetR2 is kept small. Apparently, in this case the average behavior of the cell is not representative of the single cell behavior and this necessitates the use of stochastic algorithms. Had we used a deterministic approach, fluctuations in the behavior of TetR2 could not be established.

Concerning the TetA protein, in the case of a single cell (Figure 3.2f), we note that a maximum of 239 molecules is reached. Subsequently, the number of TetA molecules becomes zero after 3.5 hours. At steady state, the maximum number of TetA molecules observed across the 1,000 cells is 348 (data not shown). It should be underlined that cells with high intracellular TetA amounts at steady state cannot survive due to the membrane collapse.

Importantly, these results are consistent with experimental observations, in that not only the number of intracellular TetA molecules is zero at steady state [87, 110], but also in that there is only a small amount of free TetR [109]. We have thus established the steady state conditions of TetR and TetA in the cell.

Our next step is to test the behavior of the system when a pulse of Tc is administered to the cell. A set of simulations was carried out varying the number of Tc molecules administered to the *E. coli* cells. More specifically, we start with the system at steady state and after 5 hours, a pulse of a wide range of Tc molecules is administered to each cell. Administration of different Tc amounts results in the production of different TetA amounts. The results are portrayed in Figure 3.3.

Focusing on the TetA amount at time 5 hours, we notice that the larger the quantity of Tc in the cell, the faster is the response of the system. This is pursuant to the fact that the *tet* operon is a well regulated system which responds fast to the addition of Tc. Furthermore, we observe a nearly linear correlation between the Tc and TetA amounts. This correlation is more clear in Figure 3.6a. An approximately linear dependence between TetA and non-toxic administered Tc amount has been observed before [111]. Given that in our case the administered Tc amount is non-toxic, this result is in agreement with the results by Korpela et al. It is worth stressing that after the removal of Tc from the cell, the system returns again to its original steady state. The time that the system needs to reach the steady state depends on the number of TetA molecules

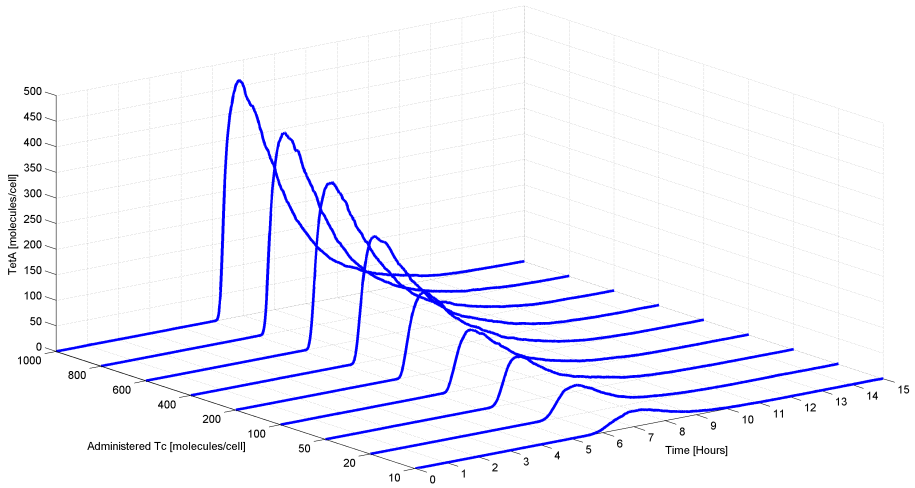

Figure 3.3: Average number of TetA molecules for different number of administered Tc molecules. Initially, there is no TetA in the cell because there is no intracellular Tc. At time equal to 5 hours, a pulse of 10, 20, 50, 100, 200, 400, 600, 800 and 1,000 Tc molecules is administered to each cell thereby inducing gene expression. Then, the amount of intracellular TetA protein increases. Increasing the amount of administered Tc results in faster production of higher TetA concentration.

in the cell. The higher the number of TetA molecules, the longer it takes for the system to reach the basal state.

Another important fact is that even with the administration of a small Tc amount (10 molecules), the gene expression is turned on automatically in order for the Tc to be removed before causing cell death. This is in line with experimental observations which suggests that induction takes place even with very low, non-toxic concentrations of Tc [87, 112, 89, 95, 85].

In addition to investigating TetA levels in response to Tc treatment, we also investigate TetR levels at these different Tc concentrations. The average number of intracellular free (unbound) TetR2 molecules practically remains constant regardless of the number of administered Tc molecules. This is an interesting result, especially when contrasted to the TetA behavior. This trend is shown in Figure 3.4.

As illustrated in Figure 3.4, the average number of TetR2 molecules decreases slightly when Tc is administered to the cells. This decrease is caused by the binding of Tc to the free TetR2 molecules. However, shortly after the decrease on the TetR2 amount, an

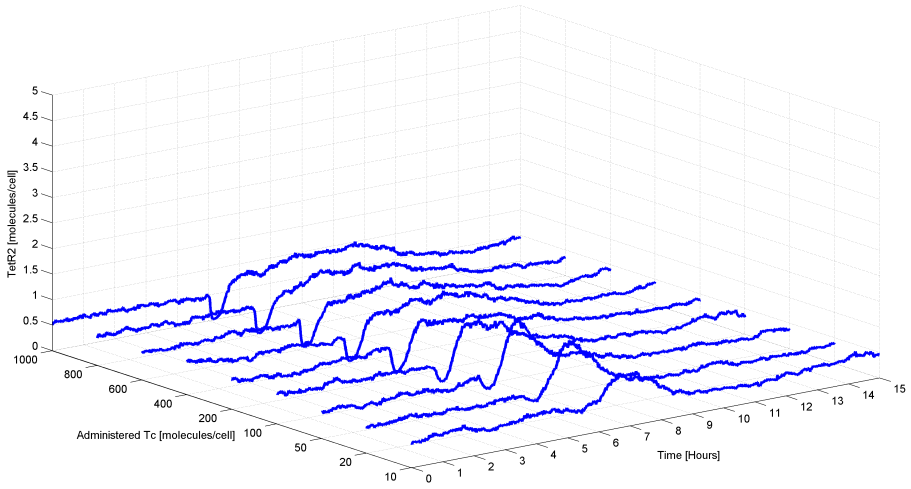

Figure 3.4: Average number of TetR2 molecules for different number of administered Tc molecules. Initially, there are about 2 TetR2 molecules in the cell. At time equal to 5 hours, a pulse of 10, 20, 50, 100, 200, 400, 600, 800 and 1,000 Tc molecules is administered to each cell thereby inducing gene expression. Then, the amount of the unbound TetR2 protein decreases because these proteins are occupied by Tc. Subsequently, and upon expression of the tetR gene, the amount of TetR2 increases. Increasing the amount of administered Tc results in higher decrease in TetR2 amount and longer required time for going back to steady state.

increase to the TeTR2 amount is observed which comes from the induction of the system which in turn causes the production of many TetR2 molecules. After this increase in the TetR2 amount, the system finally reverts to its steady state. We should mention here, that the higher the amount of administered Tc, the larger the decrease of the TetR2 amount and the longer the time that the system needs to go back to its steady state.

The dynamics of the intracellular Tc when different Tc amounts are administered to the cell are also explored. Figure 3.5 shows the average intracellular Tc amount for the 9 different cases. As expected, the average number of intracellular Tc molecules increases with the number of administered Tc molecules.

In Figure 3.6 we show the mean and the variance of the maximum number of TetA (Figure 3.6a) and intracellular Tc (Figure 3.6b) molecules for the cases that are shown in Figures 3.3 and 3.5. As evident in Figure 3.6a, the maximum TetA amount produced by each cell upon Tc administration varies significantly. It is important to notice that

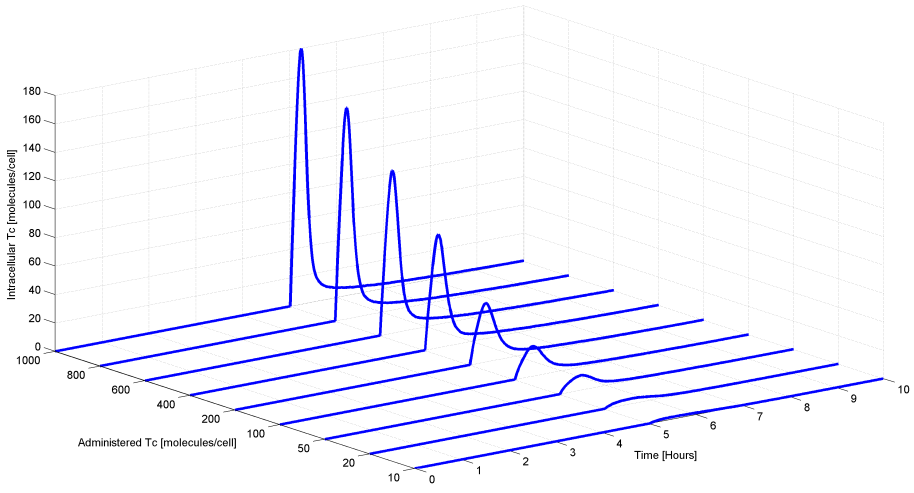

Figure 3.5: Average number of TetR2 molecules for different number of administered Tc molecules. Initially, there are about 2 TetR2 molecules in the cell. At time equal to 5 hours, a pulse of 10, 20, 50, 100, 200, 400, 600, 800 and 1,000 Tc molecules is administered to each cell thereby inducing gene expression. Then, the amount of TetR2 decreases and thereafter increases. Increasing the amount of administered Tc results in higher decrease in TetR2 amount and longer required time for going back to steady state.

even though the average value of the TetA amount in the cell is non-zero for all the 9 cases, in the first 4 cases (10, 20, 50, 100 administered Tc molecules) there are cells that produce no TetA protein upon Tc administration. These cells would probably not survive from the Tc administration since expression of the resistance protein was not activated and consequently Tc was not removed.

Analyzing Figure 3.6b, we observe that the variation of the number of intracellular Tc molecules seems to be smaller than the variation of the TetA molecules. Furthermore, we note that the higher the number of administered Tc molecules, the higher the variation of the intracellular Tc molecules. It is important to remark that cells whose intracellular Tc amounts lie in the higher regions is less probable to survive while cells with small intracellular Tc amounts (below the average) is most probable to circumvent the attack of Tc to their transcriptional machinery.

In experimental conditions, *E. coli* usually experience prolonged exposure to Tc. Thus, motivated by the need to understand the behavior of the *tet* operon when in a

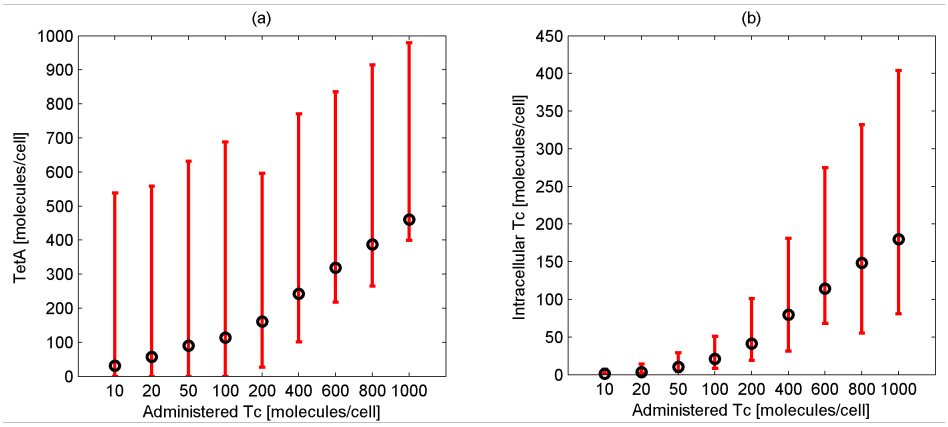

Figure 3.6: Mean (black color) and variance (red color)of the number of TetA (Figure 3.6a) and intracellular Tc (Figure 3.6b) molecules for different number of administered Tc molecules. The variation of both the number of TetA and intracellular Tc molecules is high.

solution with the antibiotic Tc, we performed a set of simulations where Tc is constantly administered to the cells. The system is initially at steady state. Then, after five hours, 500 Tc molecules are administered continuously. This is tantamount to having a constant concentration of external Tc equal to $8.3 \cdot 10^{-11}$ M under the following assumptions: 1) the volume of the cells is negligible compared to the volume of the solution, 2) the total Tc amount is ultimately uptaken by the cells, 3) the number of cells in the solution is $10^8/mL$. The average behavior of the system (Figures 3.7a,3.7c,3.7e) as well as the behavior of a single cell (Figures 3.7b,3.7d,3.7f) are illustrated in Figure 3.7.

Figure 3.7a shows that when Tc is added to the media, the average intracellular Tc amount reaches a maximum of 114 molecules within the first hour. Additionally, the average number of TetA molecules (Figure 3.7e) increases dramatically reaching a maximum of 540 molecules. Afterwards, the system reaches a steady state where there are 9 intracellular Tc molecules and 540 TetA molecules. The average amount of TetR2 (Figure 3.7b) decreases almost to the zero value when Tc is administered to the cells. This decrease caused by the binding of Tc to TetR2 which leads to occupation of the free TetR2 molecules. After the induction of the system and consequently the expression of

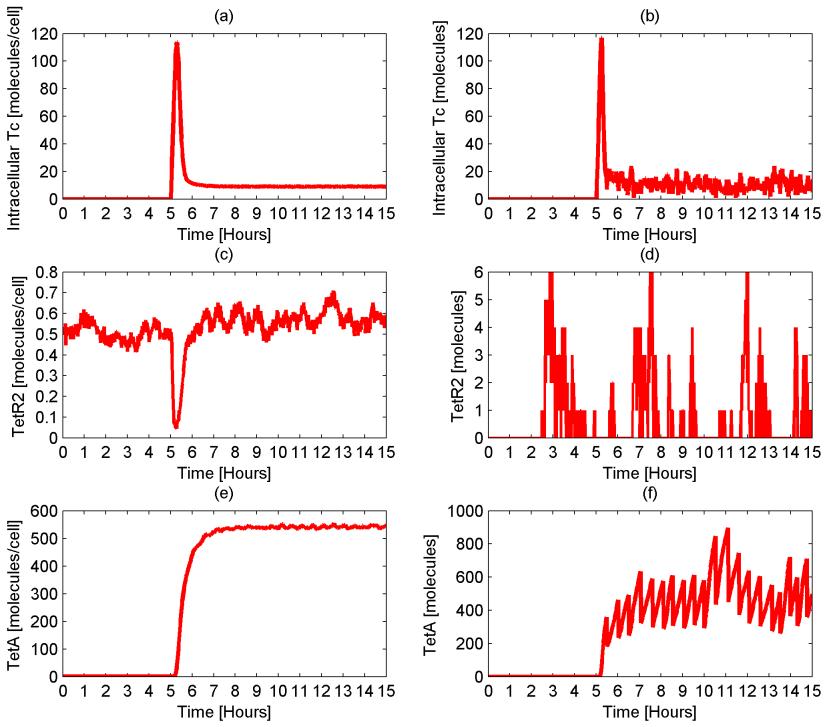

Figure 3.7: Average (Figures 3.7a,3.7c,3.7e) and single cell (Figures 3.7b,3.7d,3.7f) number of molecules of intracellular Tc, TetR2, and TetA protein when 500 Tc molecules are continuously administered to each cell. Initially, gene expression is allowed to take place. As a result, the amount of the two proteins (TetR and TetA) in the cell increases. Once TetR molecules have been produced, they repress gene expression allowing the system to reach a steady state.

the *tetR* gene, more TetR2 molecules are produced and eventually the TetR2 amount goes back to its steady state.

Notably, even with a constant high amount of external Tc, the *tet* operon does not allow high levels of internal Tc. This is attributed to the production of high TetA amounts which remove Tc from the cell, thus boosting the performance of the *tet* operon. It should be kept in mind that the intracellular Tc amounts are also decreased because of the degradation and the cell division in which the number of Tc molecules is halved every 30 ± 4 minutes. Overall, with constant Tc administration, TetA expression reaches steady state. At this state, the TetA levels become constant and intracellular Tc levels

are significantly lower than the extracellular Tc levels.

As far as the single cell behavior for Tc and TetA is concerned, it appears to be similar to the average behavior. Regarding the Tc amount in the single cell (Figure 3.7b), again after the administration of Tc we observe a peak value of approximately 117 molecules. Subsequently, the single cell amount decreases and finally fluctuates around the value of about 9 molecules. Overall, the Tc amount at steady state varies from 0 to 36 molecules among the cells (data not shown). Similarly to the average behavior, the single cell behavior with respect to TetA molecules shows an initial increase after Tc administration. This increase results in fluctuations around the value of 500 TetA molecules. At steady state, the TetA amount observed within the single cells varies between 197 and 1,399 molecules (data not shown).

As discussed previously, only a small portion (approximately 5%) of the total mRNA from the *tetR* gene is transcribed from the $tetP_{R1}$ promoter [103]. The remaining originates from the $tetP_{R2}$ promoter. In order to explore the functionality of the $tetP_{R1}$ promoter, we performed a set of simulations in which no $tetP_{R1}$ exists in the system. Initially, the system is at steady state and consequently there is neither Tc nor TetA in the cell. At time equal to 5 hours, 400 Tc molecules are administered to the wild type as well as to the cell which lacks the promoter $tetP_{R1}$. The trends of the two systems are provided in Figure 3.8.

Importantly, the results of the simulation of the wild type system are the same with the results of the system which lacks the promoter $tetP_{R1}$. The lack of $tetP_{R1}$ from the *tet* operon does not appear to affect the amount of the regulatory molecules, TetR, TetR2 and TetA, in the cell (Figures 3.8a,3.8b,3.8c). In addition, the number of the intracellular Tc molecules remains the same (Figure 3.8d). We further observed that the total behavior of the system is retained even when the *tet* operon includes no $tetP_{R1}$ promoter (data not shown). This indicates that this promoter is not essential in *E. coli*. It supports the existing hypothesis that this promoter is involved in the *tet* operon because of its functional importance in other bacteria with the *tet* operon [103]. This promoter could also have previously played a role in *E. coli*, but may remain only as a genetic artifact.

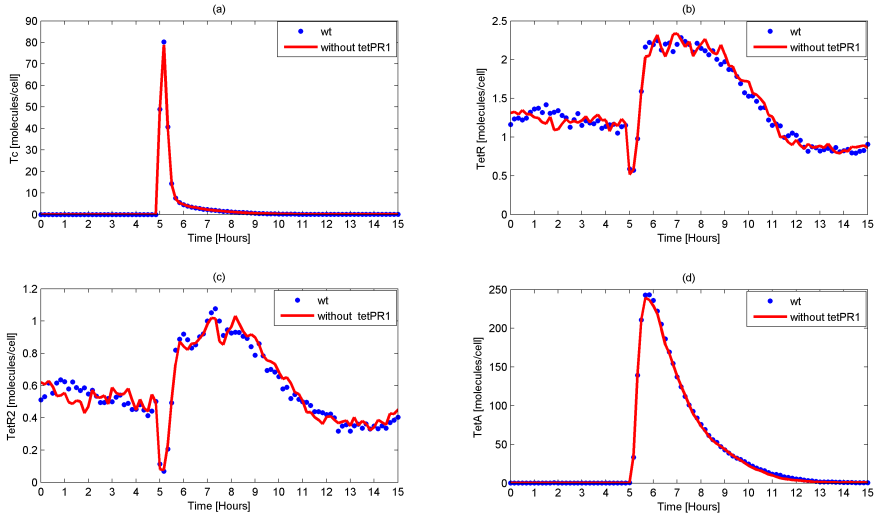

Figure 3.8: Average number of Tc (Figure 3.8a), TetR (Figure 3.8b), TetR2 (Figure 3.8c) and TetA (Figure 3.8d) molecules of the wild type (wt) *tet* operon as well as of the *tet* operon which lacks the promoter $tetP_{R1}$. At time of 5 hours, a pulse of 400 Tc molecules is administered to each cell. The behavior of the system appears to be the same with and without the promoter $tetP_{R1}$.

3.2.2 Sensitivity analysis

The *tet* operon is an excellent biological switch [85]. The key features that define its outstanding functionality include high TetA expression in the presence of Tc combined with no appreciable expression leakiness in the absence of Tc. These salient characteristics are attributed to the high affinity of Tc for the repressor TetR2 and the high affinity of TetR2 for the operator sites, *tetO1* and *tetO2*. Even though these parameters and the associated processes are not rate limiting steps, they do contribute to the fine tuning of this biological switch.

In order to explore the importance of these features, we performed a sensitivity analysis of our model to examine how the different values of the specific *tet* operon parameters influence its behavior. In this set of simulations, a pulse of a wide range (20 to 800 molecules) of Tc molecules is administered to each *E. coli* cell. With this range, we can investigate the sensitivity of the *tet* operon even when minimal amounts of Tc are administered. We can also understand the changes in the behavior of this operon

upon increasing the amount of administered Tc.

Affinity of Tc for TetR2

First, the importance of Tc's affinity for TetR2 is investigated. As discussed previously, once Tc diffuses into the cell, it binds to TetR2 with high binding strength ($K_{eq} \simeq 3 \cdot 10^9 M^{-1}$) [91]. This causes rapid induction of the system, and consequently the expression of *tetR* and *tetA*. In order to check how this binding strength influences the total behavior of the system, a set of simulations was conducted in which the affinity of Tc for TetR and TetR2 spans two orders of magnitude lower and higher than the nominal value. This could be achieved experimentally by making amino acid substitutions or deletions on TetR2 [113], thereby affecting the affinity of Tc for TetR2. In our analysis, this is realized by changing the kinetic constants that correspond to this affinity ($k_{10}, k_{12}, k_{14}, k_{16}, k_{20}, k_{22}$) 10 and 100 times.

Results of the simulations are depicted in Figure 3.9. They illustrate the maximum intracellular TetA (Figure 3.9a) and Tc (Figure 3.9b) amount that is produced when a pulse of 20, 50, 100, 200, 400, 600, 800, and 1,000 Tc molecules is administered to each cell. The various TetA levels are the result of the range of administered Tc molecules and of different Tc affinities for TetR2.

As depicted in Figure 3.9, the system is sensitive to changes in the affinity of Tc for TetR2. The maximum number of TetA molecules that are produced in the presence of Tc increases as the wild type affinity of Tc for TetR2 increases (Figure 3.9a). As expected, this has an effect on the maximum number of intracellular Tc molecules. This effect is shown in Figure 3.9b where the maximum number of intracellular Tc molecules decreases as the wild type affinity of Tc for TetR2 increases. This can be ascribed to the fact that when the affinity of Tc for TetR2 increases, the formation rate of the complex TetR2:Tc2 increases, thereby accelerating gene induction. Therefore, more TetA molecules are generated. As a consequence, Tc is excluded faster from the cell, making intracellular Tc level lower. On the other hand, a decrease in the wild type affinity of Tc for TetR2 leads to a decrease in the intracellular TetA amount (Figure 3.9a). This decrease elicits a drastic increase on the intracellular Tc amounts (Figure 3.9b).

It should be kept in mind that the lower the amount of intracellular Tc, the higher

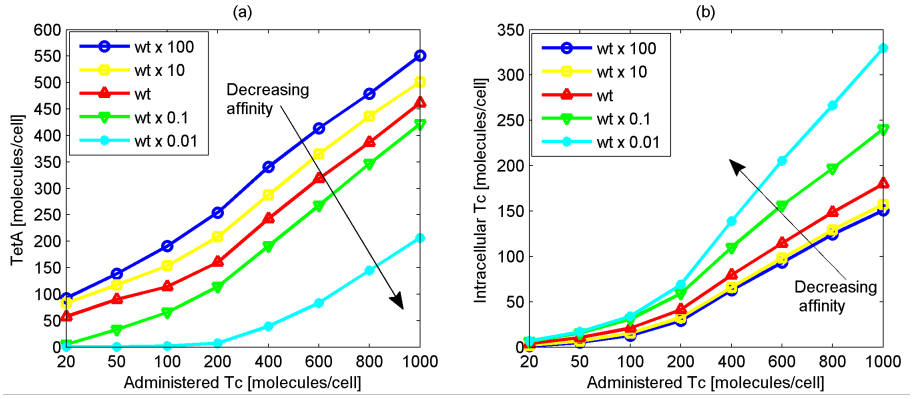

Figure 3.9: Average maximum number of TetA (Figure 3.9a) and Tc (Figure 3.9b) molecules for a range of administered Tc in the wild type (wt) system, as well as in systems where the affinity of Tc for TetR2 is 10 and 100 times lower and 10 and 100 times higher than the nominal value. An increase in the affinity of Tc for TetR2 results in the production of more TetA and less Tc present in the cell.

the probability for the cell to survive. This implies that an increase in the affinity of Tc for TetR2, could help the *E. coli* to survive. On the other hand, a decrease in the affinity of Tc for TetR2 could lead to the faster death of *E.coli*.

Figure 3.9 indicates that the larger the amount of administered Tc, the larger the differences in the amount of intracellular Tc between the eight cases (cases with different affinities). Thus, if large Tc amounts are administered, then such changes in the affinity of Tc for TetR2 will cause significantly large differences in the intracellular Tc amount among these eight cases. Again, this implies that a decrease in the affinity of Tc for TetR2 could result in high intracellular amounts of Tc in the cell, causing cell death.

Another interesting observation is that a large decrease (wt x 0.01) in the affinity of Tc for TetR2, makes the *tet* operon less sensitive to external Tc. When the affinity is 100 times lower than the nominal value, higher Tc amounts must be administered to induce gene expression. In particular, the number of administered Tc molecules must be 200 or higher before expression of *tetA* is turned on (Figure 3.9a). This is also evident when the affinity is 10 times smaller than the wild type. Then, more than 20 Tc molecules must be administered for the expression of TetA to occur. On the other hand, in the other three cases (wt, wt x 10, wt x 100), gene expression is activated with

the administration of only 20 Tc molecules. This phenomenon is observed because the higher the affinity of Tc for the repressor, the easier and faster the binding of Tc to TetR2, and consequently the induction of the *tet* operon.

For small numbers of administered Tc molecules (20, 50, 100 molecules), the difference between the eight cases is not large concerning the maximum number of intracellular Tc amount. However, regarding TetA, the difference between the eight cases is larger even for small numbers of administered Tc molecules. The high TetA amounts combined with low intracellular Tc amounts could have harmful effects for the cell since high excess of TetA in the cell could lead to cell membrane collapse. Furthermore, we observe that when the affinity is larger than the nominal value (wt x 10 and wt x 100), little difference in the amount of intracellular Tc is noticed. On the other hand, a substantial difference in the amount of intracellular Tc is observed in the two cases where the affinity is smaller than the nominal value (wt x 0.1, wt x 0.01). This implies that the system is more sensitive to a decrease in the affinity of Tc for the repressor than to an increase. This probably stems from the fact that the affinity of Tc for TetR2 is already high and not rate limiting, thus higher values do not cause dramatic changes. Furthermore, this indicates that the wild type network affinity is optimal since it is as high as possible to reduce Tc while not increasing TetA significantly.

Lastly, it is worth stressing that our changes in the affinity of Tc for TetR and TetR2 do not affect the average number of TetR and TetR2 molecules in the cell when Tc is administered (data not shown). The amount of these two molecules in the cell remains practically the same even upon applying the aforementioned changes.

Affinity of TetR2 for the operator sites

Here, we explore the significance of the high binding strength ($K_{eq} \simeq 10^{12} M^{-1}$)[95] of TetR2 to the operator sites, *tetO1* and *tetO2*. This feature enables the repressor to bind quickly and tightly to the operator sites in the absence of Tc, making this operon a thoroughly tuned biological switch. The importance of this feature is investigated through a set of simulations in which the binding strength of the repressor for the operator sites is 10, 50 and 100 times lower than the wild type value. This could be experimentally achieved by either introducing amino acid changes on the protein TetR or mutating the operator sites [114]. In our simulations, this is attained by decreasing

all the corresponding kinetic parameters ($k_3, k_5, k_{19}, k_{25}, k_{64}, k_{65}, k_{66}, k_{67}$) 10, 50 and 100 times. We decrease the rate that TetR2 binds to the operators when it is either free (k_3, k_5, k_{64}, k_{65}) or bound on Tc ($k_{19}, k_{25}, k_{66}, k_{67}$), causing a decrease to the equilibrium constant. In contrast to the previous case, in which the affinity of Tc for the repressor is investigated, we do not consider an increase to the equilibrium constant. The equilibrium constant value is already very high and a possible increase would not be experimentally meaningful due to diffusion limitations. The results of the simulations are shown in Figure 3.10 and they represent the maximum amounts of TetA and Tc in the cells.

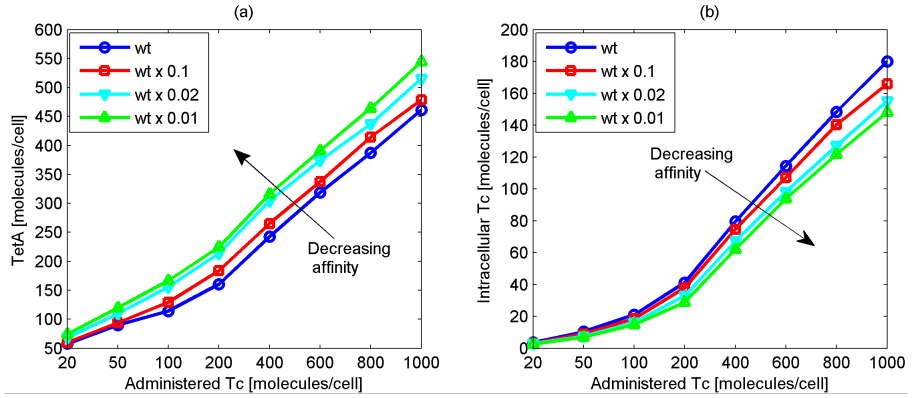

Figure 3.10: Average maximum number of TetA (Figure 3.10a) and Tc (Figure 3.10b) molecules for a range of administered Tc in the wild type (wt) system, as well as in systems where the affinity of TetR2 for the operators is 10, 50 and 100 times lower than the nominal value. Decrease in the affinity of TetR2 for the operator sites results in an increase of the generated TetA and a decrease in the intracellular Tc amount.

According to the results, the fidelity of the *tet* operon appears to be susceptible to changes in the affinity of TetR2 for the operators. A decrease in the affinity of the repressor for the operator sites results in an increase in the amount of generated TetA (Figure 3.10a). Thus, lower affinities give rise to higher TetA production. Higher TetA amount leads to faster removal of Tc from the cell thereby leading to lower intracellular Tc amounts (Figure 3.10b). The decreased amount of intracellular Tc is beneficial for the cell. However, the decrease in the affinity of the repressor for the operators results also in an increase in the TetA, TetR and TetR2 amount at steady state. This causes

the *tet* operon to loose its unique function. In other words, the trade off in having Tc removed fast from the cell is the high intracellular TetA, TetR and TetR2 amounts at steady state.

As remarked previously, *tet* operon is a unique biological switch [85]. This is attributed to the optimal design of high affinity levels which accommodates 1) small number of TetR2 molecules at steady state, which allows gene induction to occur with minimal Tc amounts, and 2) tight repression of *tetA* gene, which does not allow the production of TetA protein in the absence of Tc. Therefore, at steady state, the intracellular amount of TetR and TetR2 is small (with average values 2 and 1 molecules respectively) and the amount of TetA is zero. However, here we observe that when the affinity of TetR2 for the operator sites decreases, the amount of TetA, TetR and TetR2 increases.

In particular, when the affinity is 10, 50 and 100 times lower than the wild type value, the number of intracellular TetA molecules at steady state is 3, 7, and 10 respectively (data not shown). The lower the affinity of TetR2 for the operator sites, the higher the TetA amount at steady state. Thus, the decrease in the affinity of TetR2 for the operators results in non-zero TetA amounts at steady state. This in turn could cause cell death because TetA protein, in the absence of Tc, is toxic for the cell [110].

Additionally, when the affinity of TetR2 for the operators is 10, 50 and 100 times lower than the nominal value, the average number of intracellular TetR and TetR2 molecules is 3, 4, 5 and 2, 3, 5 respectively (data not shown). Given that the TetR and TetR2 amount in each single cell experiences large fluctuations, the higher the mean value, the higher the fluctuations in the TetR and TetR2 amount. Furthermore, the higher the intracellular TetR and TetR2 amount, the higher the required Tc amount for inducing gene expression. This happens because the administered Tc is occupied by the free TetR and TetR2 molecules and therefore cannot bind to the TetR2 bound on the operators to activate expression. Therefore, the large intracellular TetR and TetR2 amounts require large Tc amounts to induce gene expression thereby eliminating the functionality of the *tet* operon. Thus, the aforestated changes in the amount of TetR, TetR2 and TetA at steady state are detrimental to the *E. coli*. This confirms that *tet* operon is a very well tuned system and possible changes in its crucial parameters could be harmful for the *E. coli*.

It is noticeable that the differences in Tc and TetA amounts that are caused by changing the affinity of Tc for TetR2 are larger than the differences that are caused by modifying the affinity of TetR2 for the operator sites. This indicates that the *tet* operon is more sensitive to changes in the affinity of Tc for TetR2 than in the affinity of TetR2 for the operators.

3.3 Conclusions

We have developed a detailed mathematical model for the *tet* operon and performed stochastic simulations to examine the mechanisms that govern the dynamics of this interesting biological system. Conducting stochastic simulations, we investigated the average behavior of 1,000 cells, the diversification across the cells, and finally the single cell behavior. The results of the simulations are in agreement with, and explain well numerous experimental observations such as tight repression, fast gene expression, induction with small Tc amounts, and small intracellular TetR2 amounts.

Our simulations demonstrate that there is a nearly linear relationship between the administered Tc and the TetA amount. Furthermore, the results indicate large fluctuations in the amount of the repressor protein TetR2. Additionally, our findings highlight that the behavior of the *tet* operon is the same even when it lacks the promoter $tetP_{R1}$. This could imply that this promoter is redundant and not necessary in *E. coli*, although it may be functionally important in other bacteria. It could also indicate that although *E. coli* used to need this promoter, they do not need it anymore and it just exists in their genome.

Sensitivity analysis illustrates that the affinity of Tc for the repressor TetR2 and the affinity of TetR2 for the operator sites have a high impact on the behavior of the *tet* operon, suggesting optimum interaction strengths developed through natural selection.

An increase in the affinity of Tc for the repressor leads to an increase in the production of TetA protein. Increased amounts of TetA in the cell remove Tc from the cell faster, thereby keeping the levels of the intracellular Tc low.

A decrease in the affinity of the repressor TetR2 for the operators results in the production of more TetA protein upon Tc administration. Additionally, it results in an increase in the number of TetR, TetR2 and TetA molecules at steady state. This causes

a decrease in the number of intracellular Tc molecules, increasing the probability for the cell to survive. However, the existence of intracellular TetA at steady state may lead to cell death.

The need for a mechanistic understanding of bacterial resistance to Tc was identified many years ago [115, 112]. Computer simulations of the *tet* operon provide a comprehensive understanding of the interactions between the *tet* operon molecular elements. They also provide valuable information that may contribute to the design of prototype synthetic gene regulatory networks.

Table 3.1: Reaction network of the *tet* operon.

Number	Reaction
1	$2TetR \underset{k_2}{\overset{k_1}{\rightleftharpoons}} TetR2$
2	$TetR2 + tetO1 \underset{k_4}{\overset{k_3}{\rightleftharpoons}} TetR2{:}tetO1$
3	$TetR2 + tetO2 \underset{k_6}{\overset{k_5}{\rightleftharpoons}} TetR2{:}tetO2$
4	$TetR2 + nsDNA \underset{k_8}{\overset{k_7}{\rightleftharpoons}} TetR2{:}nsDNA$
5	$TcEx \xrightarrow{k_9} Tc$
6	$TetR2 + Tc \underset{k_{11}}{\overset{k_{10}}{\rightleftharpoons}} TetR2{:}Tc$
7	$TetR2{:}Tc + Tc \underset{k_{13}}{\overset{k_{12}}{\rightleftharpoons}} TetR2{:}Tc2$
8	$TetR2{:}tetO1 + Tc \underset{k_{15}}{\overset{k_{14}}{\rightleftharpoons}} TetR2{:}tetO1{:}Tc$
9	$TetR2{:}tetO1{:}Tc + Tc \underset{k_{17}}{\overset{k_{16}}{\rightleftharpoons}} TetR2{:}tetO1{:}Tc2$
10	$TetR2{:}tetO1{:}Tc2 \underset{k_{19}}{\overset{k_{18}}{\rightleftharpoons}} TetR2{:}Tc2 + tetO1$
11	$TetR2{:}tetO2 + Tc \underset{k_{21}}{\overset{k_{20}}{\rightleftharpoons}} TetR2{:}tetO2{:}Tc$
12	$TetR2{:}tetO2{:}Tc + Tc \underset{k_{23}}{\overset{k_{22}}{\rightleftharpoons}} TetR2{:}tetO2{:}Tc2$
13	$TetR2{:}tetO2{:}Tc2 \underset{k_{25}}{\overset{k_{24}}{\rightleftharpoons}} TetR2{:}Tc2 + tetO2$
14	$RNAp + Complex \underset{k_{27}}{\overset{k_{26}}{\rightleftharpoons}} RNAp{:}[tetP_{R1}]{:}tetP_{R2}{:}tetP_A{:}tetO1{:}tetO2$
15	$RNAp{:}[tetP_{R1}]{:}tetP_{R2}{:}tetP_A{:}tetO1{:}tetO2 \xrightarrow{k_{28}} RNAp^*{:}[tetP_{R1}]{:}tetP_{R2}{:}tetP_A{:}tetO1{:}tetO2$
16	$RNAp^*{:}[tetP_{R1}]{:}tetP_{R2}{:}tetP_A{:}tetO1{:}tetO2 \xrightarrow{k_{29}} RNAp^*{:}DNA(P_{R1}) + Complex$
17	$RNAp^*{:}DNA(P_{R1}) \xrightarrow[k_{31}]{k_{30}} RNAp + mRNA(tetR)$
18	$RNAp + Complex \underset{k_{33}}{\overset{k_{32}}{\rightleftharpoons}} RNAp{:}tetP_{R1}{:}[tetP_{R2}]{:}tetP_A{:}tetO1{:}tetO2$
19	$RNAp{:}tetP_{R1}{:}[tetP_{R2}]{:}tetP_A{:}tetO1{:}tetO2 \xrightarrow{k_{34}} RNAp^*{:}tetP_{R1}{:}[tetP_{R2}]{:}tetP_A{:}tetO1{:}tetO2$
20	$RNAp^*{:}tetP_{R1}{:}[tetP_{R2}]{:}tetP_A{:}tetO1{:}tetO2 \xrightarrow{k_{35}} RNAp^*{:}DNA(P_{R2}) + Complex$
21	$RNAp^*{:}DNA(P_{R2}) \xrightarrow[k_{37}]{k_{36}} RNAp + mRNA(tetR)$
22	$RNAp + tetP_{R2} + tetO1 + TetR2{:}tetO2 \underset{k_{39}}{\overset{k_{38}}{\rightleftharpoons}} RNAp{:}[tetP_{R2}]{:}tetO1{:}TetR2{:}tetO2$

Number	Reaction
23	$RNAp : [tetP_{R2}]{:}tetO1{:}TetR2{:}tetO2 \xrightarrow{k_{40}} RNAp^* {:} [tetP_{R2}]{:}tetO1{:}TetR2{:}tetO2$
24	$RNAp^* {:} [tetP_{R2}]{:}tetO1{:}TetR2{:}tetO2 \xrightarrow{k_{41}} RNAp^* {:} DNA(P_{R2}) + tetP_{R2} + tetO1 + TetR2{:}tetO2$
25	$Rib + mRNA(tetR) \xrightarrow{k_{42}} Rib{:}mRNA(tetR)$
26	$Rib{:}mRNA(tetR) \xrightarrow{k_{43}} Rib^* {:} mRNA(tetR) + mRNA(tetR)$
27	$Rib^* {:} mRNA(tetR) \underset{k_{45}}{\xrightarrow{k_{44}}} Rib + tetR$
28	$RNAp + Complex \underset{k_{47}}{\overset{k_{46}}{\rightleftharpoons}} RNAp{:}tetP_{R1}{:}tetP_{R2}{:}[tetP_A]{:}tetO1{:}tetO2$
29	$RNAp{:}tetP_{R1}{:}tetP_{R2}{:}[tetP_A]{:}tetO1{:}tetO2 \xrightarrow{k_{48}} RNAp*{:}tetP_{R1}{:}tetP_{R2}{:}[tetP_A]{:}tetO1{:}tetO2$
30	$RNAp^* {:} tetP_{R1}{:}tetP_{R2}{:}tetP_A{:}tetO1{:}tetO2 \xrightarrow{k_{49}} RNAp^* {:} DNA(P_A) + Complex$
31	$RNAp^* {:} DNA(P_A) \underset{k_{51}}{\xrightarrow{k_{50}}} RNAp + mRNA(tetA)$
32	$Rib + mRNA(tetA) \xrightarrow{k_{52}} Rib{:}mRNA(tetA)$
33	$Rib{:}mRNA(tetA) \xrightarrow{k_{53}} Rib^* {:} mRNA(tetA) + mRNA(tetA)$
34	$Rib^* {:} mRNA(tetA) \underset{k_{55}}{\xrightarrow{k_{54}}} Rib + TetA$
35	$TetR2 \xrightarrow{k_{56}} \emptyset$
36	$TetR2{:}Tc2 \xrightarrow{k_{57}} 2Tc$
37	$mRNA(tetR) \xrightarrow{k_{58}} \emptyset$
38	$mRNA(tetA) \xrightarrow{k_{59}} \emptyset$
39	$Tc \xrightarrow{k_{60}} \emptyset$
40	$Tc + TetA \underset{k_{62}}{\xrightarrow{k_{61}}} TetA$
41	$tetP_{R1} + tetP_{R2} + tetP_A + tetO1 + tetO2 \xrightarrow{k_{63}} Complex$
42	$Complex + TetR2 \xrightarrow{k_{64}} TetR2{:}tetO1 + tetP_{R1} + tetP_{R2} + tetP_A + tetO2$
43	$Complex + TetR2 \xrightarrow{k_{65}} TetR2{:}tetO2 + tetP_{R1} + tetP_{R2} + tetP_A + tetO1$
44	$Complex + TetR2{:}Tc2 \xrightarrow{k_{66}} TetR2{:}tetO1{:}Tc2 + tetP_{R1} + tetP_{R2} + tetP_A + tetO2$
45	$Complex + TetR2{:}Tc2 \xrightarrow{k_{67}} TetR2{:}tetO2{:}Tc2 + tetP_{R1} + tetP_{R2} + tetP_A + tetO1$

Table 3.2: Definition of the species that participate in the *tet* operon reaction network.

Species	Definition
TcEx	External Tc
Tc	Intracellular Tc
TetR	Tc repressor protein
TetR2	Tc repressor protein (dimer)
TetA	Tc transport protein
tetO1	Operator site 1
tetO2	Operator site 2
nsDNA	Non-specific DNA sites
$tetP_{R1}$	Promoter 1 of *tetR* gene
$tetP_{R2}$	Promoter 2 of *tetR* gene
$tetP_A$	Promoter of *tetA* gene
RNAp	RNA polymerase
Complex	Free operator and promoter sites
mRNA(*tetA*)	*tetA* gene mRNA
mRNA(*tetR*)	*tetR* gene mRNA
Rib	Ribosome
$Rib{:}mRNA(tetR)$	Ribosome bound to mRNA(*tetR*)
$Rib{:}mRNA(tetA)$	Ribosome bound to mRNA(*tetA*)
$TetR2{:}tetO1$	TetR2 bound to *tetO1*
$TetR2{:}tetO2$	TetR2 bound to *tetO2*
$TetR2{:}nsDNA$	TetR2 bound to nsDNA
$TetR2{:}Tc$	TetR2 bound to one Tc molecule
$TetR2{:}Tc2$	TetR2 bound to two Tc molecules
$TetR2{:}tetO1{:}Tc$	TetR2 bound to *tetO1* and one Tc molecule
$TetR2{:}tetO1{:}Tc2$	TetR2 bound to *tetO1* and two Tc molecules
$TetR2{:}tetO2{:}Tc$	TetR2 bound to *tetO2* and one Tc molecule
$TetR2{:}tetO2{:}Tc2$	TetR2 bound to *tetO2* and two Tc molecules
$RNAp{:}[tetP_{R1}]{:}tetP_{R2}{:}tetP_A{:}tetO1{:}tetO2$	Closed complex of RNAp bound to $tetP_{R1}$
$RNAp^*{:}[tetP_{R1}]{:}tetP_{R2}{:}tetP_A{:}tetO1{:}tetO2$	Open complex of RNAp bound to $tetP_{R1}$
$RNAp{:}tetP_{R1}{:}[tetP_{R2}]{:}tetP_A{:}tetO1{:}tetO2$	Closed complex of RNAp bound to $tetP_{R2}$
$RNAp^*{:}tetP_{R1}{:}[tetP_{R2}]{:}tetP_A{:}tetO1{:}tetO2$	Open complex of RNAp bound to $tetP_{R2}$
$RNAp{:}[tetP_{R2}]{:}tetO1{:}TetR2{:}tetO2$	Closed complex of RNAp bound to $tetP_{R2}$
$RNAp^*{:}[tetP_{R2}]{:}tetO1{:}tetR2{:}tetO2$	Open complex of RNAp bound to $tetP_{R2}$
$RNAp{:}tetP_{R1}{:}tetP_{R2}{:}[tetP_A]{:}tetO1{:}tetO2$	Closed complex of RNAp bound to $tetP_A$
$RNAp*{:}tetP_{R1}{:}tetP_{R2}{:}[tetP_A]{:}tetO1{:}tetO2$	Open complex of RNAp bound to $tetP_A$
$Rib^*{:}mRNA(tetA)$	Rib bound to the first codon of the mRNA(*tetA*)
$Rib^*{:}mRNA(tetR)$	Rib bound to the first codon of the mRNA(*tetR*)
$RNAp^*{:}DNA(P_A)$	RNAp bound to *tetA* gene
$RNAp^*{:}DNA(P_{R1})$	RNAp bound to *tetR* gene starting transcription from $tetP_{R1}$
$RNAp^*{:}DNA(P_{R2})$	RNAp bound to *tetR* gene starting transcription from $tetP_{R2}$

Table 3.3: Kinetic constants of the *tet* operon reaction network.

kinetic constant	Ref.	kinetic constant	Ref.	kinetic constant	Ref.
$k_1 = 10^{+9}$	[95][h]	$k_{24} = 5.80 \cdot 10^{-3}$	[95]	$k_{47} = 0.10$	[116][h]
$k_2 = 10$	[95][h]	$k_{25} = 0.10$	[91]	$k_{48} = 0.013$	[116]
$k_3 = 10^{+8}$	[95][h]	$k_{26} = 4.8 \cdot 10^{+4}$	[103, 116][h]	$k_{49} = 30$	[106]
$k_4 = 10^{-4}$	[95][h]	$k_{27} = 0.10$	[116][h]	$k_{50} = 30$	[106][a]
$k_5 = 5.00 \cdot 10^{+8}$	[87, 95][h]	$k_{28} = 0.013$	[116]	$k_{51} = 1182$	[117][b]
$k_6 = 10^{-4}$	[95][h]	$k_{29} = 30$	[106]	$k_{52} = 10^{+5}$	e
$k_7 = 30$	[95][h]	$k_{30} = 30$	[106][a]	$k_{53} = 100$	[106]
$k_8 = 0.10$	[95][h]	$k_{31} = 621$	[80][b]	$k_{54} = 100$	[106][a]
$k_9 = 3.30 \cdot 10^{-4}$	[118]	$k_{32} = 9.10 \cdot 10^{+5}$	[103, 116][h]	$k_{55} = 394$	[117][b]
$k_{10} = 4.80 \cdot 10^{+5}$	[91]	$k_{33} = 0.10$	[116][h]	$k_{56} = 3.85 \cdot 10^{-5}$	f
$k_{11} = 10^{-4}$	[91]	$k_{34} = 0.013$	[116]	$k_{57} = 3.85 \cdot 10^{-5}$	f
$k_{12} = 2.40 \cdot 10^{+5}$	[91]	$k_{35} = 30$	[106]	$k_{58} = 0.002$	e
$k_{13} = 2.00 \cdot 10^{-4}$	[91]	$k_{36} = 30$	[106] [a]	$k_{59} = 0.002$	e
$k_{14} = 4.80 \cdot 10^{+5}$	[91]	$k_{37} = 621$	[80][b]	$k_{60} = 2.67 \cdot 10^{-6}$	[105]
$k_{15} = 10^{-4}$	[91]	$k_{38} = 7.23 \cdot 10^{+6}$	[103, 116][h]	$k_{61} = 1$	[84][c]
$k_{16} = 2.40 \cdot 10^{+5}$	[91]	$k_{39} = 0.10$	[116][h]	$k_{62} = 4.90 \cdot 10^{-5}$	[92][d]
$k_{17} = 2.00 \cdot 10^{-4}$	[91]	$k_{40} = 0.013$	[116]	$k_{63} = 10^{+8}$	g
$k_{18} = 5.80 \cdot 10^{-3}$	[95]	$k_{41} = 30$	[106]	$k_{64} = 10^{+8}$	[95]
$k_{19} = 0.10$	[91]	$k_{42} = 10^{+5}$	e	$k_{65} = 5.00 \cdot 10^{+8}$	[87, 95]
$k_{20} = 4.80 \cdot 10^{+5}$	[91]	$k_{43} = 100$	[106]	$k_{66} = 0.10$	[91]
$k_{21} = 10^{-4}$	[91]	$k_{44} = 100$	[106][a]	$k_{67} = 0.10$	[91]
$k_{22} = 2.40 \cdot 10^{+5}$	[91]	$k_{45} = 207$	[80][b]		
$k_{23} = 2.00 \cdot 10^{-4}$	[91]	$k_{46} = 8.60 \cdot 10^{+6}$	[116, 90][h]		

Units on k: first order reactions: sec^{-1}, second order reactions: $M^{-1}sec^{-1}$, third order reactions: $M^{-2}sec^{-1}$, fourth order reactions: $M^{-3}sec^{-1}$, fifth order reactions: $M^{-4}sec^{-1}$.

[a] Scale parameter of the gamma distributed event (sec).

[b] Shape parameter of the gamma distributed event.

[c] k_{cat} of Michaelis Menten kinetics.

[d] K_M of Michaelis Menten kinetics.

[e] Rates adjusted to give 20 protein molecules per mRNA transcript.

[f] Rates adjusted for 5 hours half life.

[g] Rate adjusted for very fast reaction.

[h] Rate calculated through equilibrium constant by assuming the forward and calculating the reverse rate.

Chapter 4

Attenuated bistability in bacterial sugar utilization

4.1 Results and discussion

Bacteria commonly encode inducible pathways dedicated to the uptake and catabolism of individual nutrients (Figure 4.1). Each pathway is composed of at least one transporter protein that imports the nutrient into the cytoplasm and at least one catabolic enzyme that shunts the nutrient into central or secondary metabolism. Each pathway also consists of at least one sensory regulator that induces transporter and enzyme expression in the presence of the nutrient or a metabolic intermediate. As one example, the model bacterium *E. coli* encodes over 30 inducible pathways that are dedicated to distinct nutrients and possess similar pathway structures.

Inducible utilization pathways traditionally have been studied with bulk approaches that measure the average behavior across a population of cells [119]. However, single-cell studies of the lactose and L-arabinose utilization pathways have revealed phenotypic diversity, where the pathways were uninduced or fully induced in the presence of the inducer [120, 121, 122, 123]. This "all-or-none" response was attributed to positive feedback, where the inducer upregulates transporter expression, leading to the import of additional inducer. Critically, these studies removed catabolism of the inducer, either through the use of a non-hydrolyzable sugar analog or through the deletion of the catabolic genes. Catabolism otherwise imparts negative feedback, where the nutrient

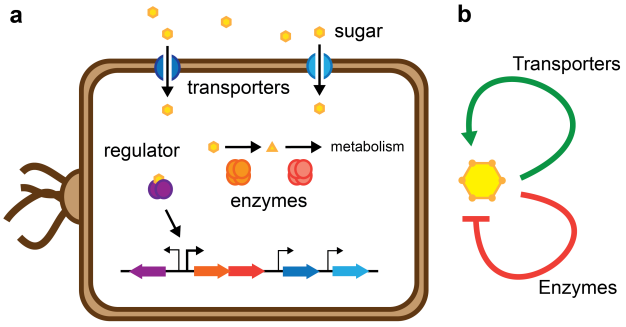

Figure 4.1: Sugar utilization in bacteria: a, The majority of sugar utilization pathways are composed of a set of transporters and catabolic enzymes that are induced in the presence of the sugar or a metabolic intermediate. b, Combination of positive and negative feedback through induction of the transporters and enzymes by the sugar.

induces the expression of the pathway enzymes that break down the nutrient. Coupling positive and negative feedback has been shown to generate diverse behaviors such as oscillations or excitability [124, 125]. Correspondingly, the few single-cell studies of the response to lactose and L-arabinose without disabling catabolism suggest varying responses [121, 126, 127, 128]. This diversity stands in contrast to theoretical studies which concluded that natural pathways only exhibit graded responses to the nutrient [129, 130].

Here, we focused on utilization pathways in *E. coli* K-12 because most of the pathway components have been identified and well characterized. To monitor the single-cell response in each pathway, we placed the green fluorescent protein gene (*gfp*) under the control of a pathway promoter on a plasmid with the low-copy pSC101 origin of replication [131]. MG1655 cells transformed with each plasmid were grown in minimal medium with glycerol and varying concentrations of the nutrient before being subjected to flow cytometry analysis. The cells were grown for 20 hours to a low cell density to ensure that the fluorescence distributions approached steady-state and to avoid depletion of the nutrient in the medium (Figure 4.2).

Remarkably, the pathways exhibited a variety of responses (Figure 4.3). The L-arabinose pathway exhibited a combination of graded and "all-or-none" responses, the

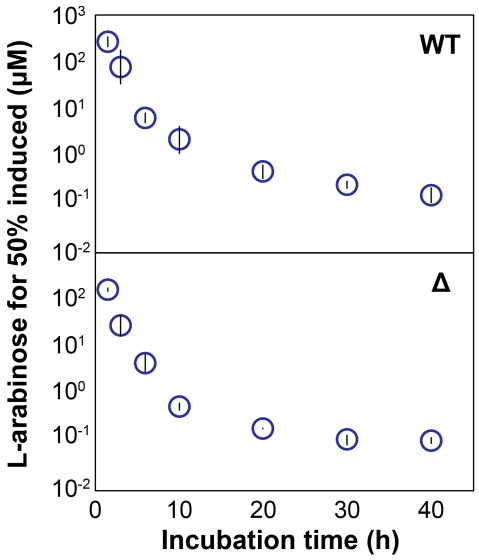

Figure 4.2: Extended incubation times and lower cell densities ensure stable distributions. Concentration of L-arabinose necessary to induce half of the population for wild type (WT) and Δ*araBAD* (Δ) cells following different incubation times. An incubation time of 20 hours was selected as the shortest time to approach a roughly stable distribution. Data points represent the geometric mean and standard deviation of at least three independent experiments.

D-xylose and L-rhamnose pathways exhibited an "all-or-none" response, and the D-galactose, lactose, and D-gluconate pathways exhibited a graded response. The diversity of responses was noteworthy considering the common structure between pathways. Two of the pathways do deviate from this structure-the D-galactose pathway is amphibolic whereas the lactose pathway is induced by the metabolic intermediate allolactose, potentially explaining the observed behavior.

To further investigate these responses, we concentrated on the L-arabinose utilization pathway (Figure 4.4), one of the best-studied utilization pathways in biology [132]. As described above, the "all-or-none" response has been ascribed to this pathway, but in the absence of L-arabinose catabolism [123]. To explore the contribution of the catabolism of L-arabinose, we deleted the L-arabinose catabolic operon araBAD (Δ*araBAD*) and repeated the flow cytometry analysis. We found that Δ*araBAD* cells were uninduced or fully induced, paralleling previous observations (Figure 4.5a). Equivalent responses

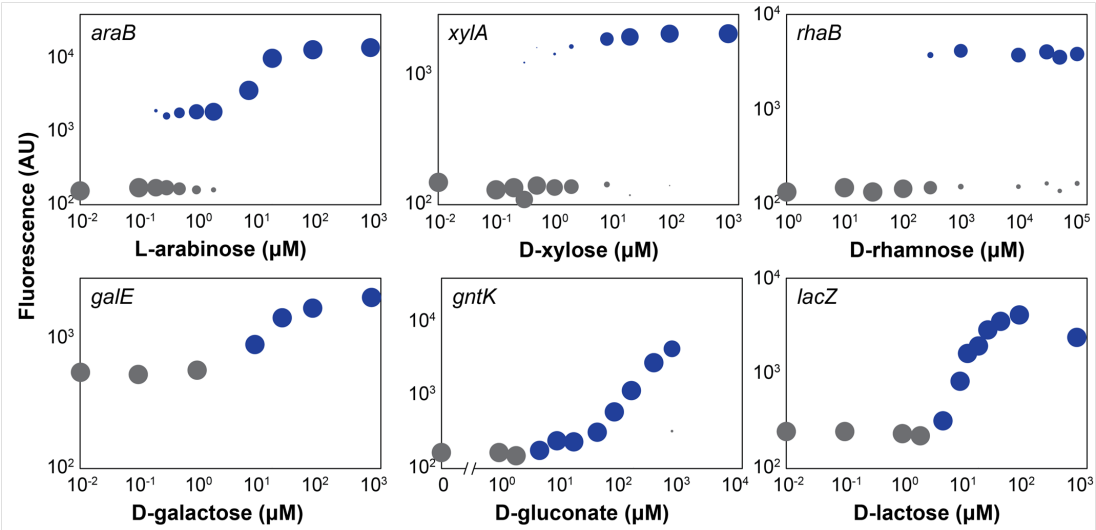

Figure 4.3: Varying responses to different sugars in *E. coli*. MG1655 cells harboring a transcriptional reporter plasmid were grown for 20 hours in M9 glycerol with different concentrations of the indicated sugar and then subjected to flow cytometry analysis. Each dot represents the mean fluorescence and the relative number of cells in the induced (blue) and uninduced (grey) subpopulations. The diameter of each dot scales with the relative number. Each dot plot is representative of at least three independent experiments.

were observed for three other L-arabinose promoters ($ParaE$, $ParaF$, $ParaJ$) in WT and in $\Delta araBAD$ cells (Figure 4.6a-c), demonstrating that catabolism of L-arabinose is necessary for the combination of graded and "all-or-none" responses to L-arabinose. Importantly, the altered response cannot be attributed to self-catabolite repression[133], as the addition of 2 mM cAMP in the medium did not have a qualitative impact on the response curves (Figure 4.6d).

A standard feature of the "all-or-none" response and bistability is hysteresis, where pre-induced systems exhibit a more sensitive response. To measure the extent of hysteresis, we supplemented overnight cultures with no L-arabinose (-ara) or sufficient L-arabinose to induce the entire population (+ara). The L-arabinose then was washed away and the cells were incubated for 20 hours in different concentrations of L-arabinose. We found that $\Delta araBAD$ cells exhibited hysteresis typical of the "all-or-none" response [120, 121], where pre-induced cells were four times more sensitive to L-arabinose than

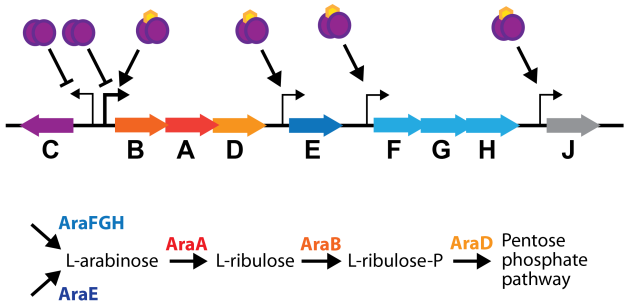

Figure 4.4: Overview of L-arabinose utilization in *E. coli*. The L-arabinose utilization pathway is composed of at least two transporter systems, three enzymes, and one transcription regulator that uptake and consume L-arabinose. The high-affinity transport system (AraF, AraG, AraH) and the low-affinity transport system (araE) actively transport L-arabinose from the periplasm into the cytoplasm. A separate putative transporter (araJ) has no obvious role in the utilization of L-arabinose (ref). The catabolic enzymes (AraB, AraA, AraD) act in series to shunt L-arabinose into the pentose phosphate pathway. The transcription regulator (AraC) represses transcription of itself and the araBAD operon in the absence of L-arabinose. In the presence of L-arabinose, AraC activates transcription of the *araB*, *araE*, *araF*, and *araJ* promoters (refs).

the pre-uninduced cells (4.1 $\pm$ 0.6, n = 3) (Figure 4.5b). In contrast, WT cells exhibited negligible hysteresis (1.0 $\pm$ 0.1, n = 3) (Figure 4.5c). We observed comparable behavior for the xylose utilization pathway, where the extent of hysteresis was 180 times larger for ΔxylAB cells (1,770 $\pm$ 570, n = 3) than for WT cells (9.9 $\pm$ 3.0, n = 3) (Figure 4.7). Taken together, the presence of negative feedback through inducible catabolism reduced hysteresis while preserving a bimodal response.

To further interrogate the contribution of negative feedback, we developed a detailed stochastic model of sugar utilization that accounts for the individual steps of gene expression. The description of the stochastic model is provided in Section 4.2.5. Importantly, the predicted response curves closely matched the response to L-arabinose in both WT and $\Delta araBAD$ strains (Figures 4.8,4.9). The stochastic model predicted a bimodal response, where strengthening negative feedback increased the range of reporter levels in induced cells and reduced sensitivity to the sugar (Figure 4.10). These results demonstrate that coupling inducible transporters and enzymes can generate a flexible, bimodal response to different nutrients.

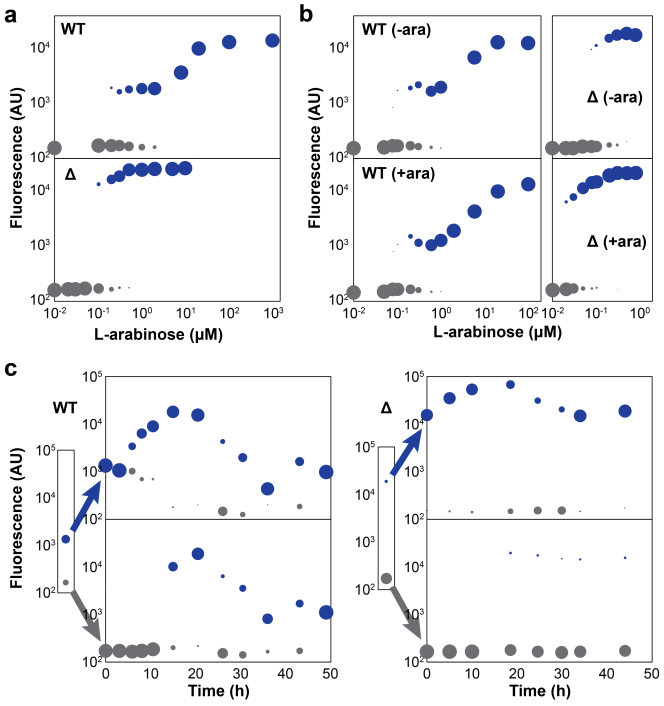

Figure 4.5: a, Response of araB promoter to L-arabinose in WT and Δ*araBAD* (Δ) cells. b, Hysteretic response to L-arabinose in WT or Δ*araBAD* cells. Overnight cultures with (+ara) or without (-ara) sufficient L-arabinose to induce the entire population (100 μM, WT; 10 μM, Δ*araBAD*) were washed and grown in the indicated concentration of L-arabinose for 20 hours. c, Dynamics of WT and Δ cells after sorting into induced and uninduced populations. The cells were grown in 0.3 μM (WT) or 0.1 μM (Δ*araBAD*) L-arabinose before and after cell sorting. All plots are representative of at least three independent experiments.

An important advantage of the stochastic model is its ability to predict the trajectories of individual cells over time. The trajectories predicted that induced cells tended to become uninduced (40%) over the 20-hour simulations, but only in the presence of negative feedback (Figure 4.11). These predictions suggest that WT cells exhibit limited memory of the induced state, potentially explaining the lack of hysteresis to L-arabinose. To test these predictions, we conducted a cell-sorting experiment. WT and Δ*araBAD* cells were grown for 20 hours in sufficient L-arabinose to induce roughly half of the population. Each population then was sorted into induced and uninduced fractions by

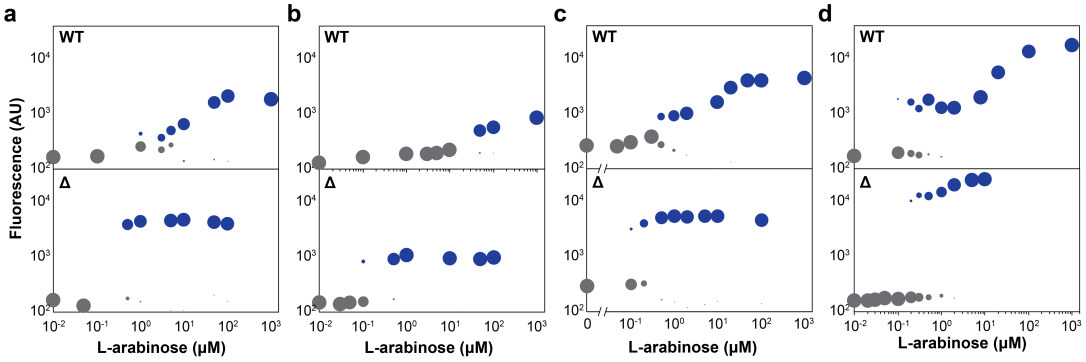

Figure 4.6: Additional response curves for L-arabinose. a, Response curve for the araE promoter in WT and Δ$araBAD$ cells. b, Response curve for the araF promoter in WT and Δ$araBAD$ cells. c, Response curve for the araJ promoter in WT and Δ$araBAD$ cells. d, Response curve for the araB promoter in WT and Δ$araBAD$ cells in the presence of 2 mM cAMP. See Figure 4.3 for an explanation of the dot plots. All plots are representative of at least three independent experiments.

fluorescence activated cell sorting and returned to the same medium conditions. Subsequently, the cells were serially back-diluted to measure the fluorescence distribution at different times while maintaining a low cell density. We found that WT cells exhibited underdamped oscillations after being returned to the original medium (Figure 4.5c). We attribute the oscillations to antimicrobials in the sheath fluid, which inhibit growth. Our collaborator is currently repeating these experiments using 1X PBS as the sheath fluid, where preliminary data suggest that the oscillations disappear. Importantly, the sorted induced and uninduced populations of WT cells converged on the same dynamical path, demonstrating the limited memory of the initial state. In contrast, Δ$araBAD$ cells converged on different, stable distributions (Figure 4.5c) similar to the hysteresis experiments (Figure 4.5b). Taken together, these results validate our model predictions and demonstrate reduced memory in the presence of sugar catabolism. In total, our findings propose that inducible utilization pathways naturally exhibit attenuated bistability, where pathways tend toward induced and uninduced states but can readily switch between states.

The observed single-cell responses in sugar utilization have clear implications for understanding sugar utilization in microorganisms, for engineering synthetic utilization

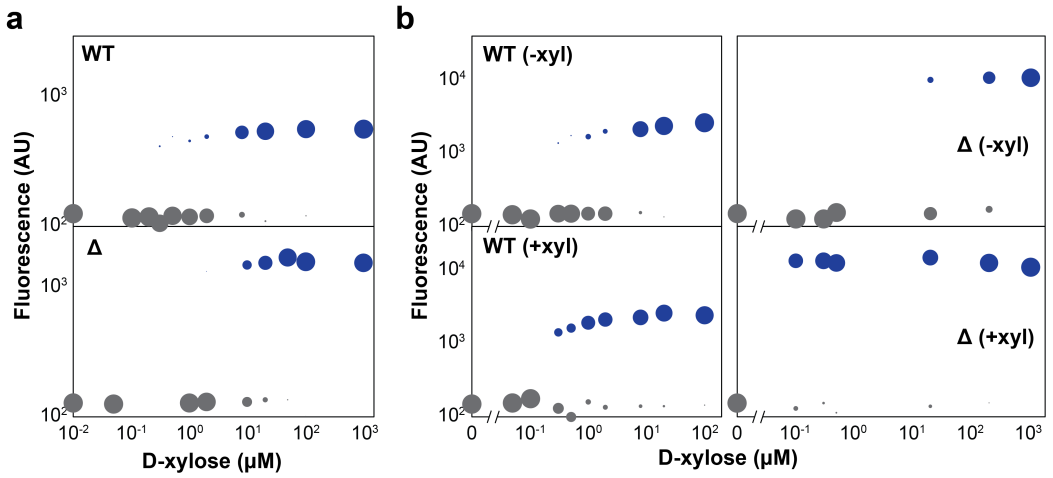

Figure 4.7: Attenuated bistability in the response to D-xylose. Hysteretic response of the xylA promoter to D-xylose in WT or Δ$araBAD$ cells. Overnight cultures with (+xyl) or without (-xyl) sufficient D-xylose to induce the entire population (1 mM, WT; 0.1 mM, Δ$xylA$) were washed and grown in the indicated concentration of L-arabinose for 20 hours. Plots are representative of at least three independent experiments.

pathways, and for generating unique single-cell responses. Sugar utilization in bacteria, archaea, and fungi often combines inducible transporters and catabolic enzymes. We found that this combination can generate unique single-cell responses, where our modeling results suggest that attenuated bistability may be a common behavior in utilization pathways [134, 135]. From an evolutionary standpoint, attenuated bistability may be beneficial by coordinating the utilization of multiple nutrients while rapidly adapting to changing conditions [136, 137]. In addition, the bimodal response would have direct implications for sugar-dependent processes such as biofilm formation [138, 139], DNA uptake [140], and virulence [141]. From an engineering standpoint, similar behaviors would be expected when constructing or transferring sugar utilization pathways for environmental bioremediation or for inducible promoter systems [142, 143, 144]. As a result, careful consideration of the biochemical characteristics of the pathway as well as ensuing single-cell studies will be required to ensure efficient utilization of environmental toxins or to achieve predictable induction. Finally, the combination of inducible production/transport and degradation/export offers a simple structure for generating diverse single-cell responses that may feature a common structure in biology.

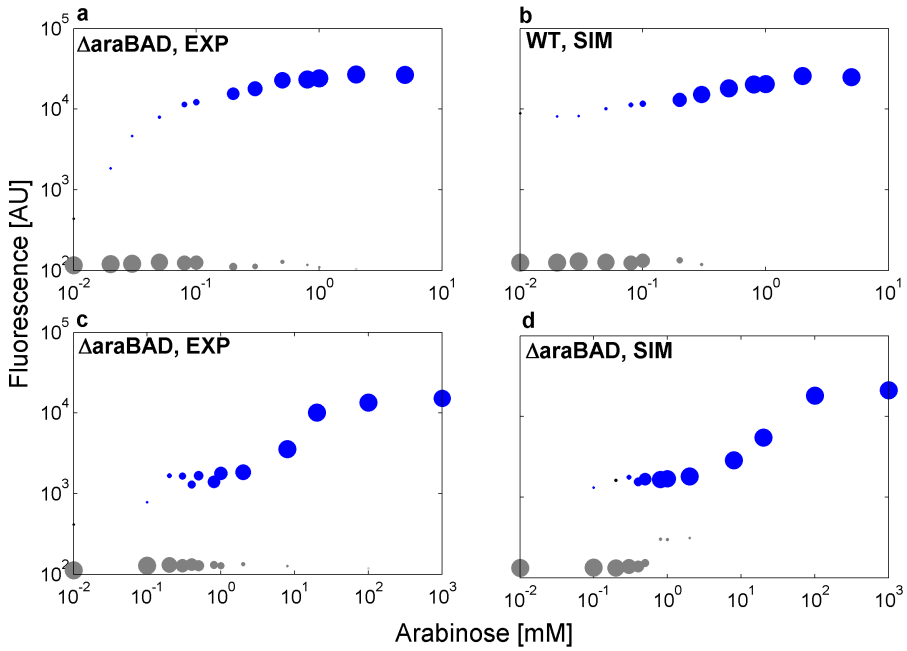

Figure 4.8: a,b, comparison between flow cytometry data and computational results for the WT system. c,d, comparison between flow cytometry data and computational results for the $\Delta araBAD$ system. See Figure 4.3 for an explanation of the dot plots.

4.2 Methods

4.2.1 Bacterial strains and plasmids

Catabolic operons from *E. coli* K-12 MG1655 were replaced with a chloramphenicol resistance cassette using standard recombineering techniques. Plasmids were obtained from the *E. coli* promoter collection (OpenBioSystems). Promoters unavailable in this collection were amplified from the MG1655 chromosome and inserted into pUA66 using standard cloning techniques.

4.2.2 Growth conditions and media

Cells in M9 minimal medium with 0.4% glycerol were shaken overnight at 250 RPM and 37^oC. Overnight cultures were back diluted into the same medium with different concentrations of the sugar and grown for 20 hours prior to flow cytometry analysis.

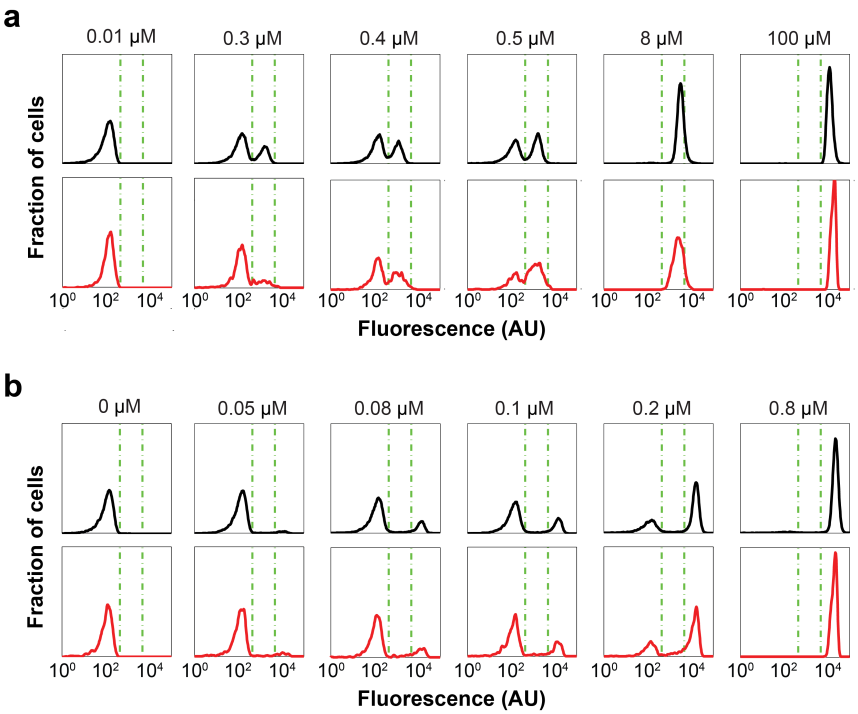

Figure 4.9: Overlaid histograms illustrate the match between the experimental results (black) and the model predictions (red). a, WT cells from Figure 4.8(a,b) and the model with degradation rate equal to 0.19. b, araBAD cells from Figure 4.8(c,d) and the model with degradation rate equal to 0.

Dilutions were made so the cell densities prior to flow cytometry analysis were ABS600 $\sim$ 0.004. To measure hysteresis, the indicated concentration of sugar was included in the overnight cultures.

4.2.3 Flow cytometry

Fluorescence measurements were performed on a BD Accuri C6 flow cytometer equipped with CFlow plate sampler, a 488 nm laser, and a 530 $\pm$ 15 nm bandpass filter. Cells were gated based on forward scatter and side scatter. Approximately 5,000 gated events were collected for each measurement.

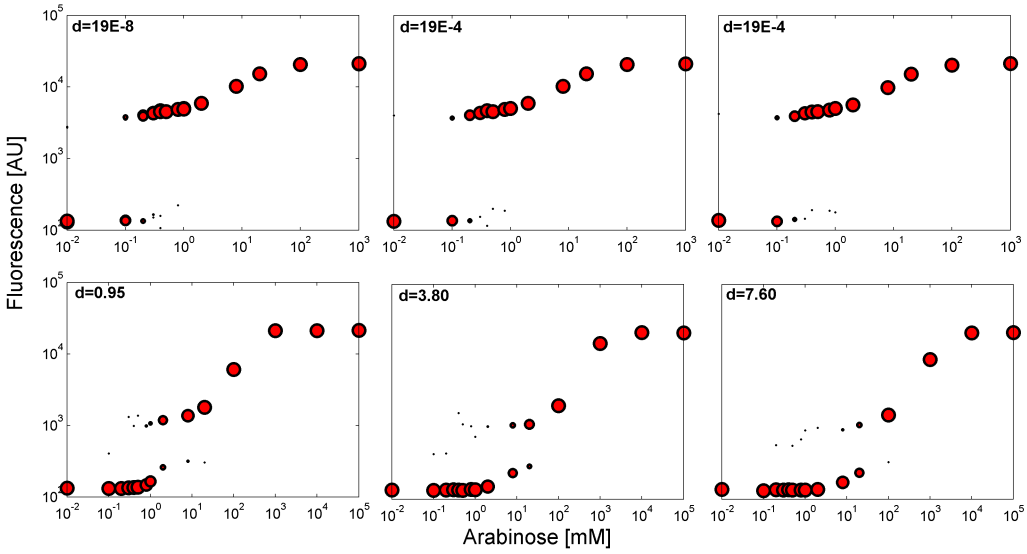

Figure 4.10: Single-cell response for differing catabolism rate as predicted by the stochastic model. Dots represent induced and uninduced subpopulations for different activities of the enzyme (d $[s^{-1}]$). All dots are shown in red for clarity.

4.2.4 Cell sorting

Cells were grown for 20 hrs in M9 minimal media with 0.4% glycerol containing 0.3 μM of L-arabinose (WT cells) or 0.1 μM of L-arabinose ($\Delta araBAD$ cells) to give induced and uninduced cells. Cells were then sorted into uninduced and induced populations on a MoFlo cell sorter (Beckman Coulter). Sorted cells were back-diluted different amounts into medium with the original concentration of L-arabinose. Each back-diluted culture was grown until reaching an ABS600 of $\sim$0.004 and then subjected to flow cytometry analysis.

4.2.5 Mathematical modeling

We developed a detailed reaction network that captures the biomolecular interactions engaged in the arabinose system. The modeling approach we followed here has been discussed previously [31, 32]. The biochemical reactions along with their kinetic parameters are detailed in Table 4.1. The definition of all the corresponding species and their initial conditions are shown in Table 4.2. The reaction network of the deletion

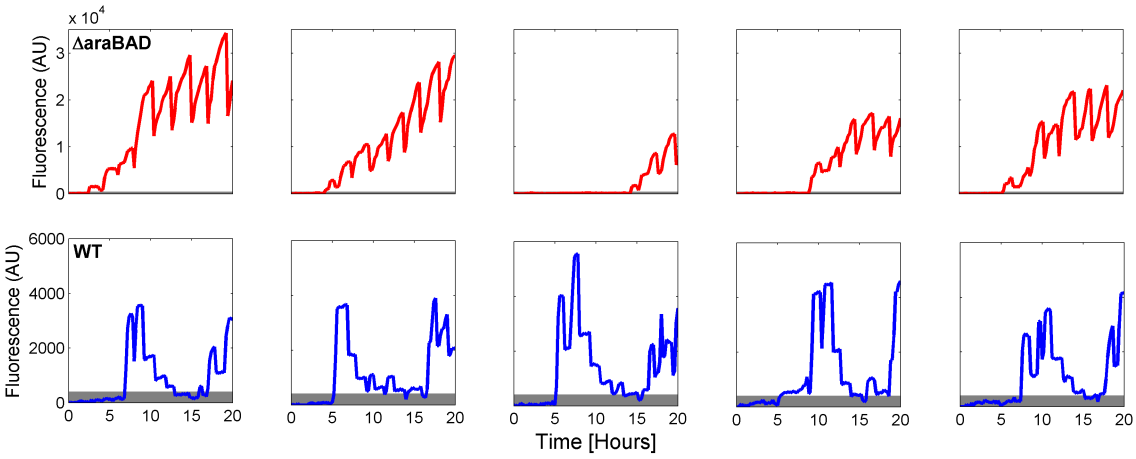

Figure 4.11: Representative traces of single-cell simulations for the $\Delta araBAD$ (red) and WT case (blue). Only in the latter case cells switch bidirectionally between the induced and uninduced (shown in gray) state.

strain differs from the wild type reaction network by a single reaction (reaction #36) which captures the catabolism of arabinose by AraBAD occurring solely in the wild type system.

The model accounts for the reversible binding of arabinose to AraC (reactions 1-8), reversible binding of Arac to DNA (reactions 9-14), transcription of $araBAD$ and $araFGH$ genes (reactions 15-19 and 23-27) and the associated mRNA translation (reactions 20-22 and 28-30), active and passive arabinose diffusion into the cell (reactions 32-33), and mRNA degradation (reactions 34-35). In addition, the model takes into consideration the difference in the cell division time between the induced and uninduced cell population. The cell division times of induced and uninduced cells used in the simulations were determined experimentally and are presented in Table 4.3. Based on the flow cytometry data, we assumed that a critical threshold for switching the cell division time was equal to 400 AraBAD Arbitrary Units (AU). In other words, cells whose AraBAD amount is lower than 400 AU follow an "uninduced" cell division time whereas cells with more than 400 AraBAD AU undergo an "induced" cell division time. The species whose amount is halved at each cell division include arab, transporter, enzyme and mRNA. Overall, the model consists of 26 species and 36 reactions. The main

assumptions of the model are the following:

1) The model assumes a constant number of Arac proteins per cell (approximately 50 [145]) and consequently does not consider Arac production in detail (gene transcription, mRNA translation). This is not an unreasonable assumption as various studies have demonstrated that AraC production takes place for at most 30 minutes post-arabinose administration and subsequently its amount goes back to its pre-induced levels [146, 147]. Bearing in mind that here we focused on the behavior of the cells 20 hours post-induction, considering constant AraC amount does not impact the presented results.

2) The model considers only one arabinose transporter (we call it AraFGH). The arabinose transporter used in the model is assumed to be a combination of the two transporters, AraFGH and AraE, and its kinetic parameters were estimated so that the simulation results match the experimental phenotype.

3) The cells are initially grown in the presence of glycerol and, therefore, we assume an abundance of CAP-cAMP molecules in each cell. To this end, CAP-cAMP is neglected in the model. Hence, the only requirement for RNA polymerase recruitment to DNA is the existence of an AraC:Arabinose complex bound to DNA.

It is widely accepted that stochasticity is ubiquitous in gene expression and its associated biological processes [61, 58, 28, 148]. To account for this stochasticity, we used Hy3S (described in 2.2.3). Similarly to previous studies [149], to match the simulation results with the GFP fluorescence we assumed a linear relationship between the number of molecules (simulations) and the GFP amount (experiments). More specifically, we hypothesized that one simulated GFP molecule corresponds to 20 GFP AU.

The kinetic parameters used in the model were chosen based on the following notes (see also Table 4.1):

a) Rate adjusted so that 0.01 uM of arabinose correspond to $4.5 \cdot 10^{-11}(M \cdot s)^{-1}$.

b) Rate estimated based on experimental data from this study.

c) Rate estimated based on experimentally measured dissociation constants, K_d [M].

d) Wild-type AraC protein always binds to the DNA with the same affinity regardless of the presence or absence of arabinose [150].

e) Binding affinity of AraC:Arabinose for DNA($araFGH$) is assumed equal to binding affinity of AraC:Arabinose for DNA($araBAD$).

f) The modeling approach followed here (formation of an open and subsequently a closed

loop) is adopted from [151].

g) This reaction is treated as a gamma distributed event, as described in [74]. The kinetic parameter k_i represents the rate of each exponentially distributed event (nucleotide transcription or amino-acid translation) whereas k_{ib} corresponds to the number of events (number of nucleotides and amino-acids).

h) Minimum possible length of the RNA transcript of the *araBAD* operon, coding the amino acid sequences for L-ribulokinase, L-arabinose isomerase, and L-ribulose 5-phosphate 4-epimerase in *E. coli* B/r.

i) Rate adjusted to give approximately 20 proteins per mRNA transcript [5].

j) *araFGH* gene expression is considered to occur only in the presence of arabinose, as in the absence of the latter the amount of AraFGH protein is approximately 1 molecule per cell [152].

k) Calculated based on the half-life.

Units of k: zero order reactions = $M \cdot s^{-1}$, first order reactions = s^{-1}, second order reactions = $(M \cdot s)^{-1}$.

Table 4.1: Reaction network capturing L-arabinose operon behavior.

Reaction	Kinetic Constant	Ref.
$\varnothing \xrightarrow{k_1} Arab$	$k1$	a
$AraFGH \xrightarrow{k_2} AraFGH + Arab$	$k_2 = 2 \cdot 10^{-2}$	b
$AraC + Arab \underset{k_4}{\overset{k_3}{\rightleftharpoons}} AraC{:}Arab1$	$k_3 = 2 \cdot 10^2, k_4 = 0.5$	[153],c
$AraC{:}Arab1 + Arab \underset{k_6}{\overset{k_5}{\rightleftharpoons}} AraC{:}Arab2$	$k_5 = 10^2, k_6 = 1$	[153],c
$AraC + DNA \underset{k_8}{\overset{k_7}{\rightleftharpoons}} AraC{:}DNA$	$k_7 = 1.5 \cdot 10^9, k_8 = 1.1 \cdot 10^{-3}$	[154],d
$AraC{:}Arab2 + DNA \underset{k_{10}}{\overset{k_9}{\rightleftharpoons}} AraC{:}Arab2{:}DNA$	$k_9 = 1.5 \cdot 10^9, k_{10} = 1.1 \cdot 10^{-3}$	[154],d
$AraC{:}Arab2 + DNA2 \underset{k_{12}}{\overset{k_{11}}{\rightleftharpoons}} AraC{:}Arab2{:}DNA2$	$k_{11} = 1.5 \cdot 10^9, k_{12} = 1.1 \cdot 10^{-3}$	[154],d,e
$RNAp + AraC{:}Arab2{:}DNA \underset{k_{14}}{\overset{k_{13}}{\rightleftharpoons}} RNAp{:}AraC{:}Arab2{:}DNA$	$k_{13} = 10^7, k_{14} = 3 \cdot 10^{-3}$	[151],c,f
$RNAp{:}AraC{:}Arab2{:}DNA \xrightarrow{k_{15}} RNAp^*{:}AraC{:}Arab2{:}DNA$	$k_{15} = 1.7 \cdot 10^{-2}$	[151],f
$RNAp^*{:}AraC{:}Arab2{:}DNA \xrightarrow{k_{16}} RNAp(B) + AraC{:}Arab2{:}DNA$	$k_{16} = 41$	[155]
$RNAp(B) \xrightarrow{k_{17}, k_{17b}} RNAp + mRNA(B)$	$k_{17} = 41, k_{17b} = 3,800$	[155, 156],g,h
$Rib + mRNA(B) \xrightarrow{k_{18}} Rib{:}mRNA(B)$	$k_{18} = 10^5$	[5],i
$Rib{:}mRNA(B) \xrightarrow{k_{19}} Rib^*{:}mRNA(B) + mRNA(B)$	$k_{19} = 14$	[157]
$Rib^*{:}mRNA(B) \xrightarrow{k_{20}, k_{20b}} Rib + araBAD$	$k_{20} = 14, k_{20b} = 1,267$	[157, 156],g,h
$RNAp + AraC{:}Arab2{:}DNA2 \underset{k_{22}}{\overset{k_{21}}{\rightleftharpoons}} RNAp{:}AraC{:}Arab2{:}DNA2$	$k_{21} = 10^7, k_{22} = 2 \cdot 10^{-3}$	[158],c,f,j
$RNAp{:}AraC{:}Arab2{:}DNA2 \xrightarrow{k_{23}} RNAp^*{:}AraC{:}Arab2{:}DNA2$	$k_{23} = 1.7 \cdot 10^{-2}$	[158],f
$RNAp^*{:}AraC{:}Arab2{:}DNA2 \xrightarrow{k_{24}} RNAp(F) + AraC{:}Arab2{:}DNA2$	$k_{24} = 41$	[155]
$RNAp(F) \xrightarrow{k_{25}, k_{25b}} RNAp + mRNA(F)$	$k_{25} = 41, k_{25b} = 3,486$	[155, 159],g
$Rib + mRNA(F) \xrightarrow{k_{26}} Rib{:}mRNA(F)$	$k_{26} = 10^5$	[5],i
$Rib{:}mRNA(F) \xrightarrow{k_{27}} Rib^*{:}mRNA(F) + mRNA(F)$	$k_{27} = 14$	[157]
$Rib^*{:}mRNA(F) \xrightarrow{k_{28}, k_{28b}} Rib + GFP$	$k_{28} = 14, k_{28b} = 1,162$	[157, 159],g
$mRNA(B) \xrightarrow{k_{29}} \varnothing$	$k_{29} = 5.8 \cdot 10^{-3}$	[152],k
$mRNA(F) \xrightarrow{k_{30}} \varnothing$	$k_{30} = 5.8 \cdot 10^{-3}$	[152],k
$Arab + AraBAD \xrightarrow{k_{31}, k_{31b}} AraBAD$	$k_{31} = 0.18, k_{31b} = 10^{-7}$	[160],b
$\varnothing \xrightarrow{k_{32}} AraBAD$	$k_{32} = 1.2 \cdot 10^{-12}$	b

Table 4.2: Definition and initial conditions (IC) of the species engaged in the arabinose system.

Species	Definition	IC
Arab	Arabinose	0
AraFGH	AraFGH transporter protein	0
Arac	Arac regulator protein	50
Arac:Arab1	Arac bound by 1 Arab molecule	0
Arac:Arab2	Arac bound by 2 Arab molecules	0
DNA	DNA of AraBAD	0
Arac:DNA	Arac bound to AraBAD DNA	1
Arac:Arab2:DNA	Arac:Arab2 bound to AraBAD DNA	0
DNA2	DNA of AraFGH	1
Arac:Arab2:DNA2	Arac:Arab2 bound to AraFGH DNA	0
RNAp	RNA polymerase	500
RNAp:Arac:Arab2:DNA2	RNAp bound to Arac:Arab2:DNA2	0
RNAp:Arac:Arab2:DNA	Closed complex of RNAp bound to Arac:Arab2:DNA	0
RNAp*:Arac:Arab2:DNA	Open complex of RNAp bound to Arac:Arab2:DNA	0
RNAp(B)	RNAp bound to araBAD gene starting transcription	0
mRNA(B)	mRNA of $araBAD$ gene	0
Rib	Ribosome	500
Rib:mRNA(B)	Rib bound to mRNA(B)	0
Rib*:mRNA(B)	Rib bound to the first codon of mRNA(B)	0
AraBAD	AraBAD enzyme	0
RNAp*:Arac:Arab2:DNA2	Open complex of RNAp bound to Arac:Arab2:DNA2	0
RNAp(F)	RNAp bound to $araFGH$ gene starting transcription	0
mRNA(F)	mRNA of $araFGH$ gene	0
Arac:Arab1:DNA	Arac:Arab1 bound to AraBAD DNA	0
Rib:mRNA(F)	Rib bound to $mRNA(F)$	0
Rib*:mRNA(F)	Rib bound to the first codon of $mRNA(F)$	0

Table 4.3: Doubling times for induced and uninduced cells.

System	Cell division time (± 10 min)	Cell division time (± 10 min)
	Induced cells	Uninduced cells
$\Delta araBAD$	129	97
WT	80	101

Chapter 5

ProTeOn and proTeOff, new protein devices that inducibly activate bacterial gene expression

5.1 Introduction

Controllable gene regulatory systems are the subject of continued, intense investigations. In addition to being critically important in explaining how phenotypes emerge from genotypes in living organisms, their components are rapidly becoming integral in efforts towards engineered gene expression control [161, 87, 95, 162, 163, 164, 165, 166, 167, 168]. Two well-studied example regulatory systems are the tetracycline (*tet*) and luminescence (*lux*) operons. The tetracycline repressor protein, TetR, and numerous TetR derivatives, afford a remarkably robust function of inducible repression and de-repression of gene expression. Thus, they have been employed in numerous synthetic biology and bioengineering applications [4, 7, 169, 12]. Switching gene expression on and off by TetR and TetR derivatives simply depends on the presence or absence of the antibiotic tetracycline (Tc). The lux operon's transcription activator protein, LuxR, is another attractive controllable molecular component for engineering applications. When induced, it activates gene overexpression by recruiting transcriptional machinery to the promoter. In addition, number of variant activators have been identified that upregulate transcription over

a range of strengths with varied dependence on inducer molecules [170, 171, 172, 173].

Herein, we present two novel synthetic molecular devices that inducibly upregulate bacterial gene expression. We have designed, built and characterized these devices using a unique methodology that is based on both computational and experimental efforts. Both devices are composed of an inducible DNA-binding domain and a transcription activation domain. In particular, TetR and the reverse-TetR (rTetR) derivative were chosen for the inducible domains while the transactivating domain of LuxR was incorporated for transcription upregulation. TetR protein and the rTetR derivative dissociate from and bind to the *tetO* operator sequence in response to Tc (respectively) [99, 85, 174, 175]. This is the behavior that made them attractive candidates for controlling our devices. In addition, LuxRΔN(2-162) (LuxRΔN) is the C-terminal, constitutive transactivating domain of the full length LuxR activator. Strong, constitutive transcription activation is achieved with this variant lacking the N-terminal residues 2-162 of the full protein [175]. These specific protein components were selected as they have been thoroughly studied and well characterized. Specifically, their protein:DNA-operator crystal structures for both domains were available to us as well as the kinetic parameters that govern their protein:protein and protein:DNA interactions.

We designed, constructed and characterized the two new devices using an original workflow that integrates experimental synthetic biology, molecular modeling, and stochastic reaction kinetic simulations. This methodology may be directly implemented for the development of other biological devices and larger regulatory networks. The devices we present here, prokaryotic-TetOn and prokaryotic-TetOff (proTeOn and proTeOff), function in an Tc-dependent manner: the proTeOn synthetic protein (PROTEON) activates gene expression in the presence of Tc, and the proTeOff synthetic protein (PROTEOFF) activates expression in the absence of Tc. Orders of magnitude higher expression levels are observed in the activating states compared to the basal expression levels of both systems. With these new devices, there is now a dial of activated expression to complement those of basal and repressed gene expression.

While the proTeOn and proTeOff systems are functionally unique, their designs are conceptually inspired by the widely used TetOn and TetOff systems, which function in eukaryotic systems [41, 44]. Gossen and co-workers fused the TetR protein to a mammalian transactivating domain, building molecular devices that activate eukaryotic gene

expression in an Tc-dependent manner. To our knowledge, there were no prokaryotic transcription factors homologous to TetOn or TetOff prior to the construction of proTeOn and proTeOff. Additionally, our kind of engineering approach to synthetic biology, using molecular dynamic simulations to guide system design and stochastic simulations to enhance system characterization, has not been previously reported in the literature.

The proposed behaviors of proTeOn and proTeOff are illustrated in Figure 5.1. As shown, anhydrotetracycline (aTc), a Tc derivative, was used as the inducer in our experiments. For proTeOn, in the absence of aTc, the inducible DNA binding domain, rTetR, does not bind the *tetO* operator and the PROTEON protein does not upregulate the target gene, green fluorescence protein (gfp). Upon activation with aTc, rTetR binds the inducer, undergoes a conformational change and binds *tetO* [174]. This binding brings LuxRΔN near its operator site, allowing it to bind luxbox and upregulate gfp transcription through RNApol recruitment to the promoter [176, 177, 178]. For proTeOff, in the absence of aTc, its inducible DNA binding domain, TetR, binds *tetO* [87, 85, 179, 180]. This binding brings LuxRΔN close to its operator site permitting it to bind luxbox and upregulate gfp transcription [176, 177, 178]. After the addition of aTc, TetR binds the small molecule, undergoes a conformational change, and releases *tetO* [87, 85, 179, 180]. Upon dissociation of TetR:*tetO*, the LuxRΔN:luxbox interaction is destabilized and transcription upregulation is terminated.

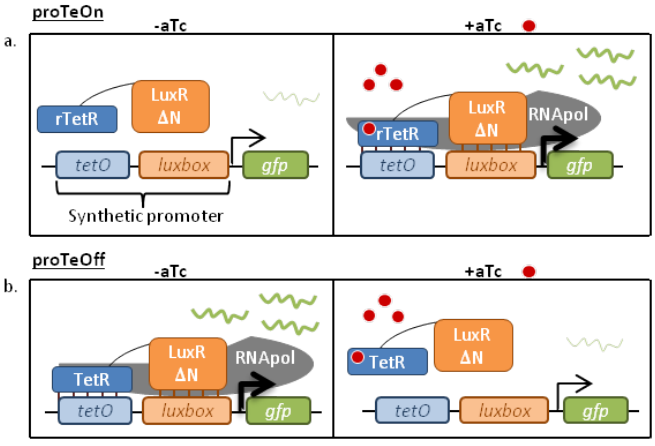

Figure 5.1: a) proTeOn behavior. b) proTeOff behavior.

To achieve these phenotypes, we first used molecular modeling to design the geometry of the synthetic transactivator/promoter pairs of each system. The important feature of proTeOn and proTeOff on which we focused was the optimized interaction between the inducible synthetic transactivator protein that upregulates gene expression and its complementary synthetic promoter. Next, we built and characterized both systems experimentally. Last, implementing stochastic simulations, we quantified the system dynamics and determined unknown kinetic parameters of key interactions. We detail the workflow in the Methods section. In the Results and Discussion section, we begin by presenting elements of the model-driven design of the devices. ProTeOn and proTeOff can be applied to robustly manage prokaryotic gene expression. Variant systems and more complex inducible gene regulatory networks can also be designed and constructed using this workflow and the proTeOn and proTeOff systems as a foundation.

5.2 Results and discussion

5.2.1 proTeOn and proTeOff designs

For proTeOn and proTeOff to perform efficiently, both protein domains (TetR/rTetR and LuxRΔN) must readily bind their operator sequences upon induction, LuxRΔN must recruit RNApol to the promoter and then RNApol must bind to the promoter and begin transcription. Numerous length and space requirements are associated with each of these steps and we systematically accommodated each of them by designing the devices prior to construction. To investigate and satisfy the above geometric constraints, we built molecular models of each system using MOE and NAMD [181]. With the chosen protein components and DNA operator sites as starting points, the entire DNA promoter sequence and the linker amino acid sequence were designed in tandem.

The synthetic promoter for both systems was designed first, and the optimized sequence is annotated in Figure 5.2a. The major aims of the promoter design were to minimize the distance between the two DNA-bound protein domains while maintaining favorable binding between the domains and their operators. The interactions between RNApol and the specific residues of the promoter to which it binds (the UP element and the -10 region) had to be accommodated as well. To meet these aims, the operator sites were integrated into the promoter such that the two domains bound along the same face

of the DNA (Figure 5.2b, side view) while maintaining a minimum distance between the two operator sites. The final design was scrutinized to ensure that neither DNA-bound protein domain was encroaching the residues of the UP element or -10 region as this would inhibit efficient RNApol recruitment.

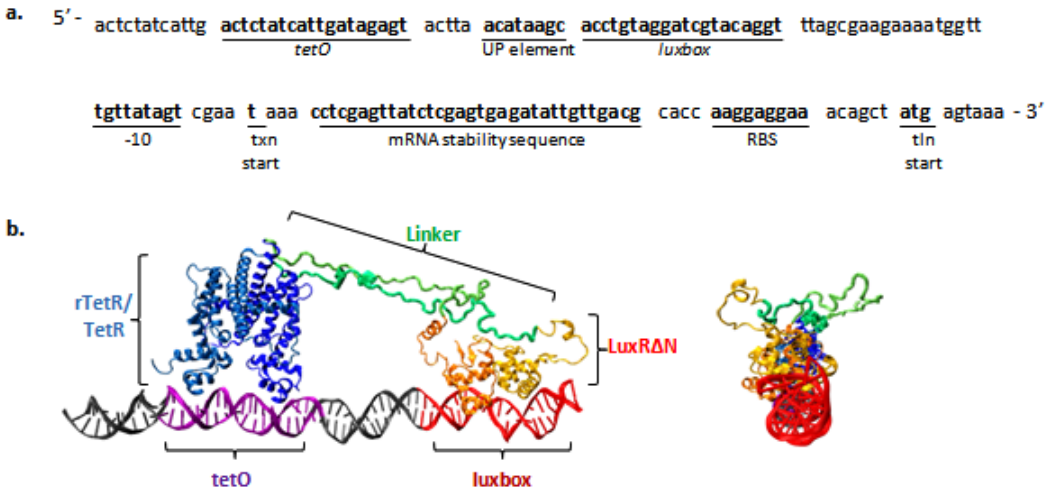

Figure 5.2: a) proTeOn and proTeOff synthetic promoter sequence. b) proTeOn molecular model. Both proTeOn and proTeOff are designed to assemble as shown. The inducible DNA binding domain (rTetR/TetR, blue) binds the *tetO* operator (purple), and the transcription activator domain (LuxRΔN, orange) binds the luxbox (red). The two domains bind their operators along the same face of the DNA double helix and are connected (TetR/rTetR's C-terminus to LuxRΔN's N-terminus) by a linker peptide (green).

Next, we optimized the peptide linker required to connect the two DNA-bound protein domains (TetR/rTetR's C-terminus to LuxRΔN's N-terminus). We did this by assuming it resembles a polymer in good solution and choosing a length that would minimize the entropic elastic tension effects. The optimal linker length was determined to be 150 $\mathring{A}$ and the final sequence is the product of 5 repeats of a 9-amino acid subunit, ARTQYSESM. The individual subunits are connected by single glycine residues, and this 49-amino acid peptide is flanked by 3 glycine residues on each side. These were chosen to ensure linker flexibility and promote correct protein domain folding. The final optimized proTeOn system is illustrated in Figure 5.2b.

5.2.2 Characterization of proTeOn and proTetOff behavior

To investigate the behavior of both systems and to determine the optimal conditions and applications of each device, we performed two sets of experiments. First, cells expressing the transactivators were induced with a range of aTc concentrations. Second, cultures were primed with these aTc amounts prior to transactivator expression. In both experiments we monitored the resulting differential expression of the target gene, *gfp*, over time. Induction experiments were repeated for all inducer concentrations and time points. The reported trends were observed across the replicates.

proTeOn

Upon administration of aTc, proTeOn upregulates target genes by one hour post-treatment and achieves 15-fold upregulation through long times. With low (10 ng/ml) aTc, proTeOn achieves steady state expression of targets 10-fold over untreated cultures just 5 hours post-treatment. With high (200 ng/ml) aTc, GFP reaches steady state expression by 10 hours post-treatment at levels 15-fold over untreated cultures. These behaviors are displayed as both expression means and population distributions in Figures 5.3 and 5.4 respectively. Basal GFP expression levels are observed from the synthetic promoter in the absence of aTc. There is no repressed state for the proTeOn system as presented here.

PROTEON expressed in the presence of aTc gives target gene upregulation within 2 hours. Steady state upregulation is achieved by 5 hours after the transactivator's initial expression, with GFP expression 10-fold over uninduced controls. PROTEON is thus rapidly transcribed, translated, folded and becomes functional when under the control of a free, non-repressed, promoter. This is shown in Figure 5.3. The mean behaviors taken across biological replicates are discussed in Appendix A.

Applications of proTeOn would be appropriate in systems when quick bursts or long term gene upregulation is desired. Such behavior is governed by rTetR's responses to induction with aTc. This control is achievable with inducer levels as low as 10 ng/ml aTc. aTc concentrations outside of this range, or scaled throughout an experiment, may be useful depending upon the desired level of activation. Notably, long term target gene upregulation increases with aTc up to 200 ng/ml. Thus, inducing with a range of aTc

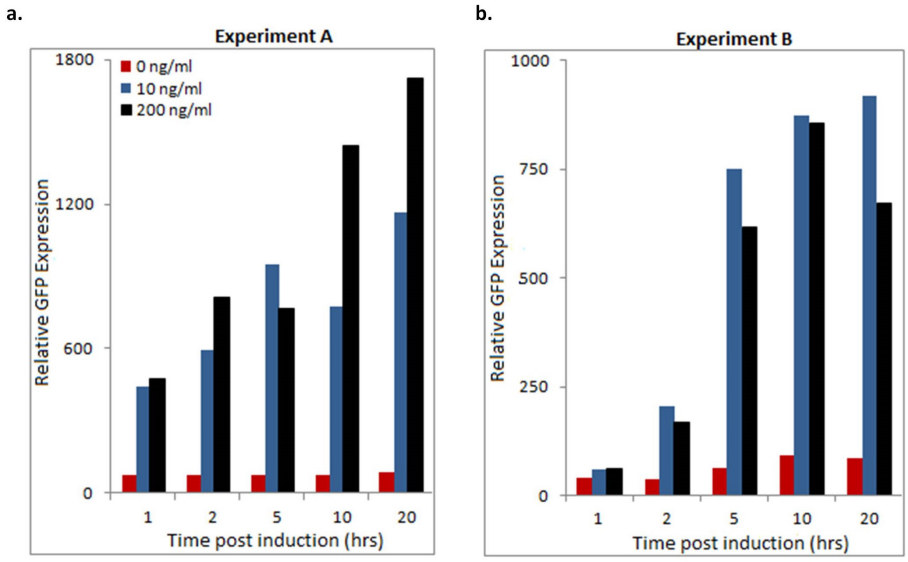

Figure 5.3: proTeOn system phenotype analysis by flow cytometry: Mean GFP expression as analyzed by flow cytometry 1, 2, 5, 10, and 20 hours post-treatment for both experimental set-ups. Maximum overexpression is approximately 10-fold above uninduced controls. a) Cells expressing PROTEON were induced with 0, 10 and 200 ng/ml aTc. b)PROTEON was expressed in cells pre-cultured with 0, 10 and 200 ng/ml aTc. Maximum upregulation is 10-fold above uninduced controls with both aTc concentrations.

concentrations, proTeOn can be tuned to robustly upregulate gene expression for both acute and long time scales.

proTeOff

In the absence of aTc, proTeOff readily activates target gene expression. Upon administration of aTc, upregulation is reduced by half just one hour post-treatment. Steady state target gene expression is only one-half and one-fifth that of untreated samples with low and high aTc (respectively). This is observable by 5 hours and remains through long times. While activation by proTeOff is significantly reduced with low aTc concentrations, a subset of activator proteins still appear to bind the promoter and upregulate transcription. However, with GFP levels at only one-fifth of the untreated controls, we

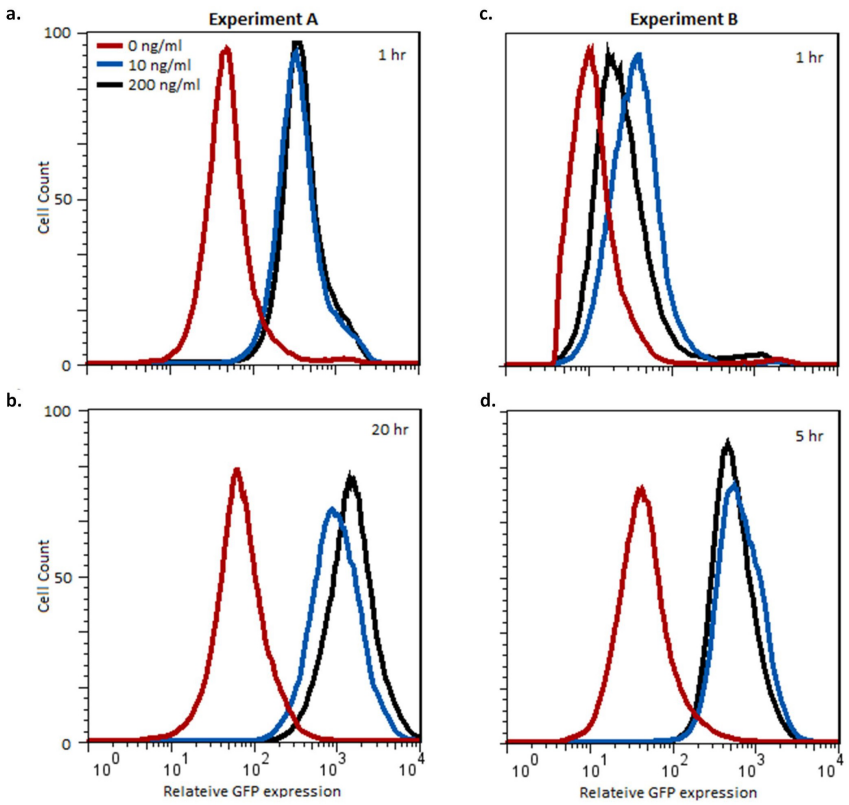

Figure 5.4: proTeOn system phenotype analysis by flow cytometry: GFP expression distribution as analyzed by flow cytometry. a,b) 1, 20 hours post-treatment, respectively, for the first experimental set-up. c,d) 1, 5 hours post-induction, respectively, for the second experimental set-up.

consider the low GFP expression observed with 200 ng/ml aTc to be the synthetic promoter's basal expression level. These low basal expression levels are observed through long times. These behaviors are presented also as expression means and population distributions in Figures 5.5 and 5.6 respectively. Please note that there is also no repressed state for the proTeOff system as presented here.

PROTEOFF expressed in the absence of aTc leads to target gene upregulation within one hour and maintains it through long times. PROTEOFF is therefore quickly transcribed, translated, folded and functional when under the control of a free, non-repressed, promoter. In the presence of low and high aTc concentrations, the device

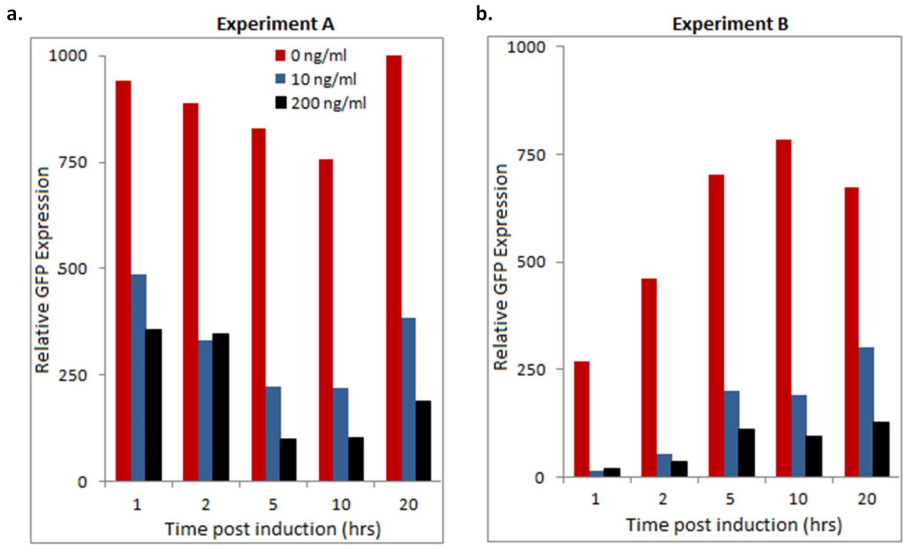

Figure 5.5: proTeOff system phenotype analysis by flow cytometry: Mean GFP expression as analyzed by flow cytometry 1, 2, 5, 10, and 20 hours post-treatment for both experimental set-ups. a) With 0 ng/ml, proTeOff upregulates GFP expression. With 10 and 200 ng/ml aTc, expression is reduced to half that of untreated samples by one hour. Minimum expression is one-half and one-fifth that of untreated samples with low and high aTc respectively, b) PROTEOFF was expressed in cells pre-cultured with 0, 10 and 200 ng/ml aTc. In 0 ng/ml aTc, proTeOff activity is observed one hour after PROTEOFF expression is induced. Low concentrations permit GFP levels to rise to one-third that of untreated cultures.

activity is maintained at less than one-fourth of untreated cultures one hour after the PROTEOFF's initial expression. An aTc concentration of 200 ng/ml maintains this very low basal activity, thus basal gene expression, through long times. Treatment with low aTc allows the target's expression to rise to one third of untreated cultures by 5 hours after the transactivator's initial expression, where it remains through long times. This elevation is due to a subset of free PROTEOFFs binding to the promoter and recruiting RNApol. These data are shown in Figure 5.5. The mean behaviors taken across biological replicates are provided in Appendix A. In addition, proTeOff can be applied to upregulate gene expression for both short and long times. Continuous long term upregulation can be achieved, as well as upregulation with intermittent periods of low expression. These behaviors are governed by TetR's response to aTc. aTc levels as

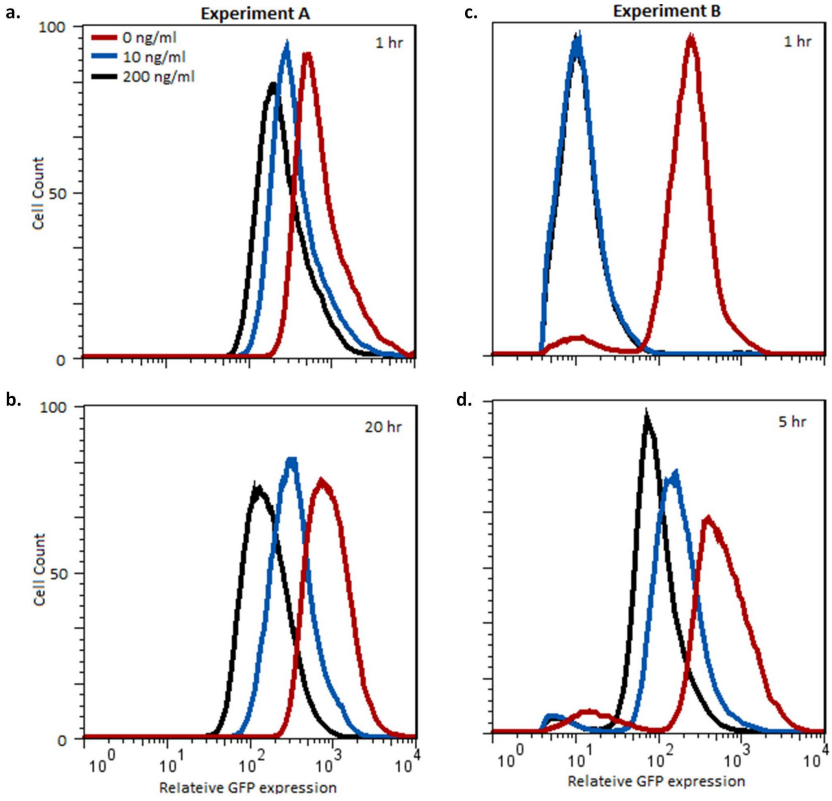

Figure 5.6: proTeOff system phenotype analysis by flow cytometry: GFP expression distribution as analyzed by flow cytometry. a,b) 1, 20 hours post-treatment, respectively, for the first experimental set-up. c,d) 1, 5 hours post-induction, respectively, for the second experimental set-up.

low as 10 ng/ml effectively achieve periods of low expression. Target gene expression drops with increasing aTc concentrations 10-200 ng/ml. aTc concentrations outside of this range or scaled throughout an experiment, may also to achieve a desired phenotype. Thus, proTeOff can be tuned to upregulate gene expression for short, long-continuous and long-intermittent time scales using a range of aTc concentrations.

5.2.3 Characterization of proTeOn and proTeOff kinetics

In addition to experimentally testing proTeOn and proTeOff, we have also assessed the kinetics of both systems by conducting stochastic simulations. This was done to

characterize the systems at a finer resolution than can be achieved in the lab alone. We aimed to quantify the strength of the interactions between the device components by developing a stochastic model that captures the time profiles of measured GFP probability distributions. We modeled the transcription, translation, regulatory and degradation events with stochastic kinetics. We built the models and conducted the simulations as described before [34, 73, 31, 182, 183, 12, 55, 74, 80, 5]. Stochastic model parameters were fit to match the untreated experimental phenotypes (0 ng/ml) and the behaviors upon induction with both low and high aTc concentrations (10 and 200 ng/ml). The mean GFP expression is captured by the model for both proTeOn and proTeOff at short times post-aTc induction (1 and 5 hours) as well as at steady state (10 hours), as shown in Figure 5.7. Overall, the mean GFP levels achieved by the simulation trajectories agree well with the experimentalobservations.

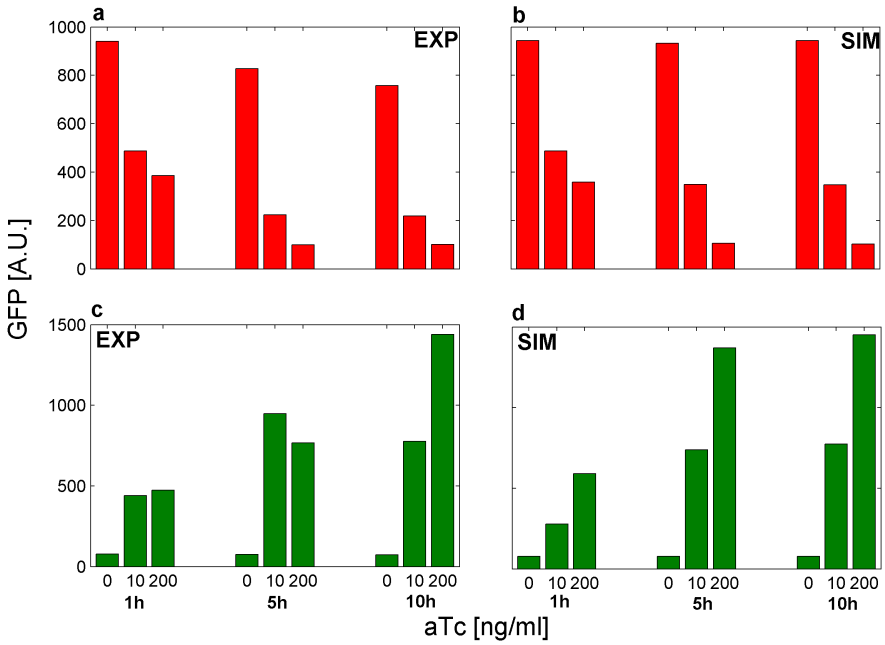

Figure 5.7: proTeOff (a,b) and proTeOn (c,d) average GFP by stochastic simulations and experimental analysis at 1, 5, and 10 hours, when 0, 10, and 200 ng/ml aTc are administered.

A discrepancy is observed between the simulation and experimental results at 5 hours with 200 ng/ml aTc. This may be attributed to high aTc levels retarding cellular

processes, such as protein overexpression [184]. By 10 hours, this effect is no longer experimentally significant and a good match is therefore observed between theoretical and experimental results. Due to their stochastic nature, the models can also capture the GFP distributions observed experimentally at short and long times. The GFP distributions at 10 hours for proTeOn and at 1 hour for proTeOff, when 0, 10 and 200 ng/ml aTc are administered, are presented in Figure 5.8. For both systems, and across all aTc concentrations, the distributions generated *in silico* match those observed *in vivo*.

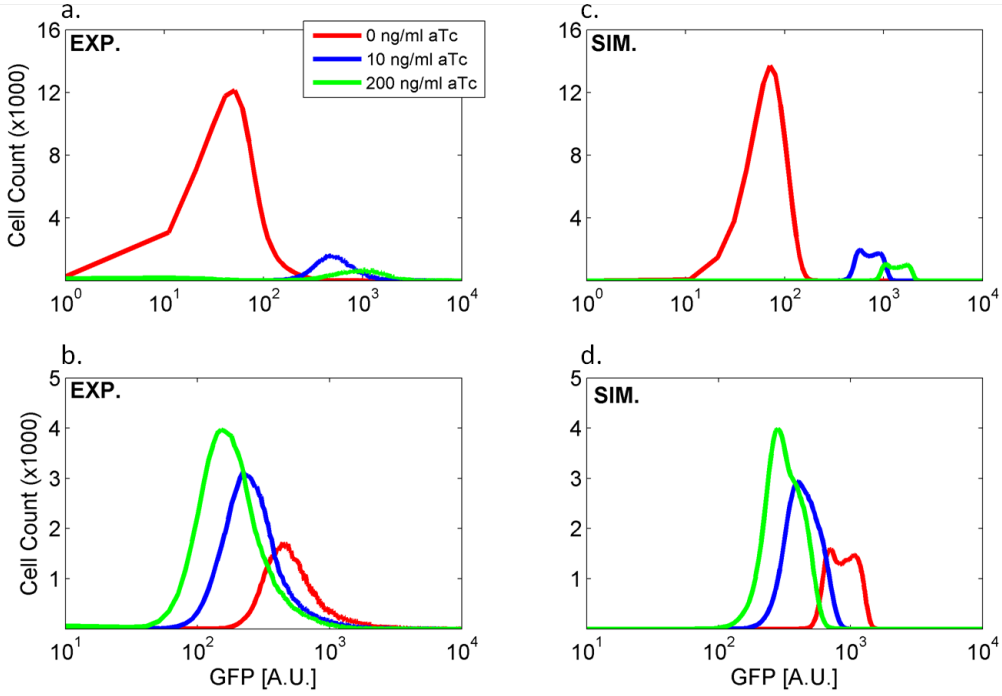

Figure 5.8: Experimental (Figures 5.8a, 5.8b) and simulated (Figures 5.8c, 5.8d) GFP distributions when 0, 10 and 200 ng/ml aTc are administered. a, c) proTeOn GFP distribution at 10 hours. b, d) proTeOff GFP distribution at 1 hour.

The protein-DNA binding strengths in the proTeOn and proTeOff systems are characterized by four kinetic parameters. These values, extracted from the models, are given in Table 5.1. Binding of aTc to PROTEON increases the affinity of the latter for the

promoter 100,000-fold. Binding of PROTEOFF to aTc leads to a decrease in the affinity of PROTEOFF to *tetO* by 10^8 times. Both PROTEON and PROTEOFF enhance gfp expression significantly when bound to*tetO* by recruiting RNApol to the promoter. In our simulations, this is realized by increasing the binding strength of RNApol to DNA. Conforming to the simulation results, binding of PROTEON to *tetO* increases the binding strength of RNApol to DNA approximately 22 times, whereas binding of PROTEOFF to *tetO* increases the binding strength around 14 times.

Table 5.1: Dissociation constants of key biomolecular interactions underlying proTeOn and proTeOff as estimated from our modeling efforts (Units: 2^{nd} order reactions: M, 3^{rd} order reactions: M^2).

Biomolecular Interactions	Binding affinity	Binding affinity
	$proTeOn$	$proTeOff$
$PROTET + tetO \rightleftharpoons PROTET : tetO$	$2.5 \cdot 10^{-5}$	$2.5 \cdot 10^{-10}$
$PROTET : aTc2 + tetO \rightleftharpoons PROTET : tetO : aTc2$	$2.5 \cdot 10^{-10}$	$2.5 \cdot 10^{-2}$
$RNApol + pro + tetO \rightleftharpoons RNApol : pro : tetO$	$3.7 \cdot 10^{-9}$	$3.7 \cdot 10^{-9}$
$RNApol + pro + PROTET : tetO \rightleftharpoons RNAp : pro : tetO : PROTET$	$1.67 \cdot 10^{-10}$	$2.56 \cdot 10^{-10}$

5.3 Conclusions

The engineered proTeOn and proTeOff systems can all be applied to tightly control gene expression with aTc in prokaryotes. Incorporating appropriate tags in the target genes' transcripts may render proTeOn and proTeOff powerful tools for cytosolic, membrane associated and secreted protein over-expression.

More broadly however, the workflow we have developed here can be implemented in the design, construction and testing of a library of variant devices and regulatory networks. Such devices include those that respond to sugars, proteins, amino acids, metabolites, toxins and other small molecules. With these, one may efficiently engineer gene expression responses as applications demand. As the collection of well characterized synthetic devices grows, we may also be able to combine them, along with naturally occurring parts, into larger gene regulatory networks to achieve more complex desired phenotypes. Feedback loops, feedforward loops, AND and OR logic gates are but a few possible networks. These devices and networks may then be used either as tools

for controlling the expression of a single gene, or as interoperable parts of larger regulatory networks that control multiple genes independently. The current proTeOn and proTeOff devices, and the kinetic and structural details that have been identified for each, are a firm stepping stone from which this work can expand. Using our workflow, integrating experimental synthetic biology, molecular modeling, and stochastic reaction kinetic simulations, the required effort and expense often associated with this kind of increase in system complexity will also be reduced. Molecular devices that can tunably regulate the expression of a single, or a handful of, genes in response to sugar, protein, amino acid, metabolite, toxin and other small molecule levels may provide biology research fields and the bioengineering industry the tools they need for efficient, diverse gene expression control.

5.4 Methods

5.4.1 PROTEON and PROTEOFF parts

PROTEON and PROTEOFF are composed of an inducible DNA binding domain and a DNA binding transcription activator, connected by a linker peptide. The reverse tetracycline repressor (rTetR) is the N-terminal, inducible DNA binding domain in PROTEON while the tetracycline repressor (TetR) is in PROTEOFF, rendering both systems responsive to aTc [87, 99, 85, 174, 179, 180]. LuxRΔN(2-162) (LuxRΔN) is the DNA-binding activator at the C-terminus of both synthetic proteins (22). LuxRΔN is the C-terminal domain of the full length LuxR transactivator. It lacks N-terminal residues 2-162 of the full length LuxR and possesses strong constitutive transactivator activity [175]. In both systems, the two domains are connected by a 150 $\mathring{A}$, 55 amino acid, peptide linker (its design is discussed below).

5.4.2 proTeOn and proTeOff synthetic promoter parts

The proTeOn and proTeOff synthetic promoter is composed of sequences from the tet and lux operons' promoters. Operator sites for TetR/rTetR [87, 85, 180] and LuxRΔN [176, 177, 180], RNApolymerase (RNApol) binding sequences (the UP element and -10 region) [178], transcription and translation start sites, an mRNA stabilizing sequence

and a ribosomal binding site [185] are included. These sequences are annotated in Figure 5.2a. GFP mutant 3 is the reporter molecule used to monitor the systems' behavior and is under the control of the synthetic promoter [186]. A second synthetic promoter, containing the RBS of the *lux* promoter and lacking the mRNA stability sequence was also tested. This promoter provided *gfp* transcript increase upon induction. However, a parallel increase in GFP protein was not observed.

5.4.3 Molecular modeling

We designed the proTeOn and proTeOff systems using molecular modeling with MOE and with NAMD [181]. We built models of the systems utilizing the known structure of the inducible DNA binding domain (rTetR or TetR) bound to the *tetO* operator (PDB code 1QPI) [99], and the transcription activator domain (LuxRΔN) bound to the luxbox (PDB code 1H0M) [187]. The linker size was determined by assuming the scaling of polymer's end-to-end-vector distance in a good solvent [188]. The specific sequence was then determined to minimize proteolysis in bacteria. We thus designed a 55-amino acid, linker peptide, which is expected to be linear, flexible, and hydrophilic [189].

5.4.4 System construction

Both synthetic activator genes were synthesized by GENEART and using standard molecular biology techniques, cloned into the expression vector pT7-FLAG1 (p1118 Sigma), at KPN1 restriction sites, in Top10 *E. coli* cells (C404010 Invitrogen). The final constructs are shown in Appendix A. The synthetic promoter and *gfp* [186] were synthesized by GENEART, on pMK, a pUC19 derived expression vector that is compatible in *E. coli*, high copy, and kanamycin resistant. This construct is illustrated in Appendix A. proTeOn and proTeOff are contained on these two plasmids and were transformed into chemically competent BL21(DE3)-T1 *E. coli* cells (B2935 Sigma) by heat shock for characterization.

Gene expression control by tetracycline in bacterial cultures BL21(DE3)-T1 cells containing the proTeOn and proTeOff systems were cultured in selective LB media, at 30^{o}C to facilitate temperature sensitive folding of LuxRΔN, agitating at 200 rpm.

Cultures were maintained in mid-log growth. We completed initial experiments to first confirm the solubility and stability of PROTEON and PROTEOFF and then to establish promoter specific gene regulation by aTc (rather than a general soluble protein upregulation). Basal, low, medium and high levels of the synthetic activators were maintained with 0, 0.25, 0.75 and 1 mM IPTG respectively. Each system was induced over a range of aTc concentrations, 1, 10, and 200 ng/ml. The total soluble protein was isolated from cultures with CelLyticB reagent (B7435 Sigma), separated by size on a 10% polyacrylamide gel and transferred to PVDF membrane. PROTEON was detected by primary mouse monoclonal anti-FLAG M2 antibody (F3165 Sigma), GFP by mouse monoclonal anti-GFP antibody [LGB-1] (ab291 Abcam), and loading control RNApol by mouse monoclonal anti-RNApol sigma 70 antibody [2G10] (ab12088). Biotinylated, polyclonal sheep anti-mouse secondary antibody, VECTASTAIN ABC kit (PK-4002 Vector Labs), and Amersham ECL Westernblotting detection reagent (RPN 2109) were used to complete the specific detection of each protein. Relative quantification of each protein was performed using ImageJ software, publically available at http://rsb.info.nih.gov, taking the ratio of protein of interest to RNApol density. PROTEON and PROTEOFF are soluble and stable at high intracellular concentrations, and gene regulation by aTc is specific to genes under the control of the synthetic promoter. This is discussed in Appendix A.

We characterized each system maintaining PROTEON and PROTEOFF expression with 0.75 mM IPTG and controlling the activity of each device with 0, 10, and 200 ng/ml aTc. Two sets of experiments were performed. In experiment set-up A, PROTEON and PROTEOFF production was induced overnight with 0.75 mM IPTG and 0 ng/ml aTc; at t = 0 hours cultures were treated with 0, 10, and 200 ng/ml aTc. In experiment set-up B cultures were treated overnight with 0, 10, and 200 ng/ml aTc and 0 mM IPTG; at t = 0 PROTEON and PROTEOFF expression was induced with 0.75 mM IPTG. Experiments A investigated the proTeOn and proTeOff system dynamics upon induction with aTc while experiments B provided insight on the PROTEON and PROTEOFF protein production and maturation dynamics. IPTG and aTc levels were maintained in the cultures at all t > 0 hours. Induction experiments were repeated for all inducer concentrations and time points, and reported trends were observed across all replicates.

Cell samples were collected for analysis by flow cytometry at t = 1, 2, 5, 10 and

20 hours, fixed with 4% paraformaldehyde for 30 min at room temperature, washed with 1x phosphate buffered saline (PBS) and stored in 1x PBS at 4^oC. Individual cells' GFP expression was measured by flow cytometry using a FACScalibur (BD Biosciences). 100,000 cells were investigated per sample with excitation at λex = 488 nm and subsequent fluorescence detection at λem = 530±30 nm. The cytometry data was collected using CellQuest (BD Biosciences) and analyzed using FlowJo (Tree Star) software. Each sample's healthy cell population was selected by first removing erroneous events (due to electronic noise) that fell below a minimum emission at λem = 530±30 nm, then secondly removing events that fell outside of the characteristic side-scatter and forward-scatter range for single *E. coli* cells. The differential GFP expression of the selected cells was analyzed and compared across samples.

5.4.5 Mean gene expression data analysis

All cytometry data was collected as described above. Prior to evaluating data across replicates, all GFP expression values were normalized to their corresponding (same system and time point) 0 ng/ml aTc samples. Basal expression (achieved with 0 ng/ml) from the proTeOn system is denoted with a value of one. In contrast, maximal expression (also achieved with 0 ng/ml) from the proTeOff system is denoted with a value of one. This basis is consistent for both experimental setups and through all times. Normalized GFP expression averages and standard errors were then calculated for replicates.

5.4.6 Stochastic modeling

We built computer models of proTeOn and proTeOff to further characterize the experimental behavior of each system. Hy3S, which is detailed in Chapter 2.2.3, was then used to simulate the system's behavior at the resolution of biomolecular interactions in individual cells. The simulations were carried out under a number of key assumptions and parameters that are discussed in the next section.

Model description

Small biological systems, either natural or synthetic, are subject to stochasticity [28, 58, 59, 190]. The behavior of such systems has been accurately described using detailed

stochastic modeling [191, 78, 12, 80, 5, 31]. Each step of biology's central dogma, including transcription, translation, degradation, dimerization, repression and induction are represented by biochemical reactions. The reaction network that portrays the behavior of both of our synthetic systems is outlined in Table 5.2. The kinetic parameters and the description of the species used in the model are detailed in Tables 5.3 and 5.4 respectively.

The first reaction captures the effective rate of aTc penetration into the cell. Reactions 2-9 model the binding of 2 aTc molecules to one PROTEON or PROTEOFF (PROTET) molecule [192]. In particular, reactions 2-5 capture the binding and unbinding of aTc to PROTET when the latter is free, whereas reactions 6-9 correspond to the same interactions when PROTET is bound to its operator site, *tetO*. Reactions 10-13 represent the binding of PROTET to *tetO* [192]. PROTET can bind to *tetO* either when free (reactions 10, 11) or bound to aTc (reactions 12, 13). Reactions 14-26 describe *gfp* transcription [116, 155]. *gfp* transcription occurs in three different ways since RNApol can transcribe *gfp* when *tetO* is: a) free (reactions 14-18) or b) occupied by PROTET:aTc2 (reactions 19-22) or c) occupied by PROTET (reactions 23-26). Three different sets of reactions are therefore used to describe transcription. In these 3 sets, the same approach is used. The first two reactions (14-15, 19-20, 23-24) capture the binding and unbinding of RNApol to *pro*. The binding of RNApol to *pro* is followed by the formation of an open complex between RNApol and DNA (reactions 16, 21, 25). Subsequently, RNApol releases the *pro* (reactions 17, 22, 26), allowing other RNApol to bind, and proceeds transcribing the entire *gfp* (reaction 18). The product of transcription, mRNA(*gfp*), is thereafter translated into GFP protein. This process is coded in reactions 27-29. The time needed for a nucleotide to be transcribed or for a codon to be translated is assumed to be exponential distributed. Consequently, the time required for the entire process of transcription and translation is considered to be gamma distributed [68]. Finally, reaction 30 captures the GFP maturation process whereas reactions 31 and 32 represent the effective rate at which GFP and mRNA(*gfp*) are degraded.

The initial conditions regarding the transcription and translation machinery define 100 RNApol and 100 ribosomes at t = 0. In addition, the number of synthetic plasmids in each *E. coli* is set equal to 20, inferring that there are 20 *tetO* and 20 *pro* sites.

The effective amount of PROTET that exists in each cell was considered constant (20 molecules) throughout the simulations since overnight IPTG treatment led to a constant PROTET concentration.

Model parameters

The conditions of the system (temperature, pH) are assumed to be constant throughout the simulations guaranteeing that the kinetic parameters used do not vary over time. It is postulated that the cell is a homogeneous, yet well stirred reactor. The volume of each cell is equal to 10^{-15} L, increases exponentially over time, and is halved during cell division. The cell division time was adjusted relative to experimentally observed behaviors with changes in intracellular aTc and GFP concentrations. Cell division time increases with increasing aTc as it retards cellular processes in higher concentrations [184]. In addition, cell division time increases with GFP concentration since this increase in protein production taxes the cellular machinery. The cell division rates are shown in Table 5.5. proTeOn and proTeOff are synthetic systems and many of the associated kinetic parameters do not exist in the literature. Thus, most of the parameters have values that are close to parameters of similar systems and have been adjusted to fit the experimental results. The penetration rate of aTc into the cell is considered equal to the penetration rate of tetracycline [118]. The binding of aTc to PROTET is assumed equal to the binding of aTc to rTetR as there is no available data for PROTET [192]. The binding of PROTET, either when free or bound to aTc, to the operator site, is considered almost equal to the binding of rTetR to *tetO* [192]. The affinity of RNApol for the *pro*, when *tetO* is free, is presumed similar to the affinity of RNApol for the *tet* promoter [116]. Further, the affinity of RNApol for the *pro* when it is occupied by PROTET or PROTET:aTc2 was fit, considering RNApol recruitment by LuxRΔN to the promoter, so that the results match the experimental phenotypes. The rates of transcription and translation do vary significantly throughout the cell population. In our simulations, we hypothesize average transcription and translation rates equal to 30 nucleotides and 100 codons per second. The GFP maturation rate was adjusted such that the simulation results were in agreement with experimental observations. Finally, the GFP degradation rate was adjusted for a 10 min protein half life, whereas mRNA degradation was adjusted to give 20 proteins per mRNA molecule.

Table 5.2: Reaction network of proteOn and proteOff.

#	Reaction
1	$aTcEx \xrightarrow{k_1} aTc$
2	$PROTET + aTc \underset{k_3}{\overset{k_2}{\rightleftharpoons}} PROTET{:}aTc$
3	$PROTET{:}aTc + aTc \underset{k_5}{\overset{k_4}{\rightleftharpoons}} PROTET{:}aTc2$
4	$PROTET{:}tetO + aTc \underset{k_7}{\overset{k_6}{\rightleftharpoons}} PROTET{:}tetO{:}aTc$
5	$PROTET{:}tetO{:}aTc \underset{k_9}{\overset{k_8}{\rightleftharpoons}} PROTET{:}tetO{:}aTc2$
6	$PROTET + tetO \underset{k_{11}}{\overset{k_{10}}{\rightleftharpoons}} PROTET{:}tetO$
7	$PROTET{:}aTc2 + tetO \underset{k_{13}}{\overset{k_{12}}{\rightleftharpoons}} PROTET{:}tetO{:}aTc2$
8	$RNApol + pro + tetO \underset{k_{15}}{\overset{k_{14}}{\rightleftharpoons}} RNApol{:}pro{:}tetO$
9	$RNApol{:}pro{:}tetO \xrightarrow{k_{16}} RNApol^*{:}pro{:}tetO$
10	$RNApol^*{:}pro{:}tetO \xrightarrow{k_{17}} RNApol^*{:}DNA(gfp) + pro + tetO$
11	$RNApol^*{:}DNA(gfp) \xrightarrow{k_{18}} RNApol + mRNA(gfp)$
12	$RNApol + pro + PROTET{:}tetO{:}aTc2 \underset{k_{21}}{\overset{k_{20}}{\rightleftharpoons}} RNApol{:}pro{:}PROTET{:}tetO{:}aTc2$
13	$RNApol{:}pro{:}PROTET{:}tetO{:}aTc2 \xrightarrow{k_{22}} RNApol^*{:}pro{:}PROTET{:}tetO{:}aTc2$
14	$RNApol^*{:}pro{:}PROTET{:}tetO{:}aTc2 \xrightarrow{k_{23}} RNApol^*{:}DNA(gfp) + pro + PROTET{:}tetO{:}aTc2$
15	$RNApol + pro + PROTET{:}tetO \underset{k_{25}}{\overset{k_{24}}{\rightleftharpoons}} RNApol{:}pro{:}PROTET{:}tetO$
16	$RNApol{:}pro{:}PROTET{:}tetO \xrightarrow{k_{26}} RNApol^*{:}pro{:}PROTET{:}tetO$
17	$RNApol^*{:}pro{:}PROTET{:}tetO \xrightarrow{k_{27}} RNApol^*{:}DNA(gfp) + pro + PROTET{:}tetO$
18	$Rib + mRNA(gfp) \xrightarrow{k_{28}} Rib{:}mRNA(gfp)$
19	$Rib{:}mRNA(gfp) \xrightarrow{k_{29}} Rib^*{:}mRNA(gfp) + mRNA(gfp)$
20	$Rib^*{:}mRNA(gfp) \xrightarrow{k_{30}} Rib + GFP$
21	$GFP \xrightarrow{k_{31}} GFP^*$
22	$GFP \xrightarrow{k_{32}} \emptyset$
23	$mRNA(gfp) \xrightarrow{k_{33}} \emptyset$
	$mRNA(gfp) \xrightarrow{k_{34}} \emptyset$

Table 5.3: Kinetic parameters used in the proTeOn and proTeOff model. Parameters in bold differentiate the behavior of the two systems. Values that were fit to match the experimental results have a reference denoted with *. Values derived from the literature "i" and then fit to match the experimental observations, have a reference denoted with "i*".

kinetic constant (proTeOn/proTeOff)	Ref.	kinetic constant (proTeOn/proTeOff)	Ref.	kinetic constant (proTeOn/proTeOff)	Ref.
$k_1 = 3.30 \cdot 10^{-4}$	[118]	$\boldsymbol{k_{12} = 4.00 \cdot 10^8 / 4.00 \cdot 10^3}$	[192]*	$k_{23} = 30$	[155]
$k_2 = 2.76 \cdot 10^8$	[192]	$\boldsymbol{k_{13} = 0.10/100}$	[192]*	$\boldsymbol{k_{24} = (6.00/3.90) \cdot 10^3}$	*
$k_3 = 1.00$	[192]	$k_{14} = 270$	[116]*	$k_{25} = 1.00 \cdot 10^6$	*
$k_4 = 1.38 \cdot 10^8$	[192]	$k_{15} = 1.00 \cdot 10^6$	[116]*	$k_{26} = 0.013$	[116]
$k_5 = 2.00$	[192]	$k_{16} = 0.013$	[116]	$k_{27} = 30$	[155]
$k_6 = 2.76 \cdot 10^8$	[192]	$k_{17} = 30$	[155]	$k_{28} = 1.00 \cdot 10^5$	[34]
$k_7 = 1.00$	[192]	$k_{18} = 30$	[155]	$k_{29} = 100$	[193]
$k_8 = 1.38 \cdot 10^8$	[192]	$k_{19} = 714$	[155]	$k_{30} = 100$	[193]
$k_9 = 2.00$	[192]	$\boldsymbol{k_{20} = (6.00/3.90) \cdot 10^3}$	*	$k_{31} = 238$	[193]
$\boldsymbol{k_{10} = 4.00 \cdot 10^3 / 4.00 \cdot 10^8}$	[192]*	$k_{21} = 1.00 \cdot 10^6$	*	$k_{32} = 1.00 \cdot 10^{-3}$	*
$k_{11} = 0.10$	[192]*	$k_{22} = 0.013$	[116]	$k_{33} = 1.20 \cdot 10^{-3}$	[80]
				$k_{34} = 2.00 \cdot 10^{-3}$	[80]

Table 5.4: Definition of the species participating in the reaction network of proTeOn and proTeOff.

Species	Definition
aTcEx	Extracellular anhydrotetracycline
aTc	Intracellular anhydrotetracycline
PROTET	PROTEON or PROTEOFF dimer
PROTET:aTc	PROTET bound to one aTc molecule
PROTET:aTc2	PROTET bound to two aTc molecules
PROTET:*tetO*	PROTET bound to *tetO*
PROTET:*tetO*:aTc	PROTET bound to *tetO* and one aTc molecule
PROTET:*tetO*:aTc2	PROTET bound to *tetO* and two aTc molecules
tetO	Operator site
RNApol	RNA polymerase
pro	Promoter site
RNApol:pro:*tetO*	Open complex of RNApol and pro when *tetO* is free
RNApol*:pro:*tetO*	Close complex of RNApol and pro when *tetO* is free
RNApol*:DNA(gfp)	RNApol bound to gfp
RNApol:pro:PROTET:aTc2:*tetO*	Open complex of RNApol and pro when *tetO* is occupied by PROTET:aTc2
RNApol*:pro:PROTET:aTc2:*tetO*	Close complex of RNApol and pro when *tetO* is occupied by PROTET:aTc2
RNApol:pro:*tetO*:PROTET	Open complex of RNApol and pro when *tetO* is occupied by PROTET
RNApol*:pro:*tetO*:PROTET	Close complex of RNApol and pro when *tetO* is occupied by PROTET
mRNA(gfp)	mRNA of gfp
Rib	Ribosome
Rib:mRNA(gfp)	Open complex of Rib and mRNA(gfp)
Rib*:mRNA(gfp)	Close complex of Rib and mRNA(gfp)
GFP	Unfolded GFP protein
GFP*	Folded GFP protein
gfp	gfp gene

Table 5.5: Cell division times for proTeOn and proTeOff when 0, 10 and 200 ng/ml aTc are administered.

aTc Concentration [ng/ml]	Cell division time (± 10 min)	Cell division time (± 10 min)
	$proTeOn$	$proTeOff$
0	32	57
10	59	35.5
200	72.5	39

Chapter 6

Stochastic simulations of a synthetic bacteria-yeast ecosystem

6.1 Background

Advances in the field of synthetic biology have enabled the design of engineered cells performing human-defined functions at a single cell resolution [3, 26]. These functions include but are not limited to oscillators [194, 5, 6], bistable switches [7], bio-logical gates [195, 12, 13], riboregulators [45, 37] and molecular devices that control gene expression [8, 80]. Despite this progress, several limitations still exist. A major shortcoming is the decreased robustness and the limited potential complexity of single cell functions. Thus, attention has been shifted to synthetic systems based on communication between cells, rather than individual isolated cell functionality. Cooperation among cells is largely mediated by quorum sensing [51] and may be promising for the development of cell-systems that robustly perform complex tasks [196, 15, 197]. These tasks range from cells rescuing or killing one another [20, 19, 198, 199] to cells synchronizing across a relatively long distance [21].

The potential advantage of microbial consortia compared to monocultures is two-fold. First, in contrast to monocultures, multicultures allow the different species to share

the various required synthetic functions or the different steps of a synthetic function. This function sharing decreases the burden in the metabolism of the cells significantly. Second, the sharing of different functions, or steps, among different cells potentially renders microbial consortia more suited for fine-tuning of their artificial functionality [17].

It is now clear that mathematical models can accurately capture the behavior of synthetic systems comprising of either bacterial or yeast cell strains and allowing cell-to-cell communication [19, 53, 198, 52, 14, 20, 200, 201, 21]. You and his co-workers designed a synthetic bacterial ecosystem where cell-cell communication controls cell density by inducing a killer gene in the bacteria [19]. To mathematically investigate the dynamics of this system, they coupled their experiments with a simple deterministic model. Shou et al. designed a synthetic yeast system where cell growth was dependent on successful cell-cell communication [53]. To further explain their system behavior, they used a mathematical model comprised of algebraic equations. Basu and his co-workers designed a synthetic system, composed of bacteria, that forms different patterns of differentiation, such as rings and clovers, driven by cell-cell communication via N-Acyl homoserine lactone (AHL) signals [52]. In addition to experimentally designing this system, they used a deterministic mathematical model to explore the behavior of this system. Balagadde et al. designed a synthetic bacterial ecosystem where cell-cell communication enables cells to exhibit predator-prey dynamics by either killing or rescuing one another [20]. They initially developed a deterministic model to thoroughly study the dynamics of their synthetic ecosystem and then introduced a constant noise term to their model aiming to explore the influence of the stochasticity in their system.

Even though communication between different species using non-AHL signals has been demonstrated previously [202], no synthetic ecosystem has been developed that is composed of bacteria and yeast which communicate with and benefit from each other using AHL signals. Such a microbial consortium could exhibit interesting dynamics, such as oscillatory behavior, that stem from the substantial differences (e.g. different volume, growth rate, gene expression process) between prokaryotes and eukaryotes. Here, we investigate the behavior of such a synthetic heterogeneous community using stochastic modeling. To this end, we have modeled and simulated a synthetic consortium composed of *Saccharomyces cerevisiae* (*S. cerevisiae*) and *E. coli* cells. This

synthetic ecosystem was found to exhibit intriguing dynamic behavior that is commonly observed in natural ecosystems. Our model, capturing the behavior of this ecosystem, has been built in such a way that it can capture the dynamics of any system with two different species communicating with AHL signals. Thus, our model may drive the experimental design of artificial ecosystems with two different species (e.g. mammalian-yeast or mammalian-bacteria) which communicate with and regulate gene expression in one another.

6.2 Methods

Design of the synthetic ecosystem

In this study, we propose the design of a synthetic yeast-bacteria ecosystem that is based on diffusible chemical signals. Examples of these signals are the RhII/RhlR and LuxI/LuxR quorum sensing signals from *Pseudomonas aeruiginosa* and *Vibrio fisheri* quorum sensing systems, respectively, which are known for their sensitivity and the absence of signal cross-reactivity [203].

Each species exists in the presence of a molecule controlling growth, Gc. This molecule could be an antibiotic, such as Kanamycin, which is effective against both *E. coli* and *S. cerevisiae* [204]. Gc inhibits cell growth and therefore each species ultimately goes extinct. However, each species contains a resistance gene which counteracts the function of Gc and is controlled by the other species via diffusible molecules. Thus, when both species are present, they induce each other's resistance gene through chemical signals, thereby rescuing one another. A schematic representation of the proposed ecosystem is illustrated in Figure 6.1.

More specifically, *S. cerevisiae* constitutively expresses a diffusible molecule, AHL1. AHL1 diffuses out of the *S. cerevisiae* cells, penetrating *E. coli* cells and binding to its cognate receptor, AHLR1. AHLR1 is constitutively produced in *E. coli* . The activated molecule in *E. coli* binds to the responsive element fused upstream of the Res promoter activating expression of Res. Subsequently, the resistance protein, Res, deactivates Gc in *E. coli*. Potential Res could be the Kanamycin resistance protein [204].

The second component of the feedback loop in *E. coli* functions in the same genetic fashion. It constitutively produces an autoinducer synthase, AHL2. Once AHL2 is

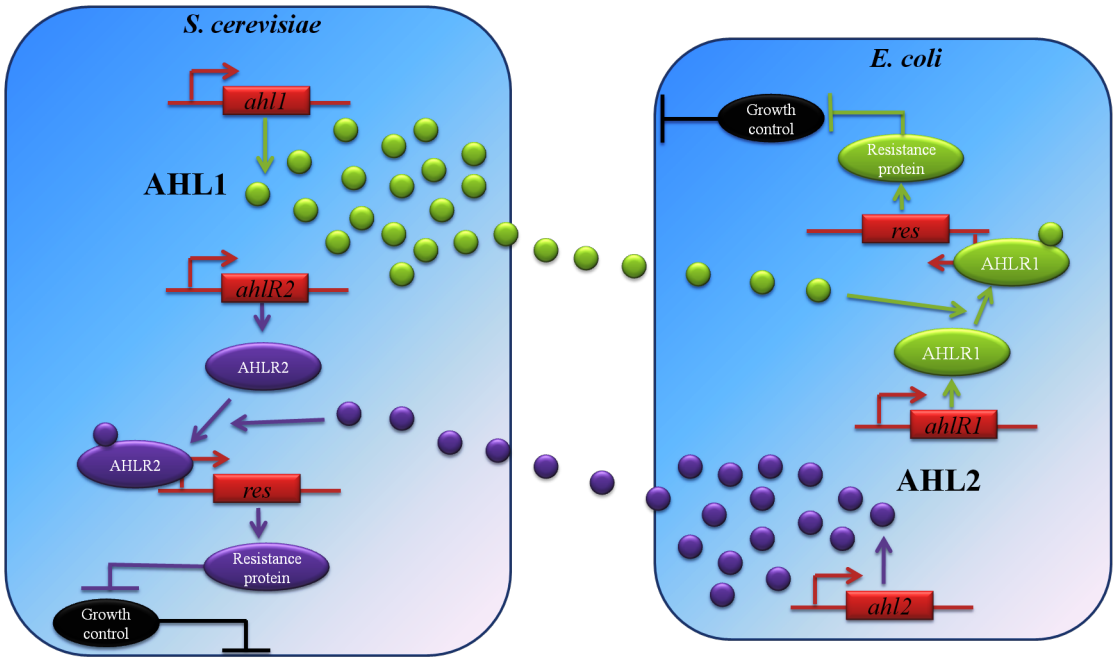

Figure 6.1: *S. cerevisiae* cells produce AHL1 thereby activating resistance gene expression in *E. coli* and cell survival. Similarly, *E. coli* cells produce AHL2 that induces resistance gene expression in *S. cerevisiae* rescuing the latter.

produced, it diffuses out of the *E. coli* into *S. cerevisiae*, and is recognized by its cognate receptor, AHLR2, which is constitutively produced in *S. cerevisiae* as a fusion protein that allows it to be activated in eukaryotic cells. This activated molecule now binds to its responsive promoter and induces expression of the resistance gene, *res*. The resistance protein, in turn, represses the function of Gc in *S. cerevisiae*.

It is important to note that for the purposes of this study, we assume synthetic bacterial molecular components function in yeast. We hypothesize that their functionality may be retained when they are used in yeast. This is not an unreasonable hypothesis since the functionality of quorum sensing bacterial elements has been demonstrated experimentally in other higher organisms [205].

Here, we aim to computationally explore the behavior of a microbial consortium consisting of two different species, and how the differences of the two species affect its dynamics. The focus is therefore on the population dynamics. The functionality of such

an ecosystem could in principle be achieved using any other molecular components with similar function.

6.2.1 Model description

As discussed in the previous section, numerous mathematical models that describe the behavior of synthetic ecosystems have been developed previously [19, 53, 198, 52, 14, 20, 200, 201]. The vast majority of these models are deterministic, ignoring the stochastic nature which is ubiquitous in biological systems [58, 28, 206]. Thus far, different methods have been described [67, 207, 55, 31, 68, 208] and extensively applied to stochastically simulate the dynamics of biological systems in general and gene networks in particular [34, 80, 12, 5, 8, 209, 29, 210, 211].

In this study, we develop a stochastic model that accounts for the intrinsic and extrinsic noise and describes the dynamics of the synthetic bacteria-yeast ecosystem depicted in Figure 6.1. The model takes into consideration the volume and the growth rate differences between *E. coli* and *S. cerevisiae*. In addition, it accounts for the gene expression dissimilarities between bacteria and yeast. Our model monitors the evolution of molecular species that usually exist in relatively high amounts allowing for the use of continuous stochastic models [55, 208]. Continuous stochastic computational approaches have also accurately described the experimental phenotype of synthetic cell communities [20]. We, therefore, employ chemical Langevin equations [208] to capture the evolution of the species participating in this synthetic ecosystem.

The model consists of 17 reactions (Table 6.4) whose dynamics are described using 9 Stochastic Differential Equations (Appendix B). The equations were integrated in Matlab using the Euler Maruyama method [212]. The type of reactions as well as the kinetic parameters used were acquired from previously published studies involving experimental work. Our model is generic (i.e it may be used to capture the dynamics of various two-species ecosystems), but for the purposes of this study we assumed specific molecular components (and their associated kinetic parameters) that have been widely used in designing synthetic ecosystems. These components are presented in Table 6.1. The current model may capture the behavior of any similar heterogeneous ecosystem by simply modifying the kinetic parameters according to the new system. The reaction network along with the kinetic parameters and the reaction rates capturing the behavior

of our system is presented in Table 6.4.

Table 6.1: Molecular components assumed in the model.

Name	Molecular Component
ahlI	*rhlI*
ahlR1	*rhlR*
AHL1	C4HSL
AHLR1	RhlR
ahl2	*luxI*
ahlR2	*luxR*
AHL2	3-oxo-C6HSL
AHLR2	LuxR
res	*kanR*
Resistance protein (Res)	Kanamycin resistance
Growth control (Gc)	Kanamycin

The first two reactions describe the cell population growth. Consistent with previous mathematical models [19, 20, 198], and because the model refers to ecology, population growth follows logistic kinetics. Bacteria were considered to grow four times faster than yeast [213]; k_1 was set four times smaller than k_2. C_{max} represents the carrying capacity of the bioreactor, i.e. the maximal population load that the bioreactor can sustain [214], and is set equal to 10^9 cells [199, 20]. Reactions 3 and 4 represent the cell death due to the presence of Gc (in our case Kanamycin). We assume a constant concentration of Gc (0.3 μM) as, according to the kinetic parameters used in our model, this concentration kills each single simulated cell colony when the two species are placed separately. Both the bacteria and yeast carry a resistance gene so the reaction rate is written such that the higher the amount of the resistance protein, the slower the cell death rate is. Similar reaction rates have been used previously to capture cell death due to killer proteins [20]. The correlation between Gc and the resistance protein is tuned through the parameter α. The parameter α was initially set equal to $5 \cdot 10^4$ molecules^{-1}, due to the lack of literature values, and subsequently the sensitivity of the ecosystem's behavior to changes in this parameter was investigated. Reactions 5 and 6 describe the production of the molecules responsible for the diffusible signals. AHL1 and AHLR2 are produced by *S. cerevisiae* whereas AHL2 and AHLR1 are produced by *E. coli*. The concentration of AHLR2 and AHLR1 is considered constant (0.5 μM) and equal to previously published values [14]. AHL1 and AHL2 production reactions

are assumed to be first order, in accordance with previous studies [198, 215, 19]. The production rate of these diffusible molecules can vary significantly depending on the promoter strength of the associated genes. Using directed evolution, a wide range of quorum sensing production rates can be achieved [216]. The optimized behavior can be also achieved using computational approaches [217]. In our model, we initially adopted k_4 and k_5 from [20] and subsequently increased their values since our system required very long time to reach steady state under these conditions. Reaction 7 captures the binding of AHL2 to AHLR2 in *S. cerevisiae*. This reaction is considered a fourth order reaction (this reaction accounts for the volume of *S. cerevisiae* cells) since it has been demonstrated that a fourth order reaction can capture the experimental phenotypes well [14]. Resistance protein (Res) production is calculated using Hill type kinetics, in accordance with experimental observations [14], and is shown in reaction 8. This reaction also accounts for gene expression differences between eukaryotes and prokaryotes. In contrast to prokaryotes, eukaryotic transcription requires many transcription factors to be recruited before its initiation. Moreover, the translation process in prokaryotes is faster than in eukaryotes [218]. These two factors introduce a delay in eukaryotic gene expression rendering it slow compared to the prokaryotic gene expression. In our model, we represent this delay using reaction 8. In fact, we assume that a complex (preRes) must first be formed before Res production can take place. A similar approach has been used previously to capture transcription in yeast [219]. After performing a set of simulations, we set k_8 equal to 5 h^{-1} since this value was found to cause a delay in our ecosystem compared to a model lacking this intermediate reaction (data not shown). The actual process of protein production is captured by reaction 9. Similarly to reaction 7, reaction 10 captures the binding of AHL1 to AHLR1 in *E. coli* (this reaction accounts for the *E. coli* cell volume). Reaction 11 is used to describe *E. coli* gene expression. Note that in this case there is no reaction describing a delay in gene expression. Finally, reactions 12-17 represent the degradation of the species participating in this network and they are all considered first-order.

6.3 Results and Discussion

6.3.1 Testing synthetic ecosystem's functionality

Initially, we explored whether *S. cerevisiae* cells can withstand Gc in the absence of *E. coli* cells and vice versa. We simulated the behavior of 50,000 yeast cells and 50,000 bacterial cells in the absence of Gc. Our simulations indicated that both *S. cerevisiae* and *E. coli* grow normally (data not shown). However, when each of the two populations is placed in a simulated bioreactor separately, in the presence of Gc, neither population is able to survive (data not shown).

Subsequently, we simulated the behavior of bacteria and yeast when both are present to test whether communication and cooperation between these two species can be successfully achieved. We simulated stochastically 100 different population colonies containing 50,000 *E. coli* and 50,000 *S. cerevisiae* cells. The results are depicted in Figures 6.2A (yeast cells) and 6.2B (bacterial cells). The different lines correspond to the cell population size in different trajectories.

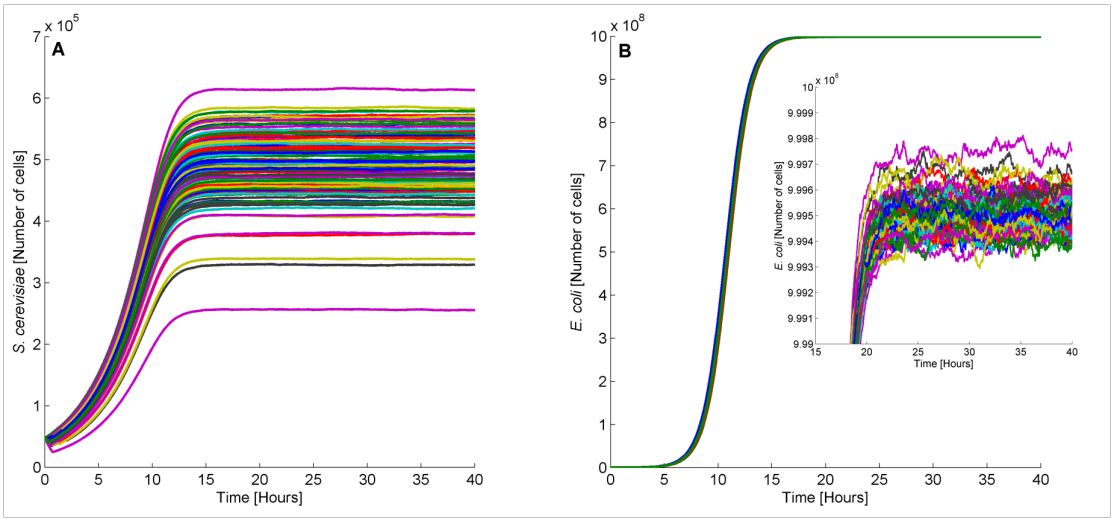

Figure 6.2: When the two species are placed together, obligatory mutualism is observed, i.e they benefit from each other and survive from Gc. The inset represents part of Figure 2B and shows the fluctuations of *E. coli* population size.

In both cases, the two different species exploit communication with one another

for successful survival in the presence of Gc. Our simulations demonstrate that yeast can successfully induce the expression of the resistance gene found in bacteria and vice versa. This is a common characteristic of ecosystems called obligatory mutualism. In other words, *S. cerevisiae* and *E. coli* cells are not able to survive separately but they are able to grow in concert. As expected, we observe that the number of *E. coli* cells is always higher than the number of *S. cerevisiae* cells. As discussed before, the reason for this is the high growth rate of bacteria relative to yeast.

Variation regarding the number of cells is observed when different colonies are simulated and is attributed to the stochasticity underlying biological functions. The average *S. cerevisiae* population (calculated over 100 trajectories) is around $4.90 \cdot 10^5$ cells and the standard deviation is equal to $5.77 \cdot 10^4$ cells. The mean *E. coli* population is about $9.99 \cdot 10^8$ cells and the standard deviation is approximately $6.68 \cdot 10^4$ cells. Note that the total number of cells cannot exceed 10^9. Both bacteria and yeast require approximately 16 hours to reach steady state. In every single bioreactor, neither bacteria nor yeast die from the presence of Gc. This demonstrates that communication can take place between *S. cerevisiae* and *E. coli* allowing for the survival of the two species.

Even though Figure 6.2 establishes cell communication and obligatory mutualism between *E. coli* and *S. cerevisiae* cells, this refers only to the case described by this set of parameters. Thus, in order to investigate which parameters promote successful communication and cooperation between *E. coli* and *S. cerevisiae* cells, and to explore the dynamics of different parameter sets, a sensitivity analysis was performed. To implement this, we systematically modified different parameters within reasonable ranges and monitored the dynamics of the system. In what follows, we present the evolution of the average *S. cerevisiae* and *E. coli* population over 100 trajectories. In some cases, we further provide all the 100 trajectories with variation in the values of key parameters examined in our analysis.

6.3.2 Ecosystem's sensitivity to parameter α

As discussed in the previous section, α represents a correlation between the molecule controlling growth and the resistance protein. More specifically, the larger the α, the smaller the amount of Res required for cells to survive from Gc (see Table 6.4). Since this parameter is of high importance for our model, and because it was the only parameter

not acquired from previously published models, we explored the influence of α on the system's behavior. To this end, we performed multiple computational experiments modifying α and investigating our ecosystem's dynamic behavior. Our simulation results showed that when α is larger than 100 nM^{-1}, the total system's behavior does not change appreciably (data not shown). For values of α smaller than 0.07 nM^{-1}, the ecosystem is driven to extinction (data not shown). Importantly, our simulations' data demonstrated that for values of α in the range of 0.07 nM^{-1} to 100 nM^{-1}, the dynamics of the system, and specifically the time the system needs to reach steady state, becomes remarkably slow. Figures 6.3A and 6.3B show the mean values, along with the standard deviation, of 100 trajectories from the stochastic simulations for α equal to 25 nM^{-1} (red), 75 nM^{-1}(green) and $5 \cdot 10^4$ nM^{-1}(blue). As observed in Figure 6.3, when the value of α is lower than 100, the system cannot reach steady state even after 10,000 hours. To our knowledge, no synthetic ecosystem exists that reaches steady state after such a long time suggesting that in order for this system to be realistic, the value of α in our model should be higher than 100 nM^{-1}. This is confirmed by the fact that our system reaches steady state approximately as fast as previously published synthetic ecosystems systems did [198, 20, 19].

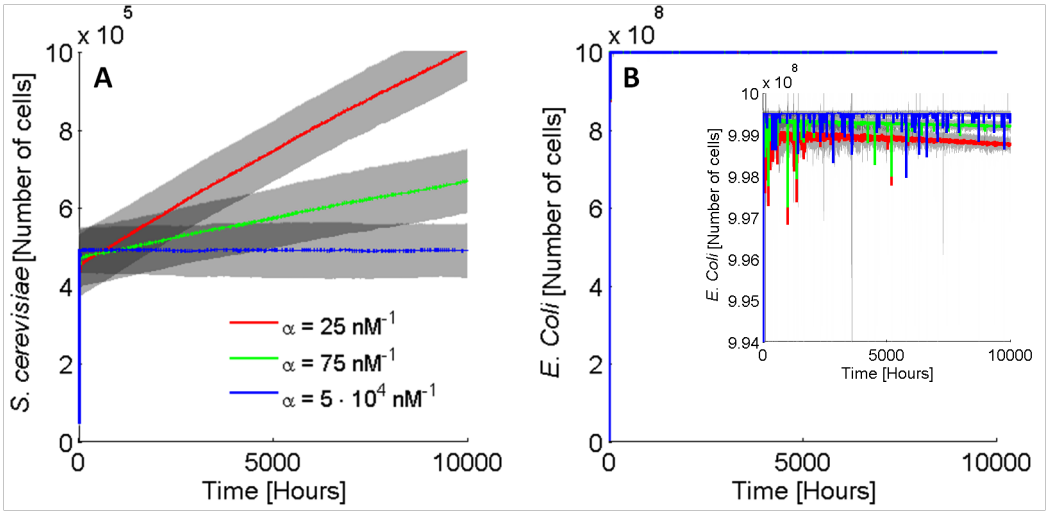

Figure 6.3: Mean values and standard deviation (grey shade) of 100 trajectories of *S. cerevisiae* (A) and *E. coli* (B) population size for different values of the parameter α.

6.3.3 Importance of Gc Concentration

Previous studies describing similar synthetic ecosystems have demonstrated the importance of the concentration of the molecule controlling growth on the system's dynamics [198]. Guided by this, we conducted a set of simulations where we modified Gc's concentration. We monitored the dynamics of the system for three different Gc concentrations. The average population values (over 100 trajectories) for each concentration are shown in Figures 6.4A and 6.4B. 100 trajectories of the two species population for the different Gc concentrations are provided in Figures 6.4C and 6.4D.

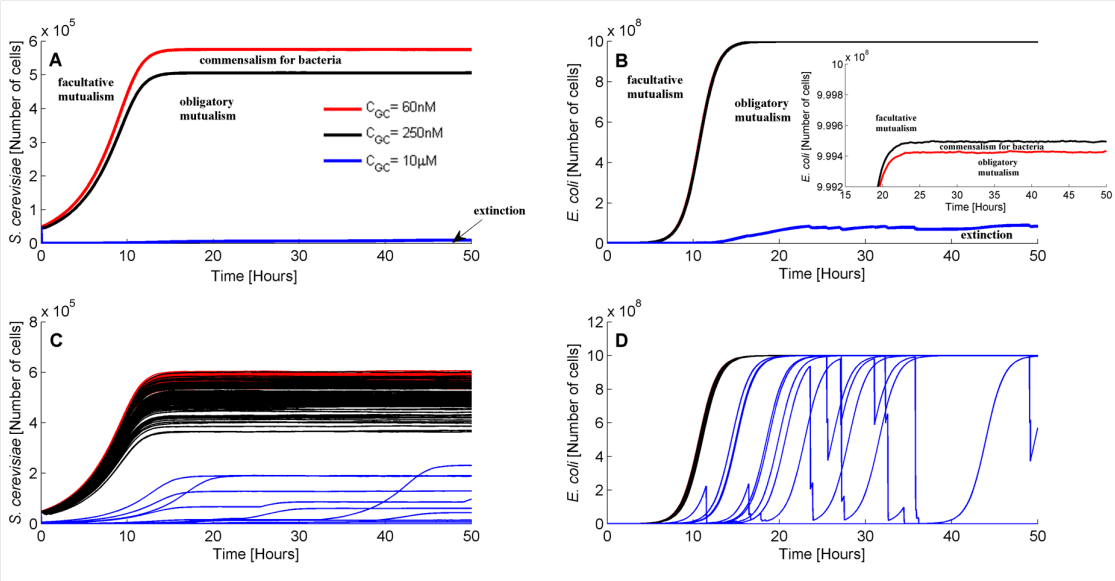

Figure 6.4: Average (over 100 trajectories) values (A,B) and 100 single trajectories (C,D) of *S. cerevisiae* and *E. coli* population size for Gc concentration equal to 60 nM (red), 250 nM (black) and 10 μM (blue). The synthetic ecosystem adopts different behaviors, that are commonly observed in natural ecosystems, in response to different Gc concentrations.

As Figure 6.4 indicates, an increase on Gc's concentration from 60 nM to 250 nM is followed by a decreased yeast population and an increased bacterial population. In other words, upon increasing Gc concentration in the bioreactor, *E. coli* cells benefit whereas *S. cerevisiae* cells are harmed. Based on the way our model was built, this is likely ascribed to the fact that yeast grow much slower than bacteria and can therefore

resist only low Gc concentrations. As the antibiotic concentration increases, yeast die faster than bacteria and the latter, even though they grow slower than they would in the absence of Gc, take advantage of the higher nutrient levels in the bioreactor. This is an interesting characteristic and could be used as a means for controlling the bioreactor's population, obviating the need of adding or removing cells. However, when Gc concentration is significantly high (e.g. 10 μM), the average value (of the 100 trajectories) of both populations decreases dramatically as many single trajectories reach zero.

Further analysis of the system's behavior indicated that changing the Gc concentration leads to an intriguing behavior commonly exhibited by natural ecosystems. More specifically, when Gc levels are low, each species can survive even in the absence of the other species. In particular, bacterial cells can withstand up to 250 nM Gc. On the other hand, yeast cells cannot survive even these Gc levels and they can only withstand Gc concentrations lower than 60 nM. Having said this, the behavior of the system for Gc levels up to 60 nM is analogous to facultative mutualism, i.e. both species benefit from but are not dependent on each other. However, when Gc's level lies between 60 nM and 250 nM, the behavior of the system is similar to commensalism for bacteria, i.e. bacteria can survive without yeast but yeast are not able to survive without bacteria.

The lethal Gc concentration for cultures with both cell types present is 20 μM. Based on this, we conclude that when Gc's levels are between 250 nM and 20 μM, the behavior of the system is homologous to obligatory mutualism as both species are completely dependent on each other and unable to survive individually. Finally, for Gc levels higher than 20 μM, we observe ecosystem's extinction. Such behaviors have been observed previously in similar synthetic bacterial ecosystems [198] and are shown in Figure 6.4. The concentrations used in Figure 6.4 represent the boundaries between different system's behavior (note that instead of 20 μM Gc, which is the boundary between obligatory mutualism and extinction, we considered 10 μM Gc). The population dynamic behavior for Gc concentrations between the ones used here lies in the area between these lines.

Figures 6.4C and 6.4D demonstrate deviation among the different cell density trajectories. Note that this deviation could not be captured using deterministic simulations. The standard deviation (calculated over the 100 trajectories) of the population size at 50 hours and for 60 nM, 250 nM, and 20 μM Gc is shown in Table 6.2.

Table 6.2: Standard deviation of population size at steady state for different Gc concentrations.

Gc [nM]	*S. cerevisiae* [10^4 cells]	*E. coli* [10^4 cells]
60	1.42	3.55
250	4.93	5.61
$2 \cdot 10^4$	3.87	2.77

6.3.4 Ecosystem's sensitivity to various carrying capacities and initial cell densities

We then explored the influence of c_{max} and the initial cell population on the synthetic ecosystem's behavior. As discussed above, the carrying capacity is the maximum number of (bacterial and yeast) cells that can exist in the bioreactor [214]. Here, we only show average values of the 100 trials since the single trajectories exhibit similar behavior as in the previous cases. Figures 6.5A (*S. cerevisiae*) and 6.5B (*E. coli*) show average population sizes for different c_{max} values.

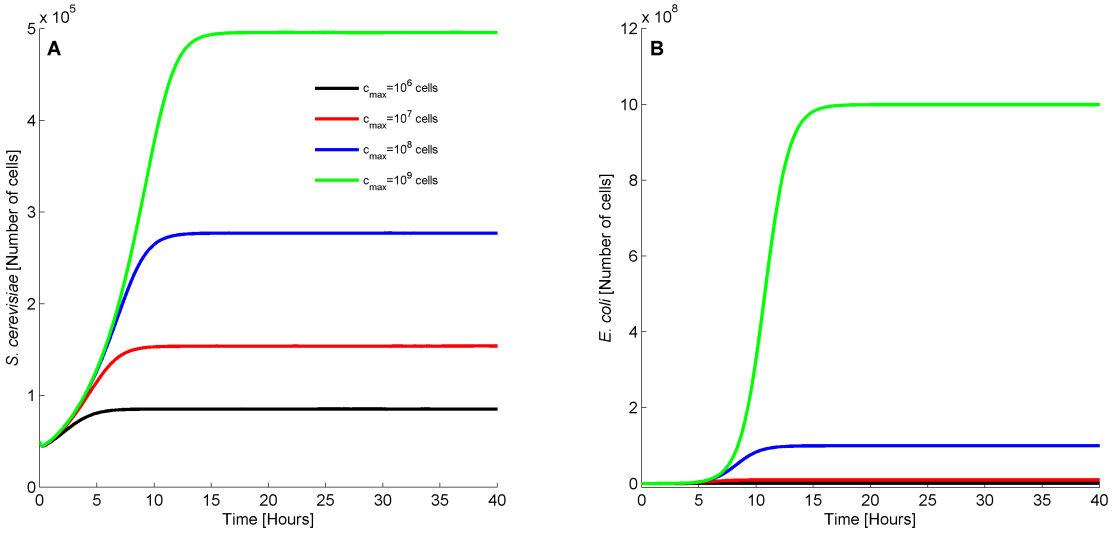

Figure 6.5: Average (over 100 trajectories) values of *S. cerevisiae* (A) and *E. coli* (B) population size for different reactor capacities. The higher the reactor capacity the higher the steady state population density of the two species is.

As expected, an increase in c_{max} causes an increase on both yeast and bacterial

steady state populations as the nutrients in the culture suffice for more cells. Thus, both species grow faster and consequently survive in the presence of Gc. It is important to note that a minimum amount of nutrients must exist in the bioreactor for the cells to grow and survive. Thus, we ran simulations decreasing c_{max} to find this minimum threshold under which the ecosystem goes extinct. According to our simulation results, the minimum c_{max} in order for all the trajectories to end up in non-zero steady states (over a period of 3,000 hours) is equal to approximately $2 \cdot 10^5$ cells (data not shown). Thus, the model suggests that our synthetic ecosystem is fully functional only for reactor capacities equal to or higher than $2 \cdot 10^5$ cells.

As in the previous cases, deviation among the different population trajectories was observed. The steady state standard deviation of the population size for various reactor capacities is provided in Table 6.3. Notably, the larger the reactor capacity, the higher is the deviation among the different population trajectories.

Table 6.3: Standard deviation of population size at steady state for different reactor capacities.

C_{max} [cells]	*S. cerevisiae* [10^4 cells]	*E. coli* [10^4 cells]
10^6	1.14	1.14
10^7	1.96	1.91
10^8	3.20	3.10
10^9	5.73	6.60

We further explored the minimum initial number of total cells required in order for the two species population to cooperate favorably and survive. To do so, we performed different simulations starting with equal *E. coli* and *S. cerevisiae* populations and monitoring the system's dynamics for 1,000 hours. Our results showed that for equal initial populations of the two species, the minimum number of *S. cerevisiae* and *E. coli* cells in the reactor should be approximately equal to 15 cells for the ecosystem to survive with Gc. Moreover, when the initial *E. coli* population is 50,000 cells, the minimum *S. cerevisiae* initial population required in order for the system to avoid extinction is 14 cells. On the other hand, when the initial *S. cerevisiae* population is 50,000 cells, the required *E. coli* initial population is 4 cells. This difference is ascribed to the fact that bacteria grow predominantly fast thereby quickly helping yeast to survive and therefore only 4 yeast cells are initially required to make the ecosystem functional. However, yeast grow and consequently rescue *E. coli* with a slower rate and therefore larger *E.*

coli population is initially required for the ecosystem to function.

6.3.5 Effects of *E. coli* cell death rates on the ecosystem's dynamics

It is clear from the aforementioned analysis that in most cases bacterial cell populations dominate yeast cell populations because of their high growth rate. We therefore introduced a bacteria degradation term in our network to enhance the competition between the population of the two species. We only considered *E. coli* degradation as bacteria grow significantly faster than yeast. This degradation could be achieved experimentally as bacteria can be engineered to stimulate their lysis in response to a human-defined signal. More specifically, introducing holin and lysozome genes that are activated via AHL signals, allows for controlling cell membrane destruction and consequently cell death [220].

Initially, we performed our analysis under the assumption that the deterministic term dominates the stochastic term, ie. the intrinsic noise of the system is negligible. The results presented in what follows were therefore produced based only on the deterministic part of the equations 1-9. A similar approach has been used previously to explore the oscillatory behavior of a synthetic ecosystem [20].

As expected, high degradation rates cause bacterial cell death followed by yeast wash out due to obligatory mutualism (data not shown). In contrast, low degradation rates allow yeast domination, as bacterial populations quickly decreases due to both Gc and degradation, thereby allowing an increase in yeast population (data not shown).

Importantly, and as observed in other synthetic ecosystems composed of species with different growth rates [20], there is a range of bacterial degradation rate where *S. cerevisiae* and *E. coli* population exhibits sustained oscillations. These oscillations originate from the antagonism between the two species population and demonstrate a predator-prey like relationship between *S. cerevisiae* and *E. coli* cells. In particular, when the *E. coli* degradation rate, d, lies between 0.30 h^{-1} and 0.72 h^{-1}, sustained oscillations are observed. For d smaller than 0.30 h^{-1}, damped oscillations are exhibited. Finally, for d larger than 0.72 h^{-1}, the two species population goes to zero. In other words, the ecosystem becomes extinct, since this high degradation rate results in bacterial death which in turn leads to yeast extinction because cell-cell communication cannot take place favorably anymore. Figure 6.6 shows the behavior of the two species

population for degradation parameters that lie in the aforementioned ranges. When d is equal to 0.25 h^{-1} (Figures 6.6A,6.6B), we observe damped oscillations that end up on a stable steady state. However, from d=0.30 h^{-1} to d=0.72 h^{-1}, sustained oscillations whose amplitude scales with the degradation rate are observed. This trend is provided in Figures 6.6C,6.6D where d=0.50 h^{-1}. Finally, when d is larger than 0.72 h^{-1} the system is driven to extinction, as depicted in Figures 6.6E and 6.6F.

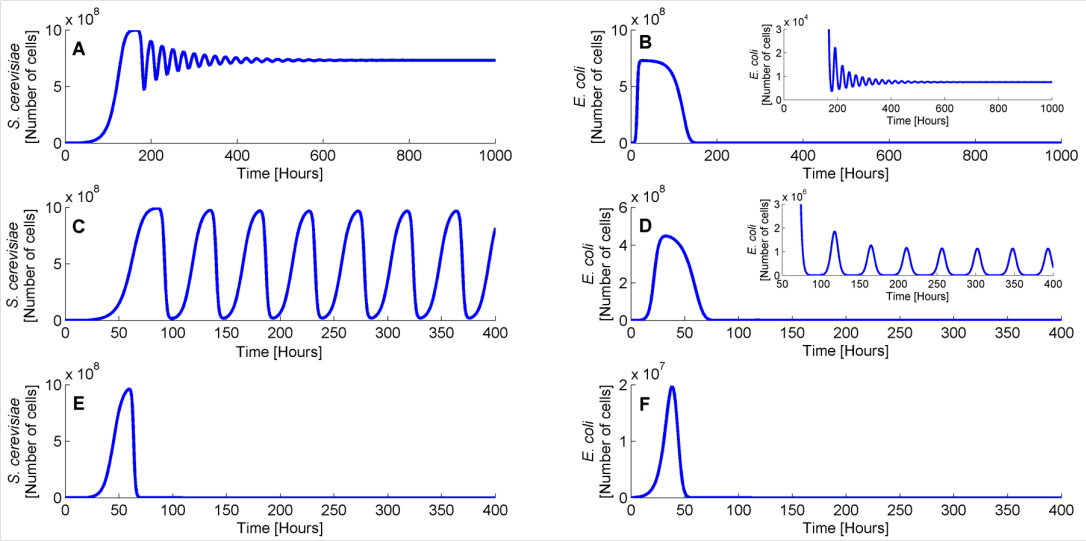

Figure 6.6: *S. cerevisiae* (A,C,E) and *E. coli* (B,D,F) population dynamics for *E. coli* degradation rate equal to 0.25, 0.50 and 0.75 h^{-1}. For d=0.25 h^{-1} (A,B) the ecosystem exhibits damped oscillations. For d=0.50 h^{-1} (C,D) the population of the two species oscillates with sustained oscillations whereas for d=0.75 h^{-1} (E,F) goes to zero.

The bifurcation diagram describing our ecosystem's oscillatory behavior is presented in Figure 6.7. A Hopf point, where sustained oscillations of the two species population initiate, is observed for d approximately equal to 0.30 h^{-1} (red). The Hopf point was further confirmed by eigenvalue analysis. The lines following the Hopf point correspond to the oscillation amplitude, as calculated from the transient analysis. Please note that for the sake of clarity, in Figure 6.7B we show the upper limit of oscillations only for d up to 0.41 h^{-1}. The complete bifurcation diagram is provided in the inset. The period of the oscillations was calculated using the FFT (Fast Fourier Transform) function in Matlab and is depicted as inset in Figure 6.7A. As evident, the period of the oscillations

scales with the *E. coli* degradation rate. This is an intriguing observation which suggests that the *E. coli* degradation rate could be used to control the period of our oscillatory ecosystem.

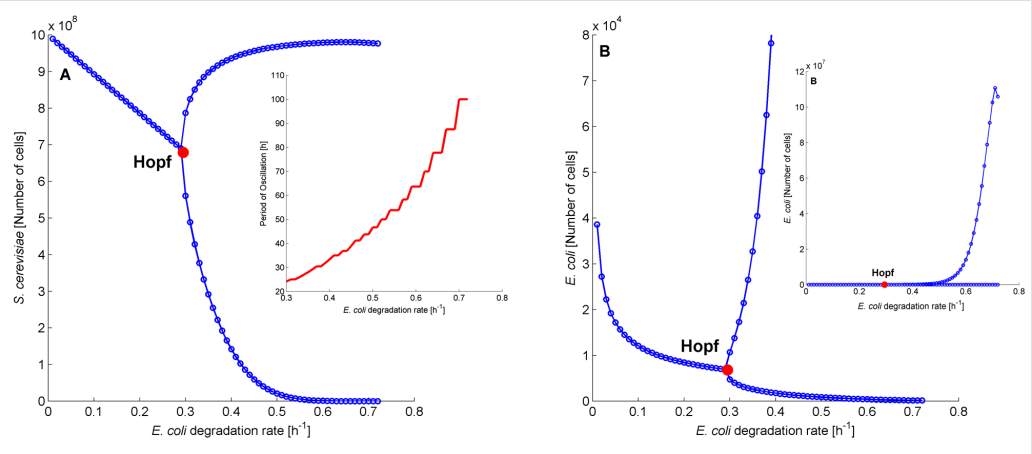

Figure 6.7: Bifurcation diagram of the *S. cerevisiae* (A) and *E. coli* (B) population versus the degradation rate of *E. coli* cells. For the sake of clarity, Figure B shows only part of the bifurcation diagram whereas the complete bifurcation diagram is illustrated in the inset. The period of oscillation of *S. cerevisiae* and *E. coli* cells for different *E. coli* degradation rates is the same and presented as inset in Figure 6.7A.

It should be stressed that including the stochastic terms in our simulations, leads to the ecosystem's extinction. This has been observed before [20] and is caused by the fact that during the oscillations, the bacterial population reaches small values and therefore noise terms destroy the sustained oscillations by driving bacterial population to zero and consequently the ecosystem to extinction (since cooperation cannot occur). In fact, the smaller the noise amplitude, the higher the probability for the system to circumvent extinction and exhibit sustained oscillations. Figure 6.8 shows a comparison between deterministic and stochastic simulations. For d=0.50 h^{-1} (Figures 6.8A and 6.8B), deterministic solution (black) provides sustained oscillations that end up in a steady state. Stochastic simulations (red) are consistent with the deterministic ones, i.e. demonstrate oscillations, but only for a small period of time and subsequently all the trajectories reach zero. Motivated by this observation, we performed several simulations (for d=0.50 h^{-1}) where we systematically decreased the noise terms amplitude.

Our simulations demonstrated that when the noise terms are 1.25% or less of the current values, the stochastic behavior matches the deterministic one, i.e. the ecosystem population exhibits sustained oscillations. When the noise terms are between 1.30% and 100% of the current values, there are always stochastic trajectories that reach zero over a period of 1,000 hours. Figures 6.8B and 6.8C show three population density trajectories of the stochastic simulation (green, red, blue) compared with the deterministic simulation (black) when the noise terms are reduced to 1.25%. As evident, the ecosystem's behavior provided by the two approaches is consistent and stochastic trajectories exhibit continuous oscillations.

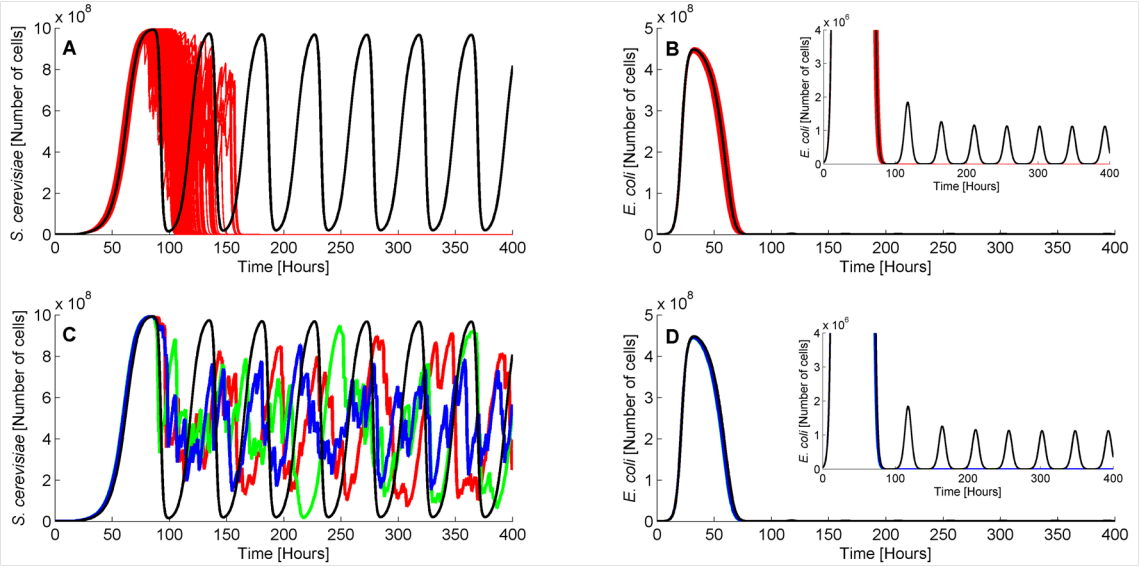

Figure 6.8: A,B: Population size of *S. cerevisiae* (A) and *E. coli* (B) for d=0.50 h^{-1} calculated using stochastic (red) and deterministic simulations (black). C,D: Population size of *S. cerevisiae* (C) and *E. coli* (D) for d=0.50 h^{-1} calculated with deterministic (black) and stochastic (red, green, blue) simulations with 1.25% of the current intrinsic noise terms.

Overall, our simulations suggest that high amplitude intrinsic noise damages the ecosystem's oscillatory behavior. On the other hand, less noisy environments stimulate the sustained oscillation of the two species population.

6.4 Conclusions

We presented the *in silico* design of the first synthetic bacterial-yeast ecosystem where communication between cells is achieved using AHL signals. The model, while developed to accurately depict these interactions, can be adapted to characterize any cell-to-cell communication and population dynamics mediated by diffusible chemical signaling.

We showed that when the two species coexist, they overcome Gc's toxicity by inducing each other's resistance gene via small molecule signalling and therefore survive. Our simulations suggest that the minimum reactor capacity required for this ecosystem to evolve is $2 \cdot 10^5$ cells. By varying the Gc concentration, the ecosystem adopts different behaviors including obligatory and facultative mutualism, commensalism and extinction. Adding an *E. coli* degradation reaction, which can be experimentally realized by engineering bacteria to induce lysis, can drive the population of the two species to predator-prey like dynamics, i.e. sustained oscillations. These oscillations can, however, be destroyed in noisy environments. Overall, we demonstrated that such kind of heterogeneous synthetic ecosystems could exhibit interesting dynamics.

As demonstrated here and in different studies [20], the development of synthetic microbial consortia using species with different characteristics (e.g. different growth rate or volume) yields systems with intriguing dynamics, such as oscillations. These systems could have various potential applications such as the delivery of two different drugs in dissimilar time intervals [17].

Our mathematical model may potentially drive the experimental design of microbial consortia with a heterogeneous population. This and similar mathematical models can further be used to predict interspecies bioreactor dynamics under numerous conditions, with differing chemical signals, and employing various population control mechanisms. Engineered interspecies system have substantial implications for complex chemical synthesis as well as future biorefinery design and optimization. Thus, the dynamics analysis presented herein may be used as the basis for the *in vivo* design of such promising synthetic ecosystems.

Table 6.4: Reaction network capturing synthetic ecosystem's behavior.

#	Reaction	Reaction rate	Kinetic constant
1	$\emptyset \xrightarrow{k_1} c_1$	$k_1 \cdot c_1\left(1-\frac{c_1+c_2}{c_{max}}\right)$	$k_1 = \frac{0.234}{h}, c_{max} = 10^9 cells$[20, 199]
2	$\emptyset \xrightarrow{k_2} c_2$	$k_2 \cdot c_2\left(1-\frac{c_1+c_2}{c_{max}}\right)$	$k_2 = \frac{0.936}{h}$ [19],$c_{max} = 10^9 cells$[20, 199]
3	$c_1 + Gc \xrightarrow{k_3} Gc$	$\frac{k'_3 \cdot c_1}{1+\alpha \cdot Res1}$	$k_3 = \frac{4 \cdot 10^6}{M \cdot h}$ [19],$\alpha = \frac{5 \cdot 10^4}{Molecules}$
4	$c_2 + Gc \xrightarrow{k_3} Gc$	$\frac{k'_3 \cdot c_2}{1+\alpha \cdot Res2}$	$k_3 = \frac{4 \cdot 10^6}{M \cdot h}$ [19],$\alpha = \frac{5 \cdot 10^4}{Molecules}$
5	$c_1 \xrightarrow{k_4} AHL1 + c_1$	$k_4 \cdot c_1$	$k_4 = 5 \cdot 10^{-6} \frac{1}{h}$[20]
6	$c_2 \xrightarrow{k_5} AHL2 + c_2$	$k_5 \cdot c_2$	$k_5 = 5 \cdot 10^{-6} \frac{1}{h}$[20]
7	$2AHL2 + 2AHLR2 \xrightarrow{k_6} AHL2{:}AHLR2$	$\frac{k'_6 \cdot AHL2^2}{V_1 \cdot Na}$	$k_6 = \frac{3 \cdot 10^{19}}{M^3 \cdot h}$[14],$V_1 = 3.7 \cdot 10^{-14} L$[213], $Na = 6.023 \cdot 10^{23}$
8	$AHL2{:}AHLR2 \xrightarrow{k_7} preRes1$	$\frac{k_7 \cdot AHL2{:}AHLR2^{n_1}}{k_{7b}^{n_1} + AHL2{:}AHLR2^{n_1}}$	$k_7 = 6 \cdot 10^{-5} \frac{M}{h}, k_{7b} = 10^{-8} M$, $n_1 = 1$[14]
9	$preRes1 \xrightarrow{k_8} Res1$	$k_8 \cdot preRes1$	$k_8 = \frac{5}{h}$
10	$2AHL1 + 2AHLR1 \xrightarrow{k_9} AHL1{:}AHLR1$	$\frac{k'_9 \cdot AHL1^2}{V_2 \cdot Na}$	$k_9 = \frac{3 \cdot 10^{19}}{M^3 \cdot h}$[14],$V_2 = 10^{-15} L$[221]
11	$AHL1{:}AHLR1 \xrightarrow{k_{10}} Res2$	$\frac{k_{10} \cdot AHL1{:}AHLR1^{n_2}}{k_{10b}^{n_2} + AHL1{:}AHLR1^{n_2}}$	$k_{10} = 6 \cdot 10^{-5} \frac{M}{h}, k_{10b} = 10^{-8} M$, $n_2 = 1$[14]
12	$AHL1 \xrightarrow{k_{11}} \varnothing$	$k_{11} \cdot AHL1$	$k_{11} = \frac{1.19}{h}$[19]
13	$AHL2 \xrightarrow{k_{12}} \varnothing$	$k_{12} \cdot AHL2$	$k_{12} = \frac{1.19}{h}$[19]
14	$AHL1{:}AHLR1 \xrightarrow{k_{13}} \varnothing$	$k_{13} \cdot AHL1{:}AHLR1$	$k_{13} = \frac{1.386}{h}$[14]
15	$AHL2{:}AHLR2 \xrightarrow{k_{14}} \varnothing$	$k_{14} \cdot AHL2{:}AHLR2$	$k_{14} = \frac{1.386}{h}$[14]
16	$Res1 \xrightarrow{k_{15}} \varnothing$	$k_{15} \cdot Res1$	$k_{15} = \frac{4}{h}$[198]
17	$Res2 \xrightarrow{k_{16}} \varnothing$	$k_{16} \cdot Res2$	$k_{16} = \frac{4}{h}$[198]

Chapter 7

Concluding remarks

How to best model biological systems still remains an open question. Herein, we adhered to Monod's postulate that the complexity of biological systems can be reduced to cascades of biomolecular interactions or reactions. These reactions capture the cellular processes that are associated with gene expression either directly or indirectly. These include but are not limited to transport, dimerization, transcription, translation, repression, and induction. The parameters utilized in the models were adopted primarily from experimental observations.

In Chapter 3, a detailed model of the *tet* operon was developed that takes into consideration the biomolecular interactions that comprise this system. It therefore provides a comprehensive picture of the dynamic behavior of *tet* operon which confers bacterial resistance to tetracycline and is extensively used in building synthetic biological systems. The kinetic parameters of this model were adopted from previously published studies. In some cases, however, the required parameters were either not determined at all or measured under conditions different from ones assumed in the model. This unavailability of parameters has always been a challenge when detailed models are developed. A future direction related to this study would involve the validation of the model predictions against experimental data which are all measured under similar conditions. Interesting predictions that could be experimentally tested include that the amount of TetA scales almost linearly with the administered aTc amount and that the existence of two *tet* promoters is redundant. This model validation would make the model a useful tool that can be utilized in the design of synthetic biological constructs

featuring *tet* components. In addition to contributing toward developing synthetic biological systems, an experimentally validated model could potentially aid in the design of novel antibiotics which circumvent this tetracycline resistance mechanism.

In Chapter 4, we paired flow cytometry experiments with mathematical modeling to explain the single-cell response to different sugars. This work is still in progress and all the experiments were performed in Beisel's lab at North Carolina State University. Modeling efforts focused on the Arabinose system from *E. coli*, although experiments were performed for various other sugar utilization pathways. In this study, the predictions of the mathematical model were first validated against experimental measurements from the present study. This allowed for estimating the few required parameters that did not exist in the literature. Indeed, mathematical modeling played a catalytic role in this study as it explained observed experimental phenotypes and motivated new sets of experiments. First, our experiments showed that the so-called "all-or-none" response exhibited in the absence of catabolism is converted to a "some-or-none" response in the natural system. With mathematical modeling the "some-or-none" response was explained and attributed to arabinose catabolism. Second, stochastic simulations of single-cells showed a non-intuitive, otherwise unpredictable response. In particular, simulations suggested that cells can transition bidirectionally from the induced to uninduced state only when catabolism is present. In the absence of catabolism, cells transition unidirectionally only from the uninduced to induced state. Had we only used experimental approaches, this response would be challenging to predict. Overall, this study sheds light on the individuality of bacterial behavior upon sugar administration. It also elucidates a phenotypic plasticity related to sugar consumption pathways in bacteria. Even though this study interrogated different sugar pathways that had not been explored previously, a future direction would include the investigation of additional sugar pathways that are still understudied. This would provide a complete understanding of single-cell response to sugars. Interesting research projects could also address whether quorum sensing is involved in sugar consumption pathways.

In Chapter 5 we presented the design and development of two "genetic switches" that tightly control the expression of a target gene. Even though these two switches were developed in *E. coli*, they are context independent. Similarly to the previous study, we coupled experimental and computational work to fully characterize the behavior of

these switches. In this case, the synthetic systems were first developed in Kaznessis lab. Thus, many required model parameters were not available in the literature. To this end, we needed to tweak our model to estimate a number of parameters. After we validated our model, we used it as a means for quantifying the key parameters governing the dynamics of these switches. Quantifying these parameters experimentally would have been cost and time-consuming. Even though this modeling process required significantly less time and effort, the "trial and error" approach used to estimate the required parameters was still computationally intensive. Therefore, future research directions could explore ways to minimize the computational cost and time of parameter estimation by coupling optimization algorithms to HY3S. Even though an initial attempt towards this direction has been already made by Kaznessis group [217], the development of an algorithm that facilitates parameter estimation, especially when it comes to large reaction networks, would be of high interest. Another potential research direction could be the development of synthetic biological circuits that give rise to interesting cellular dynamics. An attempt has been initiated few years ago by Kaznessis group, where four such circuits with interesting dynamics were designed *in silico* [80]. However, many of the parameters utilized in those models were not supported by sufficient experimental data. In our case, the availability of a validated model will allow for making powerful predictions about these circuits that can be subsequently tested experimentally in the wet lab.

Finally, Chapter 6 discusses the *in silico* design of a synthetic multicellular system composed of yeast and bacteria. Here, we developed a reduced model to capture the dynamics of the two populations. In contrast to the previous three studies where HY3S was used, in this case a set of chemical Langevin equations (see Chapter 2) was implemented. The kinetic parameters used in this model were adopted from previous experimental observations. The main drawback of this model was the assumption that the dynamics of synthetic gene networks remain the same across different species, i.e. they are context independent. This allowed us to use parameters quantified in bacteria in cases where parameters for the yeast gene network were required. This highlights again the need for experimental parameters. Therefore, an important future direction is the experimental design of this system. This work is currently underway, as this system is being designed experimentally in the Schmidt-Dannert lab at the University

of Minnesota. Building this system experimentally will help us quantify the parameters of the model and make accurate predictions about the behavior of this system under various conditions. Another research direction that could be followed is the advancement of HY3S in a way that it monitors the behavior of two populations simultaneously, as in the case of the yeast-bacteria ecosystem. Currently, HY3S can only account for single-cell behavior and cannot capture the dynamics of cell populations or interactions among different cells. Although this would be a challenging study, it would be of high interest to incorporate the basis for cell-to-cell communication in HY3S so it can be utilized in simulating multicellular systems that are receiving increasing interest.

A question that can be legitimately posed is why we need such detailed models, which require exhaustive parameter searches, rather than simply using reduced models? Indeed, there is a tendency toward using simple models which consist of only few equations and capture the experimental phenotype well. Even though these models obey Occam's razor principle, they are inefficient in elucidating how changes in individual steps of gene expression affect the final outcome, i.e. the amount of protein produced [31]. As an example, consider directed evolution that is extensively used for generating a wide spectrum of gene expression profiles [222]. In this method, mutations are introduced to DNA in a systematic way aiming to produce different protein amounts. In this process, key phenomena, such as the RNA polymerase recruitment rate to DNA, dictate the change in the amount of the produced protein. Simple models, such as one equation input-output models, are able to only describe the change in the amount of the produced protein and cannot account for the changes in distinct phenomena caused by mutations. Detailed models, however, that take each step of gene expression into consideration, capture modifications in individual steps while retaining all the other steps. As such, only detailed models would augment our understanding of how a small individual change in the transcription apparatus affects the final outcome. This understanding of how particular changes on single molecular components influence the final phenotype is of high significance in synthetic biology. Thus, detailed models should be preferred in cases where sufficient experimental data are available. However, when there are no reachable experimental data that can be plugged in the detailed models, simpler models should rather be used, as the detailed models will suffer from the non-unique choice of parameter values. Therefore, before a decision is made regarding the model

that would best describe the behavior of a system, the availability of experimental data should be evaluated and a particular type of model should be then tailored to the system of interest.

References

[1] J.C. Venter. Multiple personal genomes await. *Nature*, 464(7289):676–677, 2010.

[2] N. Nandagopal and M.B. Elowitz. Synthetic biology: integrated gene circuits. *Science*, 333(6047):1244–1248, 2011.

[3] E. Andrianantoandro, S. Basu, D.K. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2(1), 2006.

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

[5] L.M. Tuttle, H. Salis, J. Tomshine, and Y.N. Kaznessis. Model-driven designs of an oscillating gene network. *Biophysical journal*, 89(6):3873–3883, 2005.

[6] J. Stricker, S. Cookson, M.R. Bennett, W.H. Mather, L.S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, 2008.

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

[8] K. Volzing, K. Biliouris, and Y.N. Kaznessis. Proteon and proteoff, new protein devices that inducibly activate bacterial gene expression. *ACS Chemical Biology*, 6(10):1107–1116, 2011.

[9] C.L. Beisel, T.S. Bayer, K.G. Hoff, and C.D. Smolke. Model-guided design of ligand-regulated rnai for programmable control of gene expression. *Molecular systems biology*, 4(1), 2008.

[10] J. Bonnet, P. Yin, M.E. Ortiz, P. Subsoontorn, and D. Endy. Amplifying genetic logic gates. *Science*, 340(6132):599–603, 2013.

[11] T.S. Moon, C. Lou, A. Tamsir, B.C. Stanton, and C.A. Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423):249–253, 2012.

[12] K.I. Ramalingam, J.R. Tomshine, J.A. Maynard, and Y.N. Kaznessis. Forward engineering of synthetic bio-logical AND gates. *Biochemical Engineering Journal*, 47(1-3):38–47, 2009.

[13] J.C. Anderson, C.A. Voigt, and A.P. Arkin. Environmental signal integration by a modular AND gate. *Molecular systems biology*, 3(1), 2007.

[14] S. Basu, Y. Gerchman, C.H. Collins, F.H. Arnold, and R. Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037):1130–1134, 2005.

[15] A. Pai, Y. Tanouchi, C.H. Collins, and L. You. Engineering multicellular systems by cell-cell communication. *Current opinion in biotechnology*, 20(4):461–470, 2009.

[16] H. Song, S. Payne, C. Tan, and L. You. Programming microbial population dynamics by engineered cell–cell communication. *Biotechnology Journal*, 6(7):837–849, 2011.

[17] K. Brenner, L. You, and F.H. Arnold. Engineering microbial consortia: a new frontier in synthetic biology. *Trends in biotechnology*, 26(9):483–489, 2008.

[18] T. Bulter, S.G. Lee, W.W. Wong, E. Fung, M.R. Connor, and J.C. Liao. Design of artificial cell–cell communication using gene and metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(8):2299–2304, 2004.

[19] L. You, R.S. Cox, R. Weiss, and F.H. Arnold. Programmed population control by cell-cell communication and regulated killing. *Nature*, 428(6985):868–871, 2004.

[20] F.K. Balagaddé, H. Song, J. Ozaki, C.H. Collins, M. Barnet, F.H. Arnold, S.R. Quake, and L. You. A synthetic escherichia coli predator–prey ecosystem. *Molecular systems biology*, 4(1), 2008.

[21] A. Prindle, P. Samayoa, I. Razinkov, T. Danino, L.S. Tsimring, and J. Hasty. A sensing array of radically coupled genetic biopixels. *Nature*, 2011.

[22] D.K. Ro, E.M. Paradise, M. Ouellet, K.J. Fisher, K.L. Newman, J.M. Ndungu, K.A. Ho, R.A. Eachus, T.S. Ham, J. Kirby, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–943, 2006.

[23] S.K. Lee, H. Chou, T.S. Ham, T.S. Lee, and J.D. Keasling. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current opinion in biotechnology*, 19(6):556–563, 2008.

[24] M.R. Connor and J.C. Liao. Microbial production of advanced transportation fuels in non-natural hosts. *Current opinion in biotechnology*, 20(3):307–315, 2009.

[25] W.C. Ruder, T. Lu, and J.J. Collins. Synthetic biology moving into the clinic. *Science*, 333(6047):1248–1252, 2011.

[26] A.S. Khalil and J.J. Collins. Synthetic biology: applications come of age. *Nature reviews. Genetics*, 11(5):367, 2010.

[27] P. Purnick and R. Weiss. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology*, 10(6):410–422, 2009.

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

[29] H. De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology*, 9(1):67–103, 2002.

[30] D. Endy and R. Brent. Modelling cellular behaviour. *Nature*, 409(6818):391–395, 2001.

[31] Y.N. Kaznessis. Models for synthetic biology. *BMC Systems Biology*, 1(1):47, 2007.

[32] Y.N. Kaznessis. Multi-scale models for gene network engineering. *Chemical Engineering Science*, 61(3):940–953, 2006.

[33] M.D. Laubichler and G.B. Müller. *Modeling biology: Structures, behavior, evolution*. MIT Press, 2007.

[34] K. Biliouris, P. Daoutidis, and Y. Kaznessis. Stochastic simulations of the tetracycline operon. *BMC systems biology*, 5(1):9, 2011.

[35] K. Biliouris, D. Babson, C. Schmidt-Dannert, and Y. Kaznessis. Stochastic simulations of a synthetic bacteria-yeast ecosystem. *BMC systems biology*, 6(1):58, 2012.

[36] T. Afroz and C.L. Beisel. Understanding and exploiting feedback in synthetic biology. *Chemical Engineering Science*, 2013.

[37] M.N. Win and C.D. Smolke. A modular and extensible rna-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences*, 104(36):14283, 2007.

[38] S.J. Goldfless, B.J. Belmont, A.M. de Paz, J.F. Liu, and J.C. Niles. Direct and specific chemical control of eukaryotic translation with a synthetic rna–protein interaction. *Nucleic acids research*, 40(9):e64–e64, 2012.

[39] E.K. Culyba, J.L. Price, S.R. Hanson, A. Dhar, C.H. Wong, M. Gruebele, E.T. Powers, and J.W. Kelly. Protein native-state stabilization by placing aromatic side chains in n-glycosylated reverse turns. *Science*, 331(6017):571–575, 2011.

[40] M.W. Popp, S.K. Dougan, T.Y. Chuang, E. Spooner, and H.L. Ploegh. Sortase-catalyzed transformations that improve the properties of cytokines. *Proceedings of the National Academy of Sciences*, 108(8):3169–3174, 2011.

[41] M. Gossen and H. Bujard. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*, 89(12):5547–5551, 1992.

[42] M. Gossen, A.L. Bonin, and H. Bujard. Control of gene activity in higher eukaryotic cells by prokaryotic regulatory elements. *Trends in Biochemical Sciences*, 18(12):471–475, 1993.

[43] P.A. Furth, L. St Onge, H. Boger, P. Gruss, M. Gossen, A. Kistner, H. Bujard, and L. Hennighausen. Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. *Proceedings of the National Academy of Sciences*, 91(20):9302, 1994.

[44] M. Gossen, S. Freundlieb, G. Bender, G. Muller, W. Hillen, and H. Bujard. Transcriptional activation by tetracyclines in mammalian cells. *Science*, 268(5218):1766, 1995.

[45] T.S. Bayer and C.D. Smolke. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature biotechnology*, 23(3):337–343, 2005.

[46] E. Fung, W.W. Wong, J.K. Suen, T. Bulter, S.g. Lee, and J.C. Liao. A synthetic gene–metabolic oscillator. *Nature*, 435(7038):118–122, 2005.

[47] T.L. Deans, C.R. Cantor, and J.J. Collins. A tunable genetic switch based on rnai and repressor proteins for regulating gene expression in mammalian cells. *Cell*, 130(2):363–372, 2007.

[48] D.L. Shis and M.R. Bennett. Library of synthetic transcriptional and gates built with split t7 rna polymerase mutants. *Proceedings of the National Academy of Sciences*, 2013.

[49] B. Wang, R.I. Kitney, N. Joly, and M. Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature communications*, 2:508, 2011.

[50] P. Siuti, J. Yazbek, and T.K. Lu. Synthetic circuits integrating logic and memory in living cells. *Nature biotechnology*, 2013.

[51] M.B. Miller and B.L. Bassler. Quorum sensing in bacteria. *Annual Reviews in Microbiology*, 55(1):165–199, 2001.

[52] S. Basu, R. Mehreja, S. Thiberge, M.T. Chen, and R. Weiss. Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6355, 2004.

[53] W. Shou, S. Ram, and J.M.G. Vilar. Synthetic cooperation in engineered yeast populations. *Proceedings of the National Academy of Sciences*, 104(6):1877, 2007.

[54] J. Monod and A. Wainhouse. *Chance and necessity: an essay on the natural philosophy of modern biology*. Vintage Books New York, 1972.

[55] H. Salis and Y. Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *The Journal of chemical physics*, 122:054103, 2005.

[56] D.J. Wilkinson. *Stochastic modelling for systems biology*, volume 44. CRC press, 2011.

[57] V. Sotiropoulos and Y.N. Kaznessis. Analytical derivation of moment equations in stochastic chemical kinetics. *Chemical engineering science*, 66(3):268–277, 2011.

[58] H.H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814, 1997.

[59] H.H. McAdams and A. Arkin. Its a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics*, 15(2):65–69, 1999.

[60] M. Thattai and A. Van Oudenaarden. Stochastic gene expression in fluctuating environments. *Genetics*, 167(1):523–530, 2004.

[61] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 2008.

[62] N.G. Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. North holland, 1992.

[63] D.T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1):404–425, 1992.

[64] T. Jahnke and W. Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology*, 54(1):1–26, 2007.

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

[66] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

[67] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[68] M.A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, 104(9):1876–1889, 2000.

[69] M. Herajy and M. Heiner. Hybrid representation and simulation of stiff biochemical networks. *Nonlinear Analysis: Hybrid Systems*, 6(4):942–959, 2012.

[70] C.S. Gillespie. Stochastic simulation of chemically reacting systems using multicore processors. *The Journal of Chemical Physics*, 136:014101, 2012.

[71] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. Copasia complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[72] A. Funahashi, M. Morohashi, H. Kitano, and N. Tanimura. Celldesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, 1(5):159–162, 2003.

[73] A.D. Hill, J.R. Tomshine, E. Weeding, V. Sotiropoulos, and Y.N. Kaznessis. SynBioSS: the synthetic biology modeling suite. *Bioinformatics*, 24(21):2551, 2008.

[74] H. Salis, V. Sotiropoulos, and Y.N. Kaznessis. Multiscale Hy3S: Hybrid stochastic simulation for supercomputers. *BMC bioinformatics*, 7(1):93, 2006.

[75] M.H. Medema, R. van Raaphorst, E. Takano, and R. Breitling. Computational tools for the synthetic design of biochemical pathways. *Nature Reviews Microbiology*, 10(3):191–202, 2012.

[76] M.A. Marchisio and J. Stelling. Computational design tools for synthetic biology. *Current opinion in biotechnology*, 20(4):479–485, 2009.

[77] P. Wong, S. Gladney, and JD Keasling. Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnology progress*, 13(2), 1997.

[78] M. Stamatakis and N.V. Mantzaris. Comparison of Deterministic and Stochastic Models of the lac Operon Genetic Network. *Biophysical Journal*, 96(3):887–906, 2009.

[79] Z.L. Xiu, Z.Y. Chang, and A.P. Zeng. Nonlinear dynamics of regulation of bacterial trp operon: model analysis of integrated effects of repression, feedback inhibition, and attenuation. *Biotechnology progress*, 18(4):686–693, 2002.

[80] V. Sotiropoulos and Y.N. Kaznessis. Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. *BMC Systems Biology*, 1(1):7, 2007.

[81] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M. Foglierini, and L. Serrano. Noise in transcription negative feedback loops: simulation and experimental analysis. *Molecular Systems Biology*, 2(1), 2006.

[82] R.A. Goldman, T. Hasan, C.C. Hall, W.A. Strycharz, and B.S. Cooperman. Photoincorporation of tetracycline into Escherichia coli ribosomes. Identification of the major proteins photolabeled by native tetracycline and tetracycline photoproducts and implications for the inhibitory action of tetracycline on protein synthesis. *Biochemistry*, 22(2):359, 1983.

[83] T.J. Franklin and A. Godfrey. Resistance of E. coli to tetracyclines. *Biochem. J*, 94:54–60, 1965.

[84] V.A. Fischetti, R.P. Novick, and J.J. Ferretti. *Gram-positive pathogens*. Amer Society for Microbiology, 2006.

[85] W. Saenger, P. Orth, C. Kisker, W. Hillen, and W. Hinrichs. The tetracycline repressor-A paradigm for a biological switch. *Angewandte Chemie International Edition*, 39(12), 2000.

[86] S.B. Levy, L.M. McMurry, V. Burdett, P. Courvalin, W. Hillen, M.C. Roberts, and D.E. Taylor. Nomenclature for tetracycline resistance determinants. *Antimicrobial agents and chemotherapy*, 33(8):1373, 1989.

[87] W. Hillen and C. Berens. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annual Reviews in Microbiology*, 48(1):345–369, 1994.

[88] A. Yamaguchi, T. Udagawa, and T. Sawai. Transport of divalent cations with tetracycline as mediated by the transposon Tn10-encoded tetracycline resistance protein. *Journal of Biological Chemistry*, 265(9):4809–4813, 1990.

[89] W. Hillen, K. Schollmeier, and C. Gatz. Control of expression of the Tn10-encoded tetracycline resistance operon. II. Interaction of RNA polymerase and TET repressor with the tet operon regulatory region . *Journal of Moleculatr Biology*, 172(2):185–201, 1984.

[90] D.W. Daniels and K.P. Bertrand. Promoter mutations affecting divergent transcription in the Tn10 tetracycline resistance determinant. *Journal of molecular biology*, 184(4):599, 1985.

[91] M. Takahashi, L. Altschmied, and W. Hillen. Kinetic and equilibrium characterization of the tet repressor-tetracycline complex by fluorescence measurements: evidence for divalent metal ion requirement and energy transfer. *Journal of molecular biology*, 187(3):341–348, 1986.

[92] L. McMurry, R.E. Petrucci Jr, and S.B. Levy. Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in Escherichia coli. *Proc. Natl. Acad. Sci. USA*, 77(7):3974–7, 1980.

[93] I.M. Mansuy and H. Bujard. Tetracycline-regulated gene expression in the brain. *Current Opinion in Neurobiology*, 10(5):593–596, 2000.

[94] M. Geissendorfer and W. Hillen. Regulated expression of heterologous genes in-Bacillus subtilis using the Tn10 encodedtet regulatory elements. *Applied microbiology and biotechnology*, 33(6):657–663, 1990.

[95] W. Hillen, C. Gatz, L. Altschmied, K. Schollmeier, and I. Meier. Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. *Journal of molecular biology*, 169(3):707, 1983.

[96] I. Meier, L.V. Wray, and W. Hillen. Differential regulation of the Tn10-encoded tetracycline resistance genes tetA and tetR by the tandem tet operators O1 and O2. *The EMBO journal*, 7(2):567, 1988.

[97] C. Kleinschmidt, K. Tovar, W. Hillen, and D. Porschke. Dynamics of repressor-operator recognition: Tn10-encoded tetracycline resistance control. *Biochemistry*, 27(4):1094–1104, 1988.

[98] C. Kleinschmidt, K. Tovar, and W. Hillen. Computer simulations and experimental studies of gel mobility patterns for weak and strong non-cooperative protein binding to two targets on the same DNA: application to binding of tet repressor variants to multiple and single tet operator sites. *Nucleic acids research*, 19(5):1021, 1991.

[99] P. Orth, D. Schnappinger, W. Hillen, W. Saenger, and W. Hinrichs. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature structural biology*, 7(3):215–219, 2000.

[100] T. Lederer, M. Kintrup, M. Takahashi, P.E. Sum, G.A. Ellestad, and W. Hillen. Tetracycline Analogs Affecting Binding to Tn10-Encoded Tet Repressor Trigger the Same Mechanism of Induction. *Biochemistry*, 35(23):7439–7446, 1996.

[101] M. Takahashi, J. Degenkolb, and W. Hillen. Determination of the equilibrium association constant between Tet repressor and tetracycline at limiting Mg2+

concentrations: a generally applicable method for effector-dependent high-affinity complexes. *Analytical biochemistry*, 199(2):197–202, 1991.

[102] R.J. Brooker. *Genetics: analysis and principles*. McGraw-Hill Higher Education, 2008.

[103] U. Gulland and W. Hillen. The Tn10-encoded tetR mRNA has heterogeneous 5 ends in vivo and in vitro. *Gene*, 114(1):97, 1992.

[104] D.G. White, M.N. Alekshun, P.F. McDermott, and S.B. Levy. *Frontiers in antimicrobial resistance: a tribute to Stuart B. Levy*. ASM Press, 2005.

[105] A.R. English, S.Y. P'an, T.J. McBride, J.F. Gardocki, G. Van Halsema, and A.W. Wright. Tetracycline-Microbiologic, Pharmacologic, and Clinical Evaluation. *Antibiotics Annual*, pages 70–80, 1954.

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

[107] D.A. McQuarrie. Stochastic approach to chemical kinetics. *Journal of Applied Probability*, 4(3):413–478, 1967.

[108] H. Bremer, P.P. Dennis, et al. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella typhimurium: cellular and molecular biology. American Society for Microbiology, Washington, DC*, pages 1527–1542, 1987.

[109] K.P. Bertrand, K. Postle, L.V. Wray Jr, and W.S. Reznikoff. Construction of a single-copy promoter vector and its use in analysis of regulation of the transposon Tn10 tetracycline resistance determinant. *Journal of bacteriology*, 158(3):910, 1984.

[110] B. Eckert and C.F. Beck. Overproduction of transposon Tn10-encoded tetracycline resistance protein results in cell death and loss of membrane potential. *Journal of bacteriology*, 171(6):3557, 1989.

[111] M.T. Korpela, J.S. Kurittu, J.T. Karvinen, and M.T. Karp. A recombinant Escherichia coli sensor strain for the detection of tetracyclines. *Anal. Chem*, 70(21):4457–4462, 1998.

[112] B.S. Speer, N.B. Shoemaker, and A.A. Salyers. Bacterial resistance to tetracycline: mechanisms, transfer, and clinical significance. *Clinical Microbiology Reviews*, 5(4):387, 1992.

[113] C. Berens, D. Schnappinger, and W. Hillen. The role of the variable region in Tet repressor for inducibility by tetracycline. *Journal of Biological Chemistry*, 272(11):6936, 1997.

[114] C. Sizemore, Gulland U. Wissmann, A., and W. Hillen. Quantitative analysis of Tn10 Tet repressor binding to complete set of tet operator mutants. *Nucleic acids research*, 18(10):2875, 1990.

[115] I. Chopra and M. Roberts. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and Molecular Biology Reviews*, 65(2):232, 2001.

[116] E. Bertrand-Burggraf, J.F. Lefevre, and M. Daune. A new experimental approach for studying the association between RNA polymerase and the tet promoter of pBR322. *Nucleic acids research*, 12(3):1697, 1984.

[117] D. Chen, S. Bachellier, and D.M.J. Lilley. Activation of the leu-500 Promoter by a Reversed Polarity tetA Gene. *Journal of Biological Chemistry*, 273(1):653, 1998.

[118] A. Sigler, P. Schubert, W. Hillen, and M. Niederweis. Permeation of tetracyclines through membranes of liposomes and Escherichia coli. *European Journal of Biochemistry*, 267(2):527–534, 2000.

[119] S. Kaplan, A. Bren, A. Zaslaver, E. Dekel, and U. Alon. Diverse two-dimensional input functions control bacterial sugar genes. *Molecular cell*, 29(6):786–792, 2008.

[120] A. Novick and M. Weiner. Enzyme induction as an all-or-none phenomenon. *Proceedings of the National Academy of Sciences of the United States of America*, 43(7):553, 1957.

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

[122] M. Cohn and K. Horibata. Analysis of the differentiation and of the heterogeneity within a population of escherichia coli undergoing induced β-galactosidase synthesis. *Journal of bacteriology*, 78(5):613, 1959.

[123] D.A. Siegele and J.C. Hu. Gene expression from plasmids containing the arabad promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences*, 94(15):8168–8172, 1997.

[124] G.M. Süel, J. Garcia-Ojalvo, L.M. Liberman, and M.B. Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(7083):545–550, 2006.

[125] T.Y. Tsai, Y.S. Choi, W. Ma, J.R. Pomerening, C. Tang, and J.E Ferrell Jr. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science Signaling*, 321(5885):126, 2008.

[126] A. Khlebnikov, K.A. Datsenko, T. Skaug, B.L. Wanner, and J.D. Keasling. Homogeneous expression of the pbad promoter in escherichia coli by constitutive expression of the low-affinity high-capacity arae transporter. *Microbiology*, 147(12):3241–3247, 2001.

[127] R.M. Morgan-Kiss, C. Wadler, and J.E. Cronan. Long-term and homogeneous regulation of the escherichia coli arabad promoter by use of a lactose transporter of relaxed specificity. *Proceedings of the National Academy of Sciences*, 99(11):7373–7377, 2002.

[128] M. Santillán. Bistable behavior in a model of the *lac* operon in escherichia coli with variable growth rate. *Biophysical journal*, 94(6):2065–2081, 2008.

[129] D.W. Dreisigmeyer, J. Stajic, I. Nemenman, W.S. Hlavacek, and M.E. Wall. Determinants of bistability in induction of the escherichia coli lac operon. *IET systems biology*, 2(5):293–303, 2008.

[130] S. Krishna, S. Semsey, and K. Sneppen. Combinatorics of feedback in cellular uptake and metabolism of small molecules. *Proceedings of the National Academy of Sciences*, 104(52):20815–20819, 2007.

[131] A. Zaslaver, A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M.G. Surette, and U. Alon. A comprehensive library of fluorescent transcriptional reporters for escherichia coli. *Nature methods*, 3(8):623–628, 2006.

[132] R. Schleif. Regulation of the l-arabinose operon of escherichia coli. *Trends in Genetics*, 16(12):559–565, 2000.

[133] L. Katz and E. Englesberg. Hyperinducibility as a result of mutation in structural genes and self-catabolite repression in the ara operon. *Journal of bacteriology*, 107(1):34–52, 1971.

[134] S.A. Ramsey, J.J. Smith, D. Orrell, M. Marelli, T.W. Petersen, P. de Atauri, H. Bolouri, and J.D. Aitchison. Dual feedback loops in the gal regulon suppress cellular heterogeneity in yeast. *Nature genetics*, 38(9):1082–1087, 2006.

[135] M. Acar, A. Becskei, and A. van Oudenaarden. Enhancement of cellular memory by reducing stochastic transitions. *Nature*, 435(7039):228–232, 2005.

[136] M.R. Bennett, W.L. Pang, N.A. Ostroff, B.L. Baumgartner, S. Nayak, L.S. Tsimring, and J. Hasty. Metabolic gene regulation in a dynamically changing environment. *Nature*, 454(7208):1119–1122, 2008.

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

[138] M.R. Sadykov, T. Hartmann, T.A. Mattes, M. Hiatt, N.J. Jann, Y. Zhu, N. Ledala, R. Landmann, M. Herrmann, H. Rohde, et al. Ccpa coordinates central metabolism and biofilm formation in staphylococcus epidermidis. *Microbiology*, 157(12):3458–3468, 2011.

[139] Y. Chai, P.B. Beauregard, H. Vlamakis, R. Losick, and R. Kolter. Galactose metabolism plays a crucial role in biofilm formation by bacillus subtilis. *mBio*, 3(4), 2012.

[140] M.L. Scrudato and M. Blokesch. The regulatory network of natural competence and transformation of vibrio cholerae. *PLoS Genetics*, 8(6):e1002778, 2012.

[141] F. Rojo. Carbon catabolite repression in pseudomonas: optimizing metabolic versatility and interactions with the environment. *FEMS microbiology reviews*, 34(5):658–684, 2010.

[142] H. Alper and G. Stephanopoulos. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nature Reviews Microbiology*, 7(10):715–723, 2009.

[143] B. Hahn-Hägerdal, K. Karhumaa, M. Jeppsson, and M.F. Gorwa-Grauslund. Metabolic engineering for pentose utilization in saccharomyces cerevisiae. In *Biofuels*, pages 147–177. Springer, 2007.

[144] L.M. Guzman, D. Belin, M.J. Carson, and J. Beckwith. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pbad promoter. *Journal of bacteriology*, 177(14):4121–4130, 1995.

[145] R. Schleif. Arac protein, regulation of the l-arabinose operon in escherichia coli, and the light switch mechanism of arac action. *FEMS microbiology reviews*, 34(5):779–796, 2010.

[146] S. Ogden, D. Haggerty, C.M. Stoner, D. Kolodrubetz, and R. Schleif. The escherichia coli l-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. *Proceedings of the National Academy of Sciences*, 77(6):3346–3350, 1980.

[147] S. Hahn and R. Schleif. In vivo regulation of the escherichia coli arac promoter. *Journal of bacteriology*, 155(2):593–600, 1983.

[148] J.M. Raser and E.K. O'Shea. Noise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–2013, 2005.

[149] F.J. Isaacs, J. Hasty, C.R. Cantor, and J.J. Collins. Prediction and measurement of an autoregulatory genetic module. *Proceedings of the National Academy of Sciences*, 100(13):7714–7719, 2003.

[150] T. Harmer, M. Wu, and R. Schleif. The role of rigidity in dna looping-unlooping by arac. *Proceedings of the National Academy of Sciences*, 98(2):427, 2001.

[151] X. Zhang, T. Reeder, R. Schleif, et al. Transcription activation parameters at ara pbad. *Journal of molecular biology*, 258(1):14–24, 1996.

[152] C.M. Johnson and R.F. Schleif. In vivo induction kinetics of the arabinose promoters in escherichia coli. *Journal of bacteriology*, 177(12):3438, 1995.

[153] G. Wilcox. The interaction of l-arabinose and d-fucose with arac protein. *Journal of Biological Chemistry*, 249(21):6892, 1974.

[154] W. Hendrickson and R.F. Schleif. Regulation of the escherichia coli-arabinose operon studied by gel electrophoresis dna binding assay. *Journal of molecular biology*, 178(3):611–628, 1984.

[155] U. Vogel and K.F. Jensen. The RNA chain elongation rate in Escherichia coli depends on the growth rate. *Journal of bacteriology*, 176(10):2807, 1994.

[156] N. Lee and J. Carbon. Nucleotide sequence of the 5'end of arabad operon messenger rna in escherichia coli b/r. *Proceedings of the National Academy of Sciences*, 74(1):49, 1977.

[157] R. Schleif, W. Hess, S. Finkelstein, and D. Ellis. Induction kinetics of the l-arabinose operon of escherichia coli. *Journal of bacteriology*, 115(1):9, 1973.

[158] C.M. Johnson and R.F. Schleif. Cooperative action of the catabolite activator protein and arac in vitro at the arafgh promoter. *Journal of bacteriology*, 182(7):1995, 2000.

[159] J.B. Scripture, C. Voelker, S. Miller, R.T. O'Donnell, L. Polgar, J. Rade, B.F. Horazdovsky, and R.W. Hogg. High-affinity-arabinose transport operon: Nucleotide sequence and analysis of gene products. *Journal of molecular biology*, 197(1):37–46, 1987.

[160] J.W. Patrick and N. Lee. Purification and properties of an l-arabinose isomerase from escherichia coli. *Journal of Biological Chemistry*, 243(16):4312, 1968.

[161] T.J. Foster, M.A. Davis, D.E. Roberts, K. Takeshita, and N. Kleckner. Genetic organization of transposon tn10. *Cell*, 23(1):201–213, 1981.

[162] J. Engebrecht, K. Nealson, and M. Silverman. Bacterial bioluminescence: isolation and genetic analysis of functions from vibrio fischeri. *Cell*, 32(3):773–781, 1983.

[163] H.B. Kaplan and E.P. Greenberg. Diffusion of autoinducer is involved in regulation of the vibrio fischeri luminescence system. *Journal of Bacteriology*, 163(3):1210–1214, 1985.

[164] A.L. Schaefer, D.L. Val, B.L. Hanzelka, J.E. Cronan Jr, and E.P. Greenberg. Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified vibrio fischeri luxi protein. *Proceedings of the National Academy of Sciences*, 93(18):9505–9509, 1996.

[165] J.A. Kritzer. Grand challenge commentary: Beyond discovery: probes that see, grab and poke. *Nature Chemical Biology*, 6(12):868–870, 2010.

[166] C.E. Vickers, L.M. Blank, and J.O. Krömer. Grand challenge commentary: Chassis cells for industrial biochemical production. *Nature chemical biology*, 6(12):875, 2010.

[167] J.M. Callura, D.J. Dwyer, F.J. Isaacs, C.R. Cantor, and J.J. Collins. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proceedings of the National Academy of Sciences*, 107(36):15898–15903, 2010.

[168] F.J. Isaacs, J.J. Collins, et al. Plug-and-play with rna. *Nature biotechnology*, 23(3):306–307, 2005.

[169] J. Hasty, D. McMillen, and J.J. Collins. Engineered gene circuits. *Nature*, 420(6912):224–230, 2002.

[170] K.A. Egland and E.P. Greenberg. Quorum sensing in vibrio fischeri: analysis of the luxr dna binding region by alanine-scanning mutagenesis. *Journal of bacteriology*, 183(1):382–386, 2001.

[171] C. Fuqua, S.C. Winans, and E.P. Greenberg. Census and consensus in bacterial ecosystems: the luxr-luxi family of quorum-sensing transcriptional regulators. *Annual Reviews in Microbiology*, 50(1):727–751, 1996.

[172] W.C. Fuqua, S.C. Winans, and E.P. Greenberg. Quorum sensing in bacteria: the luxr-luxi family of cell density-responsive transcriptional regulators. *Journal of bacteriology*, 176(2):269, 1994.

[173] D.M. Sitnikov, J.B. Schineller, and T.O. Baldwin. Transcriptional regulation of bioluminesence genes from vibrio fischeri. *Molecular microbiology*, 17(5):801–812, 1995.

[174] O. Scholz, E.M. Henßler, J. Bail, P. Schubert, J. Bogdanska-Urbaniak, S. Sopp, M. Reich, S. Wisshak, M. Köstner, R. Bertram, et al. Activity reversal of tet repressor caused by single amino acid exchanges. *Molecular microbiology*, 53(3):777–789, 2004.

[175] S.H. Choi and E.P. Greenberg. The c-terminal region of the vibrio fischeri luxr protein contains an inducer-independent lux gene activating domain. *Proceedings of the National Academy of Sciences*, 88(24):11115–11119, 1991.

[176] J.H. Devine, G.S. Shadel, and T.O. Baldwin. Identification of the operator of the lux regulon from the vibrio fischeri strain atcc7744. *Proceedings of the National Academy of Sciences*, 86(15):5688–5692, 1989.

[177] K.A. Egland and E.P. Greenberg. Quorum sensing in vibrio fischeri: elements of the luxi promoter. *Molecular microbiology*, 31(4):1197–1204, 1999.

[178] A.H. Finney, R.J. Blick, K.o Murakami, A. Ishihama, and A.M. Stevens. Role of the c-terminal domain of the alpha subunit of rna polymerase in luxr-dependent transcriptional activation of the lux operon during quorum sensing. *Journal of bacteriology*, 184(16):4520–4528, 2002.

[179] C. Berens and W. Hillen. Gene regulation by tetracyclines. *European Journal of Biochemistry*, 270(15):3109–3121, 2003.

[180] J.L. Ramos, M. Martínez-Bueno, A.J. Molina-Henares, W. Terán, K. Watanabe, X. Zhang, M.T. Gallegos, R. Brennan, and R. Tobes. The tetr family of transcriptional repressors. *Microbiology and Molecular Biology Reviews*, 69(2):326–356, 2005.

[181] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[182] Y.N. Kaznessis. Computational methods in synthetic biology. *Biotechnology Journal*, 4(10):1392–1405, 2009.

[183] Y.N. Kaznessis. Multiscale models for synthetic biology. In *Conference proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, volume 1, page 6408, 2009.

[184] D.E. Golan, A.H. Tashjian, E.J. Armstrong, and A.W. Armstrong. *Principles of pharmacology: the pathophysiologic basis of drug therapy*. LWW, 2011.

[185] I. Meynial-Salles, M.A. Cervin, and P. Soucaille. New tool for metabolic pathway engineering in escherichia coli: one-step method to modulate expression of chromosomal genes. *Applied and environmental microbiology*, 71(4):2140–2144, 2005.

[186] B.P. Cormack, R.H. Valdivia, and S. Falkow. Facs-optimized mutants of the green fluorescent protein (gfp). *Gene*, 173(1):33–38, 1996.

[187] A. Vannini, C. Volpari, C. Gargioli, E. Muraglia, R. Cortese, R. De Francesco, P. Neddermann, and S. Di Marco. The crystal structure of the quorum sensing protein trar bound to its autoinducer and target dna. *The EMBO journal*, 21(17):4393–4401, 2002.

[188] M. Doi. *The theory of polymer dynamics*, volume 73. Oxford University Press on Demand, 1988.

[189] R.A. George and J. Heringa. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Engineering*, 15(11):871–879, 2002.

[190] M. Kærn, T.C. Elston, W.J. Blake, and J.J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6):451–464, 2005.

[191] J. Hasty, J. Pradines, M. Dolnik, and J.J. Collins. Noise-based switches and amplifiers for gene expression. *Proceedings of the National Academy of Sciences*, 97(5):2075–2080, 2000.

[192] A. Kamionka, J. Bogdanska-Urbaniak, O. Scholz, and W. Hillen. Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic acids research*, 32(2):842–847, 2004.

[193] M.A. Sørensen and S. Pedersen. Absolute in vivo translation rates of individual codons in Escherichia coli. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *Journal of molecular biology*, 222(2):265, 1991.

[194] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *J. Biol. Chem*, 274:6074–6079, 1999.

[195] M.N. Win and C.D. Smolke. Higher-order cellular information processing with synthetic rna devices. *Science*, 322(5900):456, 2008.

[196] S. Choudhary and C. Schmidt-Dannert. Applications of quorum sensing in biotechnology. *Applied microbiology and biotechnology*, 86(5):1267–1279, 2010.

[197] Y. Tanouchi, R.P. Smith, and L. You. Engineering microbial systems to explore ecological and evolutionary dynamics. *Current Opinion in Biotechnology*, 2012.

[198] B. Hu, J. Du, R. Zou, and Y. Yuan. An environment-sensitive synthetic microbial ecosystem. *PloS one*, 5(5):e10619, 2010.

[199] F.K. Balagadde, L. You, C.L. Hansen, F.H. Arnold, and S.R. Quake. Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science*, 309(5731):137, 2005.

[200] K. Brenner, D.K. Karig, R. Weiss, and F.H. Arnold. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proceedings of the National Academy of Sciences*, 104(44):17300, 2007.

[201] N. Anesiadis, W.R. Cluett, and R. Mahadevan. Dynamic metabolic engineering for increasing bioprocess productivity. *Metabolic engineering*, 10(5):255–266, 2008.

[202] W. Weber, M. Daoud-El Baba, and M. Fussenegger. Synthetic ecosystems based on airborne inter-and intrakingdom communication. *Proceedings of the National Academy of Sciences*, 104(25):10435, 2007.

[203] C. Smith, H. Song, and L. You. Signal discrimination by differential regulation of protein stability in quorum sensing. *Journal of molecular biology*, 382(5):1290–1297, 2008.

[204] T.D. Webster and R.C. Dickson. Direct selection of saccharomyces cerevisiae resistant to the antibiotic g418 following transformation with a dna vector carrying the kanamycin-resistance gene of tn903. *Gene*, 26(2-3):243–252, 1983.

[205] Y.S. You, H. Marella, R. Zentella, Y. Zhou, T. Ulmasov, T.H.D. Ho, and R.S. Quatrano. Use of bacterial quorum-sensing components to regulate gene expression in plants. *Plant physiology*, 140(4):1205, 2006.

[206] M. Kærn, T.R. Elston, W.J. Blake, and J.J. Collins. Stochasticity in gene expression. *Nature Reviews Genetics*, 6:451–464, 2005.

[207] H. Salis and Y.N. Kaznessis. An equation-free probabilistic steady-state approximation: dynamic application to the stochastic simulation of biochemical reaction networks. *The Journal of chemical physics*, 123:214106, 2005.

[208] Daniel T Gillespie. The chemical langevin equation. *The Journal of Chemical Physics*, 113:297, 2000.

[209] H. Salis and Y. Kaznessis. Numerical simulation of stochastic gene circuits. *Computers and Chemical Engineering*, 29(3):577–588, 2005.

[210] J. Hasty, D. McMillen, F. Isaacs, J.J. Collins, et al. Computational studies of gene regulatory networks: in numero molecular biology. *Nature Reviews Genetics*, 2(4):268–279, 2001.

[211] D.A. Charlebois, T.J. Perkins, and M. Kaern. Stochastic gene expression and the processing and propagation of noisy signals in genetic networks. *Information Processing and Biological Systems*, pages 89–112, 2011.

[212] P.E. Kloeden, E. Platen, and H. Schurz. Stochastic differential equations. *Numerical Solution of SDE Through Computer Experiments*, pages 63–90, 1994.

[213] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer. Bionumbers the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl 1):D750, 2010.

[214] C. Hui. Carrying capacity, population equilibrium, and environment's maximal load. *Ecological Modelling*, 192(1-2):317–320, 2006.

[215] Sun L. Kambam P.K., Henson M.A. Design and mathematical modelling of a synthetic symbiotic ecosystem. *IET Syst Biol.*, 2(1):33–38, 2008.

[216] C.H. Collins, F.H. Arnold, and J.R. Leadbetter. Directed evolution of vibrio fischeri luxr for increased sensitivity to a broad spectrum of acyl-homoserine lactones. *Molecular microbiology*, 55(3):712–723, 2005.

[217] J. Tomshine and Y.N. Kaznessis. Optimization of a stochastically simulated gene network model via simulated annealing. *Biophysical journal*, 91(9):3196–3205, 2006.

[218] R.J. Brooker. *Genetics: analysis and principles*. McGraw-Hill Higher Education New York, NY, 2005.

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

[220] L. Pasotti, S. Zucca, M. Lupotto, M Cusella De Angelis, P. Magni, et al. Characterization of a synthetic bacterial self-destruction device for programmed cell

death and for recombinant proteins release. *Journal of biological engineering*, 5(8), 2011.

[221] S. Sundararaj, A. Guo, B. Habibi-Nazhad, M. Rouani, P. Stothard, M. Ellison, and D.S. Wishart. The cybercell database (ccdb): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of escherichia coli. *Nucleic acids research*, 32(suppl 1):D293, 2004.

[222] P.A. Romero and F.H. Arnold. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12):866–876, 2009.

Appendix A

Experimental proTeOn and proTeOff characterization

PROTEON and PROTEOFF expression

Under the control of a LacI repressible T7 promoter, PROTEON and PROTEOFF expression is IPTG inducible. Cells cultured with 0, 0.1, 0.5 and 1 mM IPTG showed increasing cytosolic protein levels with increasing IPTG, as shown in Figure 2a. Even at high levels, PROTEON is efficiently translated, stable, and soluble in *E. coli*. As PROTEOFF differs from PROTEON by a single amino acid, the same is assumed for PROTEOFF protein.

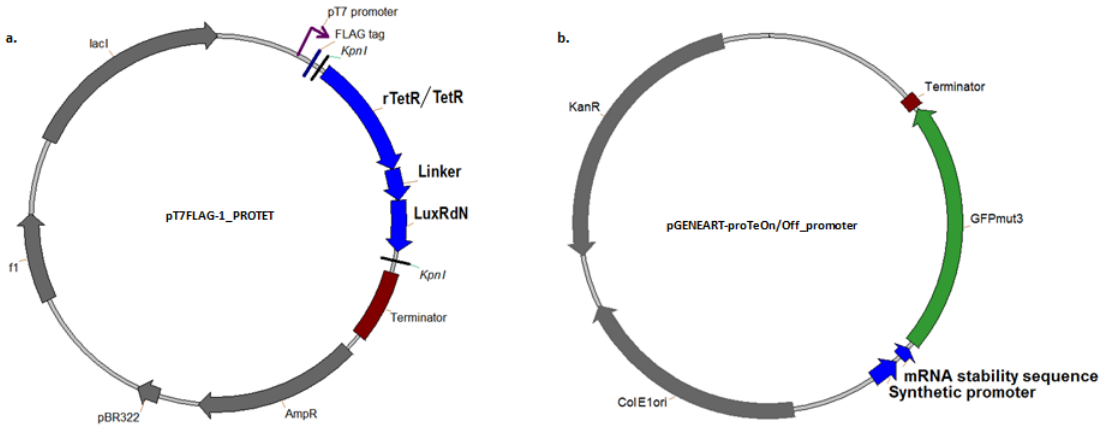

Figure 1: a) PROTEON and PROTEOFF. Both synthetic proteins, rTetR-LuxRΔN and TetR-LuxRΔN, are under the control of a LacI repressible T7 promoter on low-copy plasmid, pT7-FLAG1 (Sigma), b) proTeOn and proTeOff synthetic promoter. The synthetic promoters and GFPmut3 gene were synthesized by GENEART on, pMK, a pUC19 derived expression vector that's compatible in *E. coli*, high copy, and kanamycine resistant.

PROTEON and PROTEOFF activation

proTeOn activity is aTc inducible in *E. coli*. In the absence of aTc, PROTEON is inactive and GFP is expressed at a low basal level. Upon the addition of aTc, PROTEON is activated through a conformational change; it binds the synthetic promoter

and upregulates GFP expression. With 10 and 200 ng/ml aTc, GFP is upregulated 10 and 30 - fold respectively by Western blot. Meanwhile, the total cytosolic transactivator levels modestly decrease in the presence of aTc, as shown in Figure 2b. This decrease can be attributed to the increased demand for the cell's transcription and translation machinery with active PROTEON upregulating GFP.

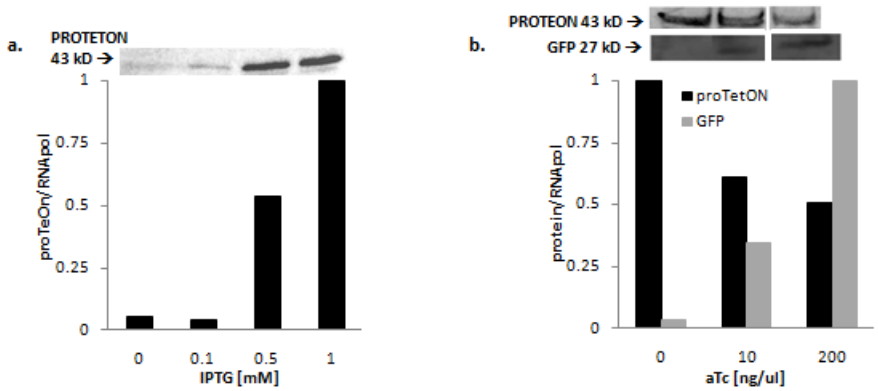

Figure 2: a) PROTEON expression relative to uninduced conditions, b) PROTEON activation.

proTeOn and proTeOff phenotype analysis

proTeOn and proTeOff induction and phenotype analysis by flow cytometry was investigated over 20 hours post-induction. Induction experiments were repeated and the behaviors compared across replicates. The general behavior observed across replicates for proTeOn is in good agreement with the specific results presented in the Results and Discussion. Overall, proTeOn upregulates target genes by one hour post-aTc treatment and achieves 15-fold upregulation through long times. Steady state expression levels are reached by 5 hours and 10 hours post-treatment with low and high (10 and 200 ng/ml) aTc respectively. Expression is generally maintained at 10 and 15-fold above uninduced controls through long times with low and high aTc levels. When expressed in the presence of aTc, PROTEON achieves target gene upregulation within 2 hours. Overall, steady state is achieved by 5 hours after the transactivator's initial expression, with target protein levels 10-fold over the uninduced controls. proTeOn's behavior is

presented in Figure 3a.

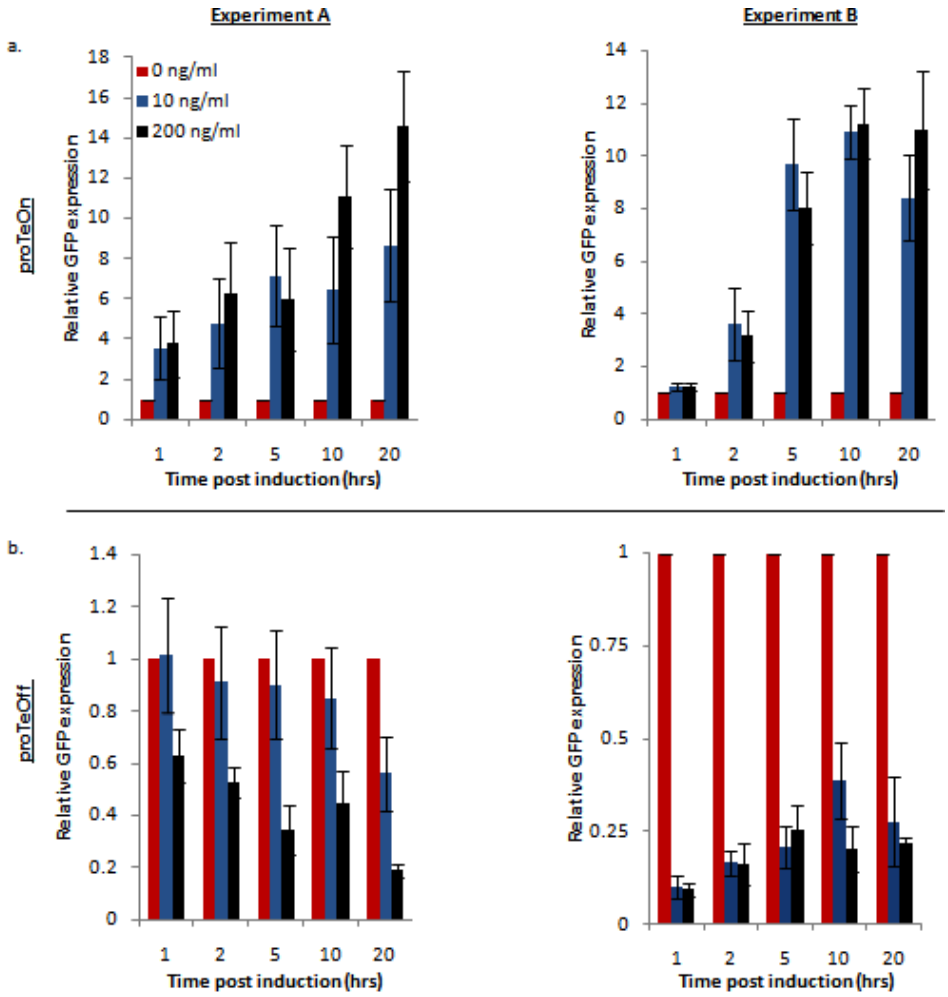

Figure 3: Mean GFP expression was analyzed by flow cytometry 1, 2, 5, 10 and 20 hours post-treatment for both experimental set-ups as described.

proTeOff upregulates target gene expression in the absence of aTc. With high (200 ng/ml) aTc, expression is reduced to half that of the untreated samples by 2 hours. In general, with low (10 ng/ml) aTc, this reduction may not be realized until 20 hours post-treatment. Low, steady state expression is achieved by 5 hours post-treatment with high aTc and maintained through long times. Minimum target gene expression levels of one-half and one-fifth that of the untreated samples are observed with low and

high aTc respectively.

When expressed in the absence of aTc, transcription upregulation by proTeOff is observed within one hour after PROTEOFF expression is induced. In low and high aTc, reduced proTeOff activity is observed across all times. In general, steady state activity is achieved by 5 hour after PROTEOFF expression is induced. Overall, proTeOff activity is one-fourth that of untreated samples for both 10 and 200 ng/ml aTc through long times. proTeOff's behavior is shown in Figure 3b.

Appendix B

Stochastic Differential Equations capturing the reaction network dynamics

$y_1 = S.\ cerevisiae\ cells$

$y_2 = E.\ coli\ cells$

$y_3 = C4HSL$

$y_4 = 3oxoC6HSL$

$y_5 = preKanR1$

$y_6 = KanR1$

$y_7 = 3oxoC6HSL : LuxR$

$y_8 = C4HSL : RhlR$

$y_9 = KanR$

$$dy_1 = \left(k_1 y_1\left(1 - \frac{y_1 + y_2}{c_{max}}\right) - \frac{k_3 \cdot Kan \cdot y_1}{1 + \alpha \cdot y_6}\right)dt + \sqrt{k_1 y_1}dw_1 - \sqrt{k_1 y_1 \frac{y_1 + y_2}{c_{max}}}dw_2 - \sqrt{\frac{k_3 \cdot Kan \cdot y_1}{1 + \alpha \cdot y_6}}dw_3 \quad (1)$$

$$dy_2 = \left(k_2 y_2\left(1 - \frac{y_1 + y_2}{c_{max}}\right) - \frac{k_3 \cdot Kan \cdot y_2}{1 + \alpha \cdot y_9}\right)dt + \sqrt{k_2 y_2}dw_4 - \sqrt{k_2 y_2 \frac{y_1 + y_2}{c_{max}}}dw_5 - \sqrt{\frac{k_3 \cdot Kan \cdot y_2}{1 + \alpha \cdot y_9}}dw_6 \quad (2)$$

$$dy_3 = \left(k_4 y_1 - \frac{k_9 \cdot RhlR^2 \cdot {y_3}^2}{Na \cdot V_2} - k_{11} y_3\right)dt + \sqrt{k_4 y_1}dw_7 - \sqrt{\frac{k_9 \cdot RhlR^2 \cdot {y_3}^2}{Na \cdot V_2}}dw_8 - \sqrt{k_{11} y_3}dw_9 \quad (3)$$

$$dy_4 = \left(k_5 y_2 - \frac{k_6 \cdot LuxR^2 \cdot {y_4}^2}{Na \cdot V_1} - k_{12} y_4\right)dt + \sqrt{k_5 y_2}dw_{10} - \sqrt{\frac{k_6 \cdot LuxR^2 \cdot {y_4}^2}{Na \cdot V_1}}dw_{11} - \sqrt{k_{12} y_4}dw_{12} \quad (4)$$

$$dy_5 = \left(\frac{k_7 y_7^{n_1}}{k_{7b}^{n_1} + y_7^{n_1}} - k_8 y_5 \right) dt + \sqrt{\frac{k_7 y_7^{n_1}}{k_{7b}^{n_1} + y_7^{n_1}}} dw_{13} - \sqrt{k_8 y_5} dw_{14} \tag{5}$$

$$dy_6 = \left(k_8 y_5 - k_{15} y_6 \right) dt + \sqrt{k_8 y_5} dw_{14} - \sqrt{k_{15} y_6} dw_{15} \tag{6}$$

$$dy_7 = \left(\frac{k_6 \cdot LuxR^2 \cdot {y_4}^2}{Na \cdot V_1} - k_{14} y_7 \right) dt + \sqrt{\frac{k_6 \cdot LuxR^2 \cdot {y_4}^2}{Na \cdot V_1}} dw_{11} - \sqrt{k_{14} y_7} dw_{16} \tag{7}$$

$$dy_8 = \left(\frac{k_9 \cdot RhlR^2 \cdot {y_3}^2}{Na \cdot V_2} - k_{13} y_8 \right) dt + \sqrt{\frac{k_9 \cdot RhlR^2 \cdot {y_3}^2}{Na \cdot V_2}} dw_8 - \sqrt{k_{13} y_8} dw_{17} \tag{8}$$

$$dy_9 = \left(\frac{k_{10} y_8^{n_2}}{k_{10b}^{n_2} + y_8^{n_2}} - k_{16} y_9 \right) dt + \sqrt{\frac{k_{10} y_8^{n_2}}{k_{10b}^{n_2} + y_8^{n_2}}} dw_{18} - \sqrt{k_{16} y_9} dw_{19} \tag{9}$$