RICE UNIVERSITY

Protein Degradation in Synthetic Gene Circuits

by

Erin O'Brien Gilbert

A Thesis Submitted

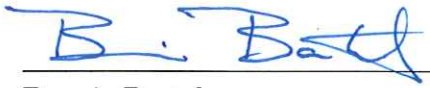
in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

Matthew R. Bennett, Chair
Associate Professor of Biosciences and Bioengineering



Bonnie Bartel
Ralph and Dorothy Looney Professor
Professor of Biosciences

Michael Gustin
Professor of Biosciences

Kathleen Matthews
Stewart Memorial Professor of Biochemistry and Cell Biology in Biosciences

Laura Segatori
Associate Professor of Chemical and Biomolecular Engineering, Bioengineering, and Biosciences

Houston, Texas

April, 2018

ABSTRACT

Protein Degradation in Synthetic Gene Circuits

by

Erin O'Brien Gilbert

Synthetic gene circuits are built with mathematical predictions and are further characterized experimentally. Most synthetic circuits utilize degradation tags to normalize and speed up the rate of degradation of circuit components. Despite their widespread use, the effect of degradation tags on circuit dynamics has not been well studied. This work aims to characterize the degradation rate of the ssrA degradation tag variants on a single substrate level and determine their role in overall network dynamics within a synthetic gene circuit at the single cell resolution. Small differences in the protein degradations rates indicating that the parameter space for the degradation tags is critical for achieving desired circuit dynamics. Mathematically and experimentally, this work demonstrates varying the rate of degradation can ultimately dictate the output oscillations for the circuit dynamics. The ultimate goal of this work is to create a better understanding of the role degradation plays in synthetic gene circuit dynamics.

Acknowledgments

My graduate experience at Rice has fundamentally changed who I am as a woman and as a scientist. There are many people who have influenced this growth, and I am extremely grateful that I have had strong network of support, both personally and scientifically.

First, I would like to thank my husband, Travis, for all of his love and support throughout my graduate career. Additionally, our families have been a source of relentless love and support. My parents, sisters (Colleen, Caitlin, and Regan), and my Aunt Jenny have been my cheerleaders throughout this time period.

One the greatest gifts graduate school has given me is a network of strong, intelligent female scientists. I am better person today because of my friendship with these women. Evi, Faiza, Anisha, Kim, and Marcella, you are all brilliant and thoughtful women. I am honored to be considered your friend and colleague. Your belief in me, especially when I am doubtful of myself, is one of the reasons I have made it to the finish line.

Lastly, I would like to thank Matt for all his mentorship and patience throughout these years. I'm grateful for all of the advice and inspiration inside and outside of the lab.

Contents

Figures

Tables

Chapter 1

Introduction to Synthetic Biology

Synthetic biology is a field of interdisciplinary research involving applied mathematics, engineering, and biology. This new area of science uses a bottom-up approach to engineer novel genetic circuits that allow the study of dynamics in native gene networks or build circuits that generate new functions for the host organism. Ideally, this field would generate a library of standardized, orthogonal, and modular biological parts. Typically, the design process would begin with developing a new circuit on paper with the appropriate parts: promoters, repressors, and activators. Next, a computational model would be built to coarsely tune the dynamic concentration range of each part and make overall dynamic predictions about the circuit. The most time-intensive step would involve building the circuit on the bench. Often, the computational tuning predictions are incorrect, leading to an "ad hoc" method of building the circuit. This library of available biological parts is completely dependent on network architecture, environmental conditions, and the host genetic background. The lack of modularity in the biological parts available to the synthetic biologist creates a significant time lag in bringing a new synthetic circuit design to fruition. Substantial effort to characterize, both mathematically and experimentally, this library of parts, as well as to expanding the modular components.

Using synthetic gene circuits, this research aims to understand endogenous gene networks by breaking complex networks into simple components. These synthetic gene

circuits are built from mathematical predictions and are characterized experimentally.

1.1 The Molecular Biology Foundation of Synthetic Biology

A critical publication for the foundation of synthetic biology was the examination of the dynamic response of the *lac* operon in *E. coli* to the cell environment by Francois Jacob and Jacques Monod in 1961 [16]. Jacob and Monod then envisioned the ability to assemble novel regulatory systems from native molecular components [16, 17]. A subsequent publication that further resolved the molecular details of transcriptional regulation in bacteria further cemented the idea of programmable gene expression. In combination with the expansion of knowledge of bacterial gene regulation and the invention of the polymerase chain reaction (PCR) method, the capability to engineer artificial gene regulation became technically possible [17, 18]. Another monumental step towards building a novel synthetic circuit was the ability to sequence entire bacterial genomes, which expanded the library of cellular components. The ability to sequence whole genomes created the field of systems biology, requiring both biologists and computer scientists working together to resolve large sequencing datasets in order to reverse-engineer cellular networks [19–21]. Systems biology utilized a top-down approach to resolving gene regulatory networks and demonstrated that, like engineered systems, these networks had distinct functional modules [22]. The expansion of potential molecular parts slowly generated a more rational, bottom-up approach to gene regulatory dynamics, forming the basis of a formal molecular biological engineering discipline that could forward-engineer synthetic gene networks that could demonstrate non-native cellular dynamics [17, 23]. The physical realization of this bottom-up approach to molecular biology occurred when collaborative efforts by

engineers, physicists, and computer scientists capitalized on the opportunity to use previous molecular biology accomplishments to create a simple gene regulatory network that would carry out a simple function, analogous to an electrical circuit [24,25].

1.2 The First Synthetic Gene Circuits

The Goodwin Oscillator

The Goodwin oscillator was the first synthetic gene oscillator to be theoretically developed [26] (although it was not experimentally built until much later [5]). This simple oscillator has a single gene that represses itself (Fig. 1.1 (A)). The protein concentration of the repressor oscillates between high and low states; as the system is induced, the amount of protein increases, binding to the promoter and shutting down protein expression. Then, as the repressor concentration decreases, the probability of a repressor binding to the promoter decreases, allowing the expression from the promoter to increase, and the systems restarts. Oscillatory behavior was achieved through *in silico* analysis. This early work explored a variety of mathematical modeling techniques and determined that, within biologically relevant parameter ranges, oscillations can be demonstrated using stochastic simulations [27, 28]. These initial simulations found that the time delay was an important characteristic of oscillatory systems. Biologically, this delay originates from processes in the cell, such as, transcription, translation, and protein folding [27]. However, further examination of the simulations suggested that too long a time delay decreased the stability and robustness of the oscillatory behavior [29].

Experimentally, the Goodwin oscillator was built using the LacI inducible promoter,

$P_{Llac0\text{-}1}$, [10] driving the expression of the repressor LacI [5] (Fig. 1.1 (B)). A negative feedback loop is formed when LacI represses the $P_{Llac0\text{-}1}$ promotor, inhibiting it's expression. To observe circuit behavior, the fluorescent protein, the gene encoding the green fluorscent protein (GFP) was also placed under $P_{Llac0\text{-}1}$ control. Additionally, to have observable dynamic behavior, the ssrA degradation tag, which is recognized by native *E. coli* proteases, was placed at the C-terminus of each protein. Adding these degradation tags ensures rapid protein turnover, enabling dynamics to occur on a reasonably fast timescale. Oscillatory behavior was erratic and noisy, but agreed with mathematical predictions; however, the oscillatory periods were longer than predicted (10 to 20 minutes predicted verses 30 minutes observed). These results highlight the need for building mathematical models to simulate circuit behavior, and the importance of understanding parameter ranges and limitations to guide the biological implementation of these synthetic gene circuits.

The Toggle Switch

The first synthetic gene circuit built was a bistable toggle switch composed of the two repressors LacI and cI [30]. The bistable topology of this circuit is built by having the *lacI* gene under the regulation of the *cI* promoter and the *cI* and *gfp* genes under the control of the *lacI* promoter. Isopropyl β-D-1-thiogalactopyranoside (IPTG), the ligand for LacI, was added to induce the *cI* expression state. This releases the repression of cI and GFP expression, and the amount of green fluorescence increases. Bistable behavior is exhibited because the system can only exist in one state or the other, not both. The system toggles between one state or the other based on an environmental stimulus. When IPTG is removed from the system, the system persists (often referred to as cellular "memory") in the LacI state; however, once heat

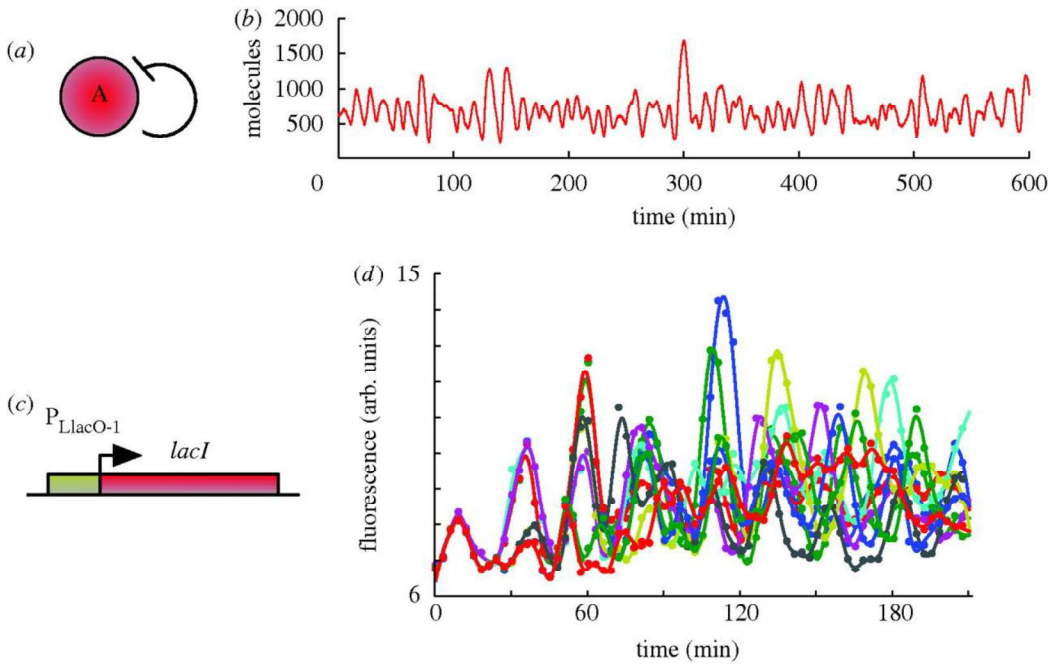


Figure 1.1 : **The Goodwin oscillator. (A)** This oscillator is composed of a single gene that represses itself. **(B)** Mathematical simulations predicted short, noisy oscillations. (C) *In vivo* implementation of the circuit (D) demonstrated predicted noisy behavior with longer oscillations. Figure adapted from [1] with permission.

is introduced (in the absence of IPTG), the cI state is activated. cI is expressed and begins repression of LacI, causing a sharp transition into the cI state. Experimentally, this state is observed when the amount of green fluorescence begins to decline. This synthetic gene circuit was demonstrated our ability to design, build, and implement non-native behavior from engineered DNA.

The Repressilator

The first synthetic gene oscillator, the repressilator, was built the same year as the bistable toggle switch [31]. Building on the Goodwin oscillator foundation, the repressilator was composed of a ring of three repressors, LacI, TetR, and cI (Fig. 2.7 (A)). The topology of three interconnected repressors is also seen in electronics (ring oscillators) and in neuroscience (neural ring networks) [1]. The *cI* promoter regulated the expression of *tetR*, the *tetR* promoter regulated the expression of *lacI* and the reporter, *gfp*, and the *lacI* promoter regulated the expression of *cI*. Temporal oscillations are generated from this repressor ring design. Specifically, LacI inhibits the production of TetR and GFP, relieving the inhibition of TetR on the cI promoter, causing increased production of the cI repressor. As cI accumulates, it inhibits the expression of LacI, relieving the inhibition of LacI on TetR, and allowing the concentration of TetR to increase. TetR represses the production of cI, relieving the repression of LacI expression, and starting the cycle all over again.

The *in vivo* implementation of the repressilator demonstrated cell cycle-independent, large, variable-amplitude oscillations in 40% of cells [31]. Mathematical simulations correctly predicted this behavior. The amount of GFP gradually increased over time, which was thought to be a consequence of GFP stability, but was not mathematically

accounted for (Fig. 2.7 (B)). Additionally, all dynamics ceased once the cell entered the stationary phase, indicating that circuit behavior was coupled to global regulation and the effects of cell growth and division, allowing the system to relax to an equilibrium state [1, 31].

Both the toggle switch and the repressilator demonstrated a rational, mathematical model-based approach to design and build these synthetic gene circuits. However, agreement between model and experimental outcome was only achieved after tedious, iterative, experimental tuning of the synthetic gene circuit. These pivotal synthetic gene circuits established a foundational workflow that included rational design, mathematical modeling, experimental construction, experimental testing of circuit, and then extensive experimental troubleshooting to achieve the desired circuit behavior [17]. The necessity for extensive experimental troubleshooting identifies a noticeable gap between the mathematical understanding of the design and the experimental realities, highlighting the importance of mathematically characterizing individual circuit parts, such as the work done in this thesis. Increasing the accuracy of these mathematical models can more efficiently guide the experimental construction, and eliminate the amount of troubleshooting and re-engineering needed. A more detailed explanation of these mathematical models is provided in the next chapter.

1.3 Understanding Synthetic Gene Oscillators and The Importance of Topology

Positive Feedback Influence Oscillator Dynamics

The previous synthetic gene circuits described above only incorporated negative regu-

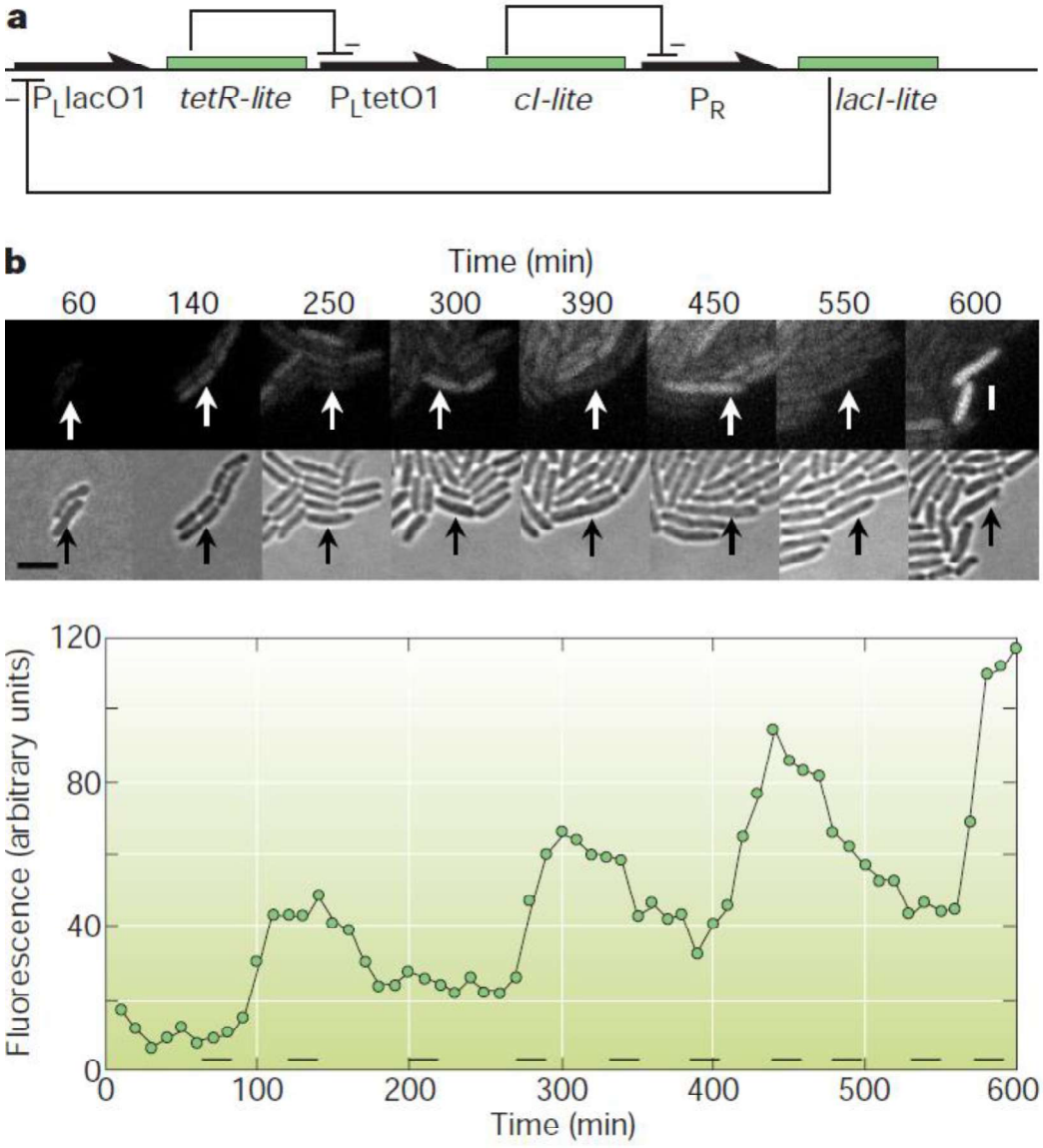


Figure 1.2 : **The Repressilator oscillator. (A)** The Repressilator is composed of a ring of three repressors, LacI, TetR, and cI. **(B)** *In vivo* implementation of the repressilator agreed with mathematical predictions and produced variable amplitude oscillations in a portion of the cell population [2]. Reprinted with permission.

lation between circuit genes. To further understand the fundamentals of the contribution of gene regulation to overall circuit dynamics, the addition of positive regulation, or feedback, was examined. *In silico* mathematical simulation studies of the addition of different types of positive feedback is comprehensively investigated in the next chapter.

In 2003, Atkinson *et al.* built a synthetic gene oscillator composed of an activator, NRI, and a repressor, LacI, where *glnG* (which encodes NRI) activates the transcription of *lacI* under the glnKp promoter. LacI then represses the expression of *glnG* under the glnApO promoter. However, an additional positive feedback element is built into the system by having NRI activate its own expression (Fig. 1.3 (A-B)). Circuit dynamics were recorded at the population level by utilizing the β-galactoside reporting system, and at the single-cell level by fusing CFP to the LacI promoter. The oscillator yielded three damped oscillation cycles at the population and the single cell levels, agreeing with mathematical modeling predictions. Originally, each gene in the system was integrated into the *E. coli* chromosome as a single copy; however, it was possible to increase the number of oscillations to four by increasing the copy number of the activator, NRI, near a section of the chromosome that undergoes four-fold higher transcription [11]. This amplified negative feedback through transcriptional control showed excellent agreement between mathematical simulations and experimental dynamics, at both the population and single-cell levels. Additionally, it showed that the addition of a positive feedback auto-activation loop caused the system to produce damped oscillations (Fig. 1.3 (C-D)). However, this oscillator did not show sustainable or robust oscillations. This laid the foundation for exploring the topology of the dual feedback oscillator.

The Fussenegger Oscillators

To generate sustain oscillations, mathematical analysis determined that the activator needed faster dynamics compared to the repressor. As discussed in the previous section, this can be accomplished through negative feedback. Another way to integrate this characteristic is to add a delay to the negative feedback loop [1]. Both of these integrations allow the activator to increase to a critical concentration before the repressor settles the system into a repressed steady state, thereby promoting sustained oscillator dynamics [1]. These ideas were actualized in the oscillators built in Tigges *et al.* [3, 32], commonly referred to as the Fussenger oscillators.

The Fussenegger oscillators were the first synthetic gene oscillators to be built in an eukaryotic system (Fig. 1.4 (A)). These oscillators took advantage of the higher complexity of the eukaryotic system and built regulation into the oscillator through sense and anti-sense interactions. The first iteration of this oscillator had two genes, which both had regulatory sense and anti-sense transcripts. The translated protein from the sense transcript activates its own expression and that of the second gene. This second gene acts a repressor by activating the anti-sense transcript from the first gene, which represses the translation of sense transcript by hybridizing to it, preventing protein production (Fig. 1.4 (B)). This interaction creates an amplified negative feedback loop, but varies from the previously discussed amplified negative feedback oscillator due to the added delay of repression from anti-sense transcript dynamics [3].

Experimentally, the sense transcript positive feedback loop was formed by the tetracycline-dependent transactivator protein (tTA) under the $P_{hCMV*-1}$ promoter. This promoter

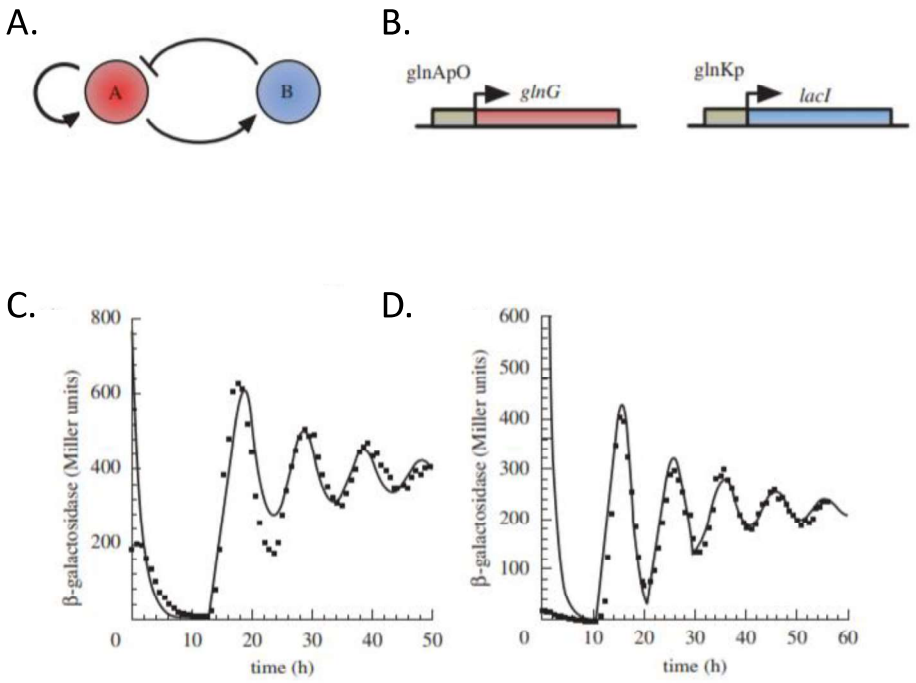


Figure 1.3 : **Positive feedback influences oscillator behavior.** **(A)** This oscillator has amplifed negative feedback topology. **(B)** *In vivo* implementation of this oscillator which **(C-D)** produced damped oscillations which was accurately predicted by mathematical simulations. Figure adapted from [1] with permission.

also controlled the expression pristinamycin-dependent transactivator protein (PIT). Therefore, as tTA promoted its own transcription, it also activated the expression of PIT. The ability of tTA to bind the $P_{hCMV*-1}$ promoter could be inhibited by the presence of tetracycline. The negative feedback loop was formed as PIT bound to the P_{PIR} promoter, which controlled the expression of the *tTA* antisense transcript. The ability of PIT to bind the P_{PIR} promoter could be inhibited by pristinamycin. Oscillator dynamics were observed through the expression of GFP under the control of $P_{hCMV*-1}$. This added delay in the negative feedback loop produced undamped oscillations, proving the importance of delay in synthetic gene oscillators. In addition, much of the observed behavior matched mathematical simulations, even when varying the circuit experimental conditions, testifying that oscillator dynamics could be mathematically predicted (Fig. 1.4 (C)). A "low-frequency" variant of this oscillator was later built, in which a delay in the negative feedback loop is attributed to direct siRNA interference, rather than to anti-sense RNA, which still produces undamped oscillations (Fig. 1.4 (D)) [32]. Importantly, the use of tunable amounts of RNAs suggested a relationship between gene dosage and the observed dynamics, offering synthetic biologists a mechanism by which to tune oscillator dynamics [1].

The Metabolator

The Metabolator was the first synthetic gene oscillator to use metabolites to drive circuit behavior. On a gross scale, the metabolator has two genes. One gene, gene A, converts a metabolic pool (M_2) to metabolic pool (M_1). The expression of gene A is activated by the presence of (M_2). (M_2) also represses gene B, which converts (M_1) to (M_2). Lastly, there is an influx into (M_1) and an efflux from (M_2) [1] [4]. Simply put, gene A activates the expression of gene B while repressing its own expression.

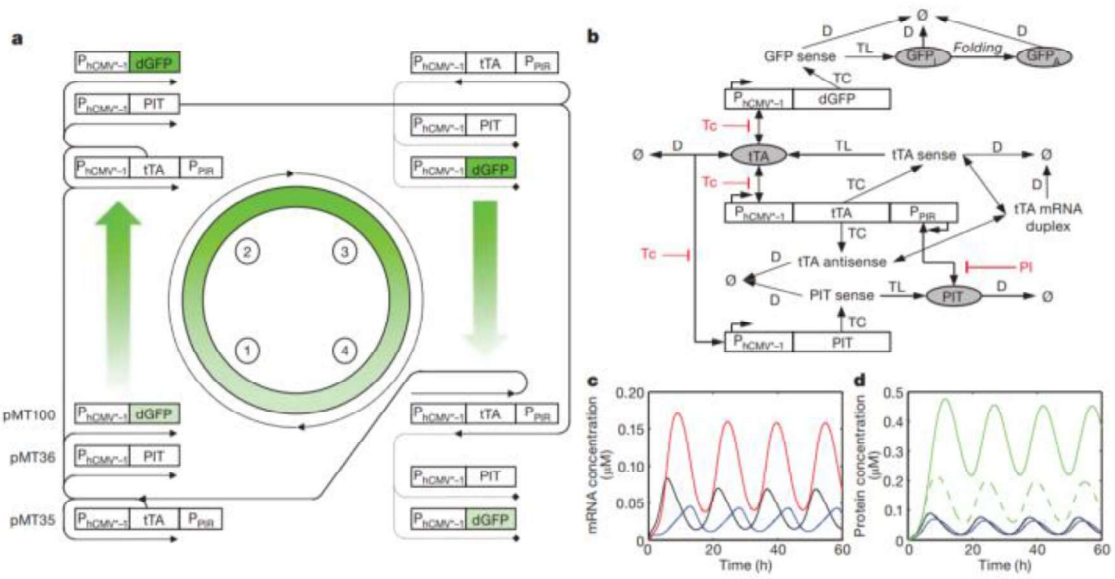


Figure 1.4 : **The Fussenegger oscillator. (A)** The general *in vivo* implementation of the Fussenegger oscillator and the **(B)** network interactions of the synthetic gene oscillator. **(C)** Undamped oscillations were archived through a negative feedback loop and matched mathematical prediction. **(D)** A low frequency version of the oscillator still produced undamped oscillations. [3]. Reprinted with permission.

Gene B activates the transcription of gene A, while repressing its own transcription.

This oscillator was implemented experimentally in *E. coli* using acetyl coenzyme A (acetyl-CoA) for (M_1) and acetyl phosphate for (M_2). Phosphate acetyltransferase, gene A converts acetyl-CoA into acetyl phosphate. Then, through indirect repression, acetyl phosphate triggers the phosphorylation of Nitrogren Regulation 1 (NR_1), which activates the glnAp2 promoter that expresses LacI. The production of phosphate acetyltransferase is controlled by the $p_{Llac0\text{-}1}$ [10] promoter, which is repressed when LacI is present. Additionally, acetyl-CoA synthetase is also under the glnAp2 promoter, and is activated in the presence of NR_1. The influx of acetyl-CoA into the system is through sugars present in the growth media and the efflux of acetyl phosphate from the system occurs through the conversion of acetate kinase to acetate, which is then protonated and becomes permeable across the cell membrane (Fig. 1.5 (A-B)).

The metabolator produced oscillations that were independent of cell division ((Fig. 1.5 (C-E)). Interestingly, because key components of the circuit were metabolites, oscillations could be varied by changing the sugar in the medium from glucose to fructose, mannose, and glycerol. However, glycerol did not have a high enough influx of acetyl CoA to sustain oscillations. Oscillations could also be suppressed by either adding high concentrations of acetate or allowing for natural accumulation. This behavior was correctly predicted by mathematical simulations and showed that oscillations could be controlled by external sources [1].

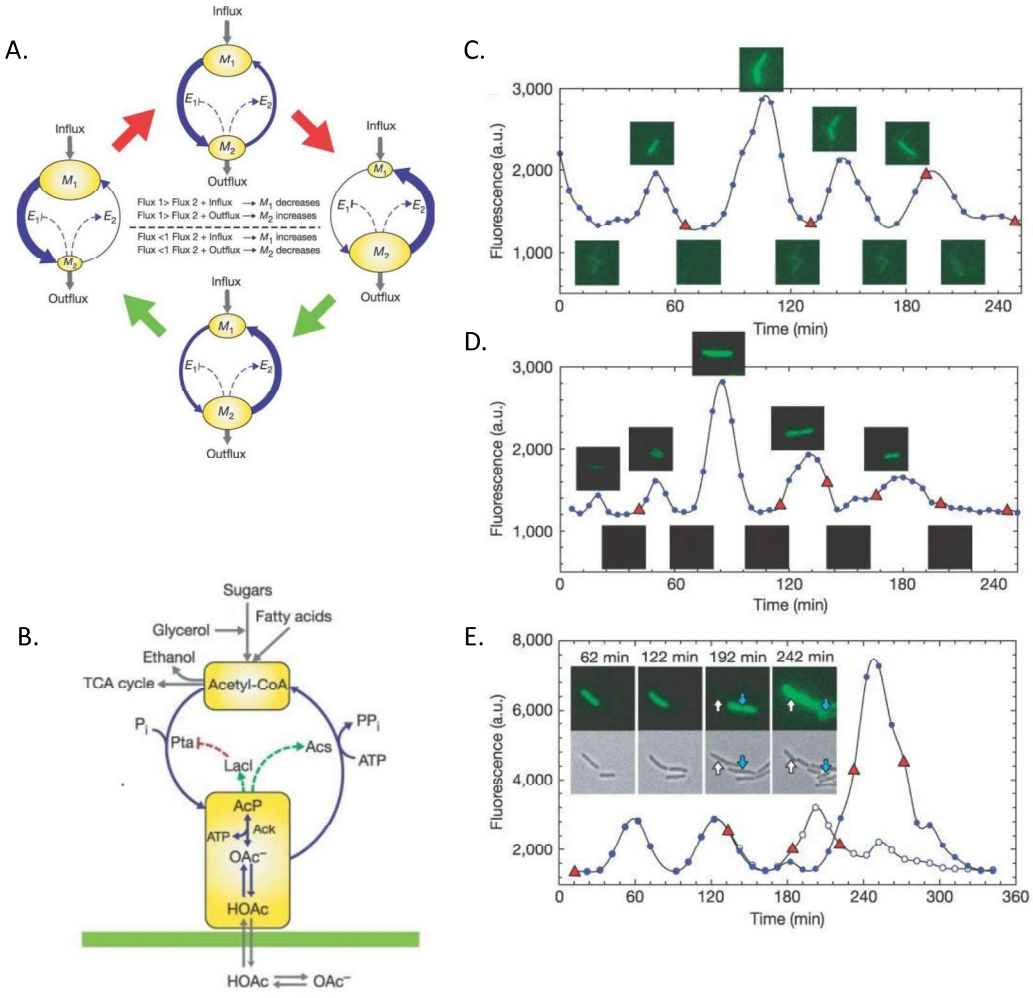


Figure 1.5 : **The Metabolator oscillator. (A)** The general network interaction of the metabolator oscillator flux, **(B)** the general metabolic interactions of the synthetic gene oscillator, and **(C-E)** the resulting single cell trajectories of observed cell cycle independent oscillations. Figured adapted from [4] with permission.

1.4 Dual Feedback Oscillator

Previously built synthetic gene oscillators did not produce robust observable sustained oscillation dynamics. Oscillation dynamics, if reported, were only observed in a small percentage of the cell populations. The Smolen oscillator [33] theoretically proposed a two-component oscillator composed of an activator, gene A, and a repressor, gene B. An *in vivo* implementation of this oscillator, termed the dual-feedback oscillator, was built by Stricker *et al.* [5]. The dual feedback oscillator was an oscillator built using the repressor LacI, the activator AraC, and the fluorescent reporter GFP. The oscillator was built using a hybrid that incorporated elements of the AraC and LacI promoters. This creates competition between the two transcription factors. When AraC "wins", all three genes are transcribed. When LacI 'wins', transcription of all three genes is repressed. In this way, the hybrid promoter creates oscillation of GFP florescence. *In silico* analysis provided critical insights into the physical manifestation of the dual feedback oscillator; for example, that oscillations were more likely to occur if the degradation rate of the activator is at least two times faster than that of the repressor [1, 34]. This analysis is discussed in more detail in Chapter 2 of this thesis.

Experimentally, the dual feedback oscillator was built on two plasmids, in which identical copies of the $P_{lac/arc\text{-}1}$ promoter controlled the expression of oscillator genes; *araC*, *lacI*, and *yemGFP* (monomeric yeast-enhanced GFP), and which were transformed into *E. coli* cells lacking LacI and AraC. Each protein was tagged with the C-terminal ssrA degradation tag. Circuit dynamics were initiated by the addition of IPTG and arabinose (Fig. 1.6 (A)). An important characteristic of this circuit is the robustness of the cell cycle-independent oscillations. Unlike previous implementations of synthetic gene oscillators, oscillations were observed in 99% of the cells.

Additionally, model simulations and observed dynamics matched well, demonstrating a range of conditions where oscillator dynamics were tunable (Fig. 1.6 (B)). Oscillations with 2 mM IPTG and 0.7% IPTG generated an oscillatory period of 40 minutes. This could be tuned to 13 to 58 minutes by changing the concentration of IPTG and arabinose, temperature, and types of growth media (LB or minimal media).

It is well understood that raising the temperature 10 °C, doubles the biochemical reaction rates in a cell; this is termed Arrhenius scaling of reaction rates. Therefore, the oscillatory period of the dual feedback oscillator decreased by roughly 50% for every 10 °C increase [35]. Resistance to temperature, or temperature compensation, is a common feature found in natural oscillators [36]. A more detailed investigation of the relationship between temperature and oscillation dynamics revealed that the dual feedback oscillator could be engineered to be resistant to temperature changes [37]. These studies demonstrate that dual feedback oscillator dynamics can be attributed to a number of different parameters. A greater understanding of the external (experimental conditions) and internal (circuit "parts") influences on genetic circuits, both experimentally and mathematically, provides insight to build future novel synthetic gene circuits. This thesis work focuses on examining the influence of degradation rates of circuit proteins on dual feedback oscillator oscillator dynamics.

One aspect of synthetic biology aims to expand the number of biological tools available, by taking advantage of native genomic resources (promoters, transcription factors, proteases, etc.) and optimizing these resources for specific synthetic gene networks. To obtain dynamic behavior, systematic, controlled protein turnover is required. Without protein degradation, it would be impossible to elicit oscillating

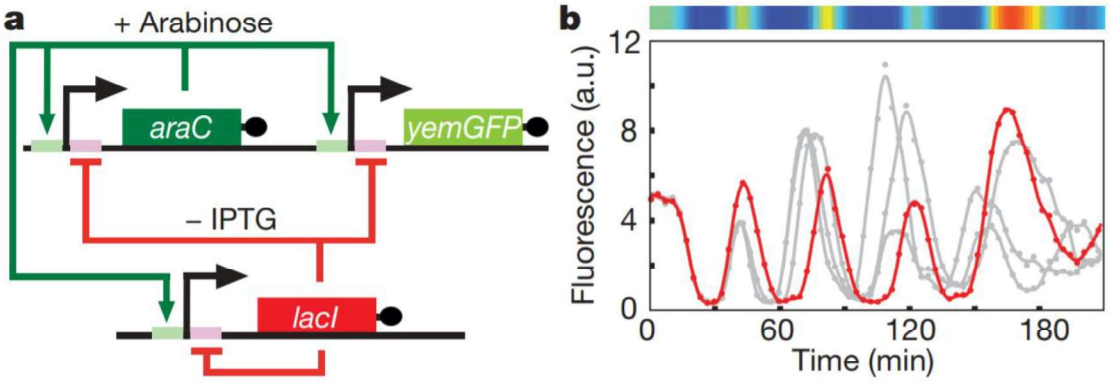


Figure 1.6 : **The dual feedback oscillator. (A)** The general *in vivo* implementation of the dual feedback oscillator. **(B)** Single cell trajectories of the dual feedback oscillator experimental data. Figure adapted from [5] with permission.

behavior from the dual feedback oscillator. Many synthetic gene circuits widely but blindly use ssrA degradation tags. These tags take advantage of the native ClpXP degradation pathway of *E. coli* and consist of an 11-amino acid tag (AANDENYALAA) that is attached to the C-terminus of a protein. Anderson *et al.*, using GFP tagged with ssrA variants, found that changing the last three amino acids had a significant impact on protein degradation rates (Fig. 1.7). However, these rates are a population average and are not calculated at single-cell resolution. Finer resolution of these degradation rates is needed to better mathematical model degradation in synthetic gene circuits. The work presented in this thesis quantifies the degradation rates of each ssrA variant on a single cell level as well as demonstrates the relationship between increasing degradation rates and changes in dual feedback oscillator oscillation periods. To accurately quantify these degradation rates, it was necessary to thoroughly understand the protein biochemical dynamics that contribute to changes in oscillatory behavior.

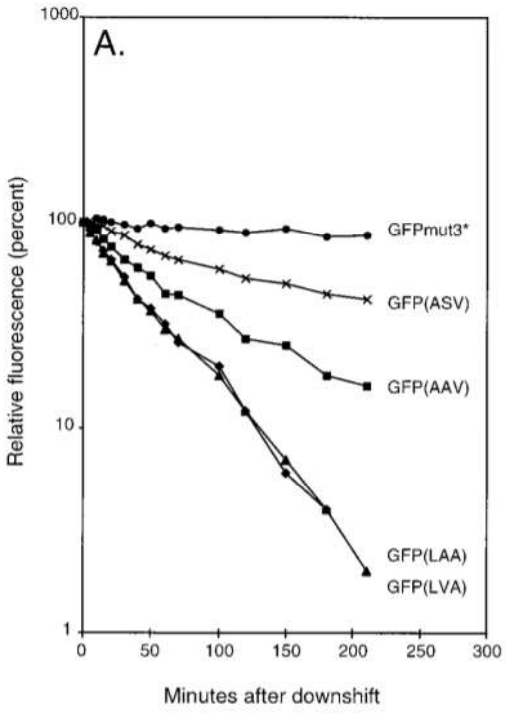


Figure 1.7 : **ssrA degradation variants.** Varying the last three amino acids of the ssrA degradation tag has large effects on fluorescence decay. Changing these amino acids decreases the binding ability of the tagged fluorescent protein to ClpX, resulting in a sustained fluorescence signal. GFP tagged with the ssrA variants, AAV and ASV, is significantly more stable that GFP tagged with the wildtype sequence of ssrA. Figure adaptred from [6] with permission.

1.5 ClpXP Protein Degradation in Synthetic Gene Circuits

All organisms have native protein degradation machinery, consisting of proteases, that removes misfolded or aggregated proteins from the cell to prevent the damage or disruption of normal cellular processes. These native degradation machineries simplify synthetic circuit design and construction because they do not need to be built or cloned into the host organism. Targeted protein degradation needs to only occur for proteins involved in circuit dynamics for synthetic gene circuits. There are two components that need to be considered when building a synthetic gene circuit. What native protease will be used, and how will the proteins be recognized by the protease for targeted degradation?

Most synthetic gene circuits utilize the ssrA protein degradation tag system for rapid protein turnover for faster circuit dynamics. The ssrA degradation system is a quality control protease pathway in *E. coli* that prevents the buildup of mistranslated proteins [38] that target the protease, ClpXP. This ssrA-mediated ClpXP degradation is an essential process in *E. coli*, as 1 in 200 protein synthesis events terminates with ssrA tagging [38]. SsrA (small stable RNA A molecule) was first discovered in *E. coli*; it was determined to be composed of 362 nucleotides and have structural and functional similarities to alanyl-tRNA synthetase [39]. The first evidence that the ssrA gene in *E. coli* had a role in protein quality control was discovered when truncated proteins found in inclusion bodies had an 11-amino acid modification (AANDENYALAA) at the C-terminus of the truncation; this addition did not occur in cells when the ssrA gene was deleted [39].

ClpXP is a well-conserved protease complex that was first discovered in *E. coli* and

has since been extensively characterized [40]. ClpXP is composed of two distinct proteins, an AAA+ ATPase, ClpX and a peptidase, ClpP. Together they form the barrel-shaped protease, ClpXP. Additionally, the adapter binding protein, SspB, is responsible for delivering ssrA-tagged proteins to the ClpXP complex and anchoring the bound protein to ClpX. Anchoring the ssrA-tagged protein reduces slipping and increases ClpX's grip on the docked protein, allowing the maximum amount of force to be transferred to the tagged protein for unfolding and eventual digestion [41, 42]. The mechanics of the coordinated efforts of ClpP, ClpX, and SspB are described below.

1.5.1 ClpX AAA+ ATPase

The Structure-Function Relationship of ClpX

AAA+ enzymes are defined by their ability to harness the energy of ATP binding and hydrolysis to perform mechanical work to power numerous biological reactions and processes. Of specific importance for this work, this mechanical work is the force required to degrade fully folded proteins that are often in stable tertiary and quaternary structures [43].

The homohexameric ring structure of ClpX is critical for performing the mechanical functions necessary to degrade stable proteins. A single ClpX monomer contains three important domains: the N-domain, which contributes to the overall stability of the complex and is responsible for adapter and substrate recognition, the large AAA+ domain where ClpP binds, and the small AAA+ domain [44–47]. The AAA+ domains form the hexomeric ring and perform all the mechanical functions of ClpX [44, 47–49].

ATP binding and hydrolysis occurs in the hinge regions between the small and large AAA+ domains [8,50]. Interestingly, the N-domain forms a dimer independent of the rest of the protein. In contrast to the rigid AAA+ domains that form the hexomeric ring and carry, the N-domains are flexibly tethered [8, 44, 47, 51].

The orientation between the AAA+ subunits can considerably vary in monomers of the ClpX hexomer [8] (Fig. 1.8 (A)). Functional differences arise from this variation between monomers. The ability to bind nucleotides dictates the difference between the two possible domain-domain orientations. Of the six monomers, four are "loadable" (or L) subunits where the domains can bind ATP (Fig. 1.8 (B)). The two "unloadable" (or U) monomers have an approximate 80° rotation between the small and large AAA+ domains that eliminate the ATP binding site. These subunits are arranged in an L-L-U-L-L-U pattern resulting in an asymmetric structure, where a maximum of four ATP molecules can bind [8] (Fig. 1.8 (C)). This redundancy in ATP hydrolysis functions to increase the rate of unfolding and translocation by ClpX. Martin *et al.* demonstrated, using engineered single chain molecules of ClpX, that a construct with only one subunit is capable of ATP hydrolysis and is sufficient to drive the conformational changes need to carry out the mechanical functions to unfold and translocate substrates [48].

In clockwise fashion, each small AAA+ domain is tightly compressed against the large domain of the neighboring subunit [8]. The ClpX hexamer is stabilized by vast subunit-subunit connections where rigid-body interactions bury nearly 2000 $Å^2$ of the surface area. The region between the two AAA+ domains acts as a hinge. Prior to ATP binding, the ring is topologically closed; however, ATP binding triggers confor-

mational changes in the orientation of the small and large AAA+ subunits, which propagates around the entire ring to aid in substrate unfolding and translocation [40].

ClpX Mechanism of ssrA and SspB Recognition

ClpX recognizes proteins substrates that have a large variety of short peptide sequences on both the N- and C-terminal of proteins, as well as the ability to bind many adapter proteins [42, 52]. For the purpose of this work, only the mechanisms of ssrA and SspB recognition will be discussed.

To remove incomplete peptide sequences stalled on ribosomes, an 11-amino acid peptide sequence is added to an incomplete nascent protein in eubacteria. The transfer-messenger RNA (tmRNA) system is triggered when a ribosome stalls on the transcript, and then the 11-amino acid peptide tag (AANDENYALAA) is translated from the ssrA in tRNA-like fashion [39,53]. The ssrA peptide tag has two important recognition sequences. The adapter protein, SspB, recognizes the first seven amino acids of the ssrA tag (AADENY), whereas the C-terminal alanines with the negatively charged α-carboxyl group determine ClpX recognition [41, 53, 54].

SspB delievers ssrA-tagged substrates to ClpXP, working in concert with the degradation complex and thereby modulating proteolysis [40, 41, 55]. Although the ClpX hexamer demonstrates proteolytic ability outside of the ClpXP-SspB complex, degradation is greatly enhanced with the binding of SspB [41]. Active SspB is a dimer with two important structural features. First, the SspB dimer has a pocket that recognizes the first seven amino acids of the ssrA tag [41, 56–58]. Second, each monomer in the dimer has a highly flexible C-terminal tail that docks with the N-domain of

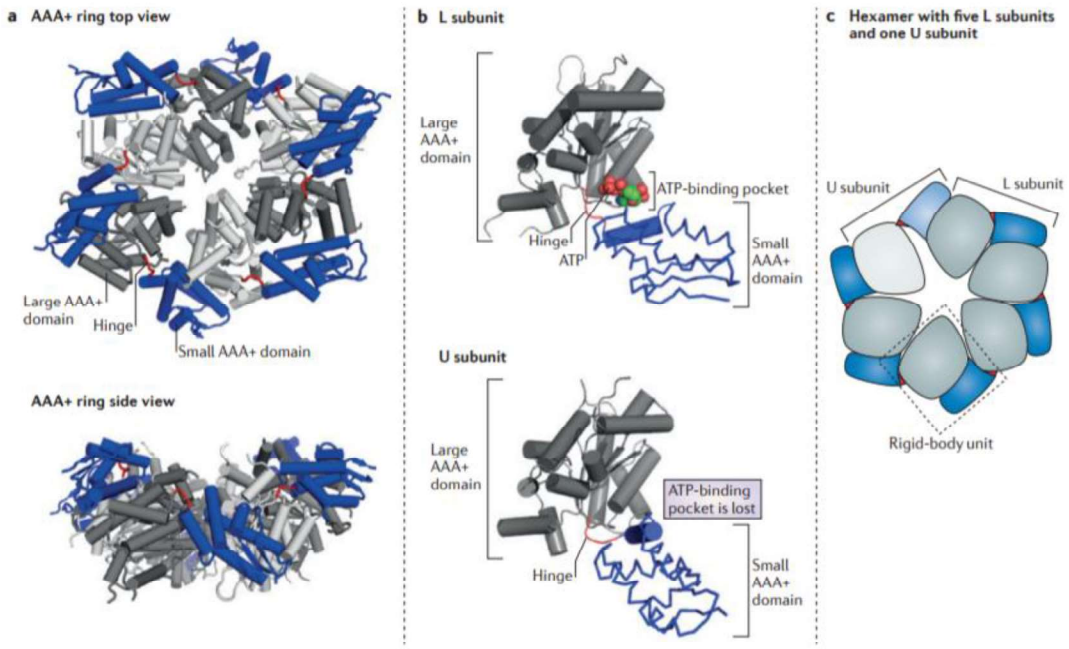


Figure 1.8 : **The structure of ClpX. (A)** The protein structure of the assembled hexomeric ring of ClpX. **(B)** Four of the six monomers are loadable (L) subunits, where tagged substrates can bind and two of the monomers are unloadable (U), providing a rigid body to support degradation activity. **(C)** Schematic of the orientation of the loadable and unloadable monomers in the hexomeric ring. [7]. Reprinted with permission.

ClpX [59, 61]. For efficient degradation, both tails of the SspB dimers must make multivalent contacts with ClpX, ensuring high-affinity tethering of the ssrA-tagged substrate [60]. It should be noted that it is essential that the ssrA degradation tag be engaged with SspB and ClpX at the same time. Orchestrating binding in this way results in strong, tight binding of the substrate to the degradation machinery, and ensures efficient degradation in low ClpXP concentrations. Briefly, the affinity of each SspB tail to the N-domain of ClpX is about 20 μM, and that of ssrA binding to the ClpX pore is about 1 μM, but the coordination of all three binding events increases the affinity of ClpX to the SspB-ssrA substrate complex to 15 nM [60, 61].

Unfolding and Translocation Mechanics of ClpX

The main role of ClpX is to use mechanical force to translocate (spool) proteins through the axial pore and into the ClpP chamber. This translocation through ClpX's small pore often prompts the unfolding of the larger ssrA-tagged substrate. ClpXP must exert an incredible amount of force to disrupt the stable delivery complex to unfold the tagged substrate and facilitate degradation. This force termed a "power stroke". When ClpX does not have a good grip on the substrate, this power stroke is not effective because it cannot transmit the force efficiently [40]. This stable delivery complex attached to the ATPase emphasizes the mechanical force that ClpXP exerts on an ssrA-tagged substrate, to stimulate disruption of the structure to eventually be fed through the protease. Bolon *et al.*, permanently bound GFP-ssrA to the SspB dimer through a disulfide bond. Unable to mechanically separate the ssrA substrate from SspB, the protease degraded both the ssrA substrate and SspB [60]. ClpXP is natively capable of degrading disulfide-bonded substrates by translocating two or more peptides through the pore [60,62]. The hinge region between the large and small

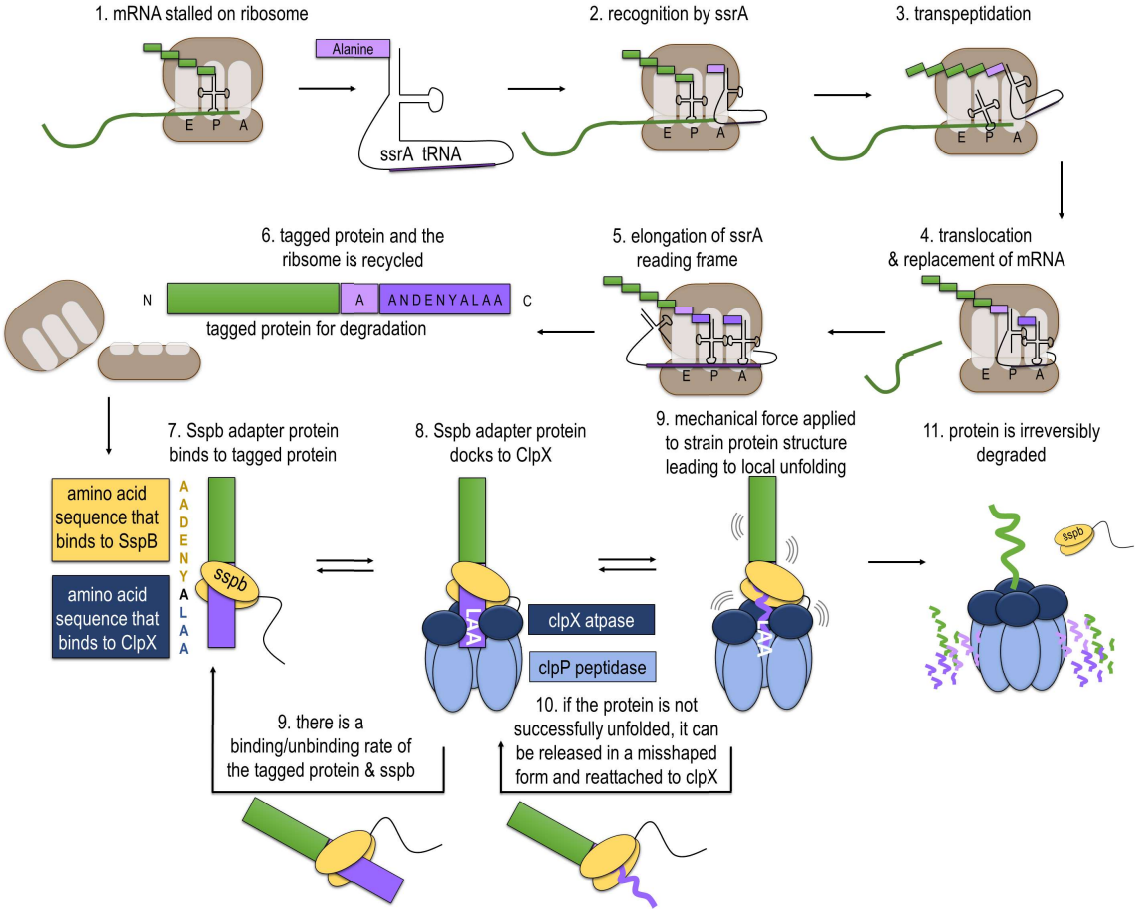


Figure 1.9 : **Mechanism of ssrA target degradation by the ClpXP degradation machinery.** When a stalled ribosome has a peptide, it is tagged on the C-terminus by the ssrA protein sequence. The stalled ribosome signals for the ssrA mRNA, and the ribosome reads the tag sequence, which is subsequently added to the peptide. The tagged protein is released by the ribosome and is targeted by the ClpX subunit. The ClpX subunit then denatures the peptide sequence and translocates the peptide into the ClpP subunit. The ClpP subunit degrades the peptide [8].

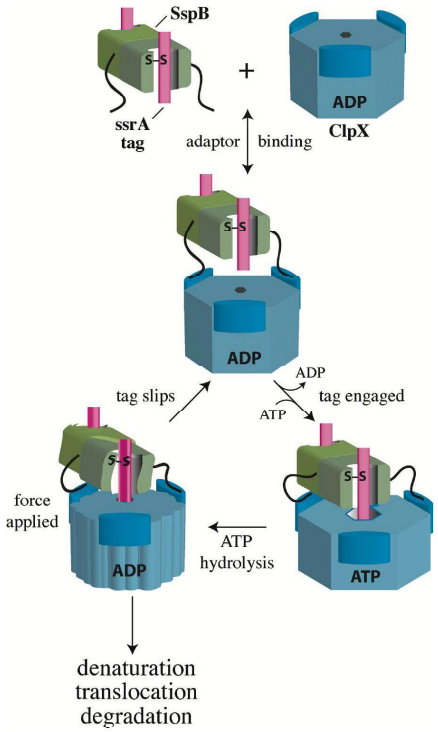


Figure 1.10 : **ClpX mechanism of ssrA and SspB recognition.** When a tagged substrate is released from the ribosome, SspB recognizes the ssrA tag and binds to the first six amino acids. SspB aids in substrate binding to ClpX by anchoring to ClpX. However, if the tag slips, SspB remains bound to the tagged substrate and attempts another binding event. After degradation begins, SspB is forced off as the substrate translocated into ClpP [9]. Reprinted with permission.

AAA+ domains of unloadable ClpX monomers unfolds to widen the pore to conform to larger substrates and subsequently refolds to close the pore.

After the ssrA is translocated, ClpX grips the subsequent peptide chain through van der Waal's interactions with carbonyl oxygens along the peptide chain [63]. The conformations of unfolded peptides, and the amino acid step size by which the peptide strand is translocated, vary by protein sequence.

1.5.2 ClpP Peptidase

Structure and Function of ClpP

ClpP is the proteolytic component of the ClpXP protease. To prevent unregulated degradation, ClpP is expressed as a proenzyme and undergoes autoproteolysis to remove an N-terminal peptide sequence to generate the mature enzyme [64]. ClpP is a homoheptamer ring that forms a barrel shape. This final barrel shape is constructed from two homoheptamer ring structures that stack to form a double-ring tetradecamer [65, 66]. The mature assembly is then capped or doubly capped by an ATPase cohort, which prevents rogue degradation [51]. Small axial pores allow entrance into the proteolytic chamber. The active and inactive confirmations of ClpP result from the distance between these axial pores [64, 66–70]. The pores regulate degradation because they are large enough to allow small peptides to enter. To prevent clogging, the pores exclude folded proteins and specifically slow the entry of large unfolded proteins. The proteolytic chamber has a diameter of roughly 50 Å [71].

ClpP will cleave any unfolded peptide chain that is translocated into its proteolytic chamber. The proteolytic cleavage requires ATP hydrolysis It should be noted that the peptidase does not digest the peptide into single amino acids, rather it chops the peptide chain into small enough fragments to exit the chamber. Free floating exopeptidases then degrade these smaller fragments into single, free amino acids [72, 73]. ClpP preferentially cleaves non-polar residues, specifically Leu, Gly, Met, Ala, and Tyr [40].

The structure of the proteolytic chamber gives insights to the mechanisms of peptide cleavage. The proteolytically active sites of ClpP reside within a barrel shaped chamber, formed by the face-to-face stacking of two heptameric Clp rings [51]. Most

importantly, the high local concentration of active sites and the orientation and distance of these 14 active sites from each other inside the chamber govern the efficiency of cleavage. The high local concentration close to 350 mM of active sites means that proximity drives active site binding therefore decreases the need for high specificity, ensuring that weakly interacting sequences can be degraded. It is easy to understand the lack of a need for specificity for a protease responsible for degrading unwanted, misfolded proteins. The specificity of degradation is regulated via the addition of the ssrA degradation tag. To further illustrate this point, each active site must be 25 Å away from three other active sites, meaning this distance could be bridged by eight amino acids. Consequently, a substrate can bind to multiple active sites and can be simultaneously degraded. These tandem-cleavage events increase the rate at which ClpP can digest a peptide [74, 75].

1.5.3 ClpXP-SspB Degradation Machinery

Structure and Function of ClpXP

Although both ClpX and ClpP have ring structures, the difference in the number of monomers, six and seven, respectively, causes an asymmetric fit when ClpX caps ClpP [40]. However, the axial pores of ClpX align with the pores of the ClpP rings, providing entrance into the proteolytic chamber of ClpP. This formation of ClpXP influences the function of each of the subunits [8, 39, 49, 54]. Despite being doubly capped by ClpX, translocation is coordinated through ClpP and can only occur from one of the two ClpX rings [76]. The tight binding of these two proteins absolutely requires ATP; no ClpXP proteases are detected in the absence of ATP [51,54,77]. It is not clear if conformational changes of ClpP occur concurrently with ATP binding and

hydrolysis by ClpX. It is possible that the binding of ClpX stabilizes the active conformational state of ClpP during ATP-driven substrate unfolding and translocation. Additionally, it is possible that structural conformational states in ClpX and ClpP are coupled to facilitate substrate translocation, degradation, and product release [40].

Steady State Kinetics of ClpXP

Evidence of this influence of ClpP on ClpX behavior is found in the steady-state kinetic parameters for ClpX hydrolysis of ATP when bound to Clp with or without a protein substrate [39, 78]. Succinctly, V_{max} linearly correlates with K_M. Biochemically, this linear relationship indicates that ATP molecules that bind to the ClpXP complex are hydrolyzed, and the resulting ADP molecules dissociate [78]. Individual ClpX monomers hydrolyze ATP at a rate of 100 to 600 min^{-1}. This rate varies depending on ClpP binding and the presence or absence of a substrate [78,79]. However, it is unknown whether, once four ATPs are bound to all loadable ClpX, the hydrolysis of ATP and dissociation of ADP occur in concert with the other molecules or independently of one another [40]. Additionally, it is not known whether monomers alternate loadable or unloadable roles. Recall that in a ClpX hexamer, only four of the ClpX monomers are loadable—capable of ATP binding.

The tight binding of the adapter protein SspB to ClpX reduces the amount of ATP required for the substrate to bind to ClpX by 30-fold, compared to just the ssrA-tagged peptide sequence by itself [60]. This cooperative binding of SspB and ClpX, ATP binding, and the rate of hydrolysis all increase cooperatively as the ATP concentration increases [80]. Bound ATP and ADP facilitates the binding of ssrA-tagged substrates to ClpX loadable monomers. In addition, ATP binding of one ClpX monomer

enhances the hydrolysis activity of its clockwise loadable neighbor — the one that donates the large subunit of the shared rigid body unit. Therefore, in the L-**L**-U-L-**L**-U orientation of the ClpX hexamer, the **L**- monomer between the L- and U- subunits is capable of the highest hydrolysis activity.

Degradation Mechanics of ClpXP

ClpX is the rate limiting step for degradation, as ClpP is able to rapidly cleave an unfolded peptide sequence (36). Previous work has shown that ClpX translocates peptide sequences highly processively, 58 amino acids at a time, with an average rate of 1800 residues min^{-1} [75]. Once translocation begins, it is a robust process. For example, at high ATP concentrations and low resistance load, translocation of a filamin domain consisting of 100 amino acids occurs in 15 steps without any slipping [75]. Each step is powered by the hydrolysis of one ATP molecule. In the case of high resistance load, such as 20 pN, ClpXP can perform a minimum of 3 kcal/mol of mechanical work per power stroke [40].

For incredibly stable substrates, unfolding is the rate-limiting step for degradation (34). Successful unfolding of stable substrates may require several attempts through power strokes fueled by ATP hydrolysis [40]. The super-stable 127 domain of titan requires 200 pN of force to denature with AFM and 500 ATP hydrolysis events to be degraded by ClpXP [81]. After a successful local unfolding event, the peptide segment is translocated, and a pause is observed where no movement from the peptide is detected. This pause can be a few to several hundred seconds (36). When the substrate resists unfolding, it is possible to for ClpXP to lose its grip, and the substrate can dissociate [81]. If this occurs prior to the degradation of the ssrA tag,

ClpXP can simply rebind and try again. These results indicate that during this pause, ClpXP hydrolyzes more ATP to unfold the next local stable structure in the substrate. The most productive unfolding occurs when a power stroke coincides with a transient thermal destabilization of the local structure, indicating that although stochastic, unfolding results from a single ATP-dependent power stroke [40]. Taken together, these actions force ClpXP to preferentially engage easily unfoldable proteins in a mixture of substrates [81].

1.6 Focus of This Thesis

Synthetic biology seeks to understand endogenous gene networks by breaking complex gene networks into simple components. There has been extensive research on the construction of new promoters, activators, and repressors; however, there has been little research on the role and importance of the degradation of circuit proteins in circuit dynamics. Degradation tags are essential for synthetic gene circuit dynamics by accelerating protein degradation and preventing the buildup of circuit proteins that could be detrimental to circuit behavior. The ability to vary the protein degradation rate allows further control of the dynamics of the circuit. The overall goal of this work is to quantify the degradation rate of five ssrA degradation tag variants, and their consequential effect on overall circuit dynamics. A deeper understanding of how degradation affects circuit dynamics will make it possible to reduce the build time on the bench, and increase the usefulness of mathematical models to guide important parameters for this building time.

Chapter 2

Modeling Synthetic Gene Oscillators

This thesis chapter is a result of the review paper:

O'Brien, E. L., Itallie, E.V., and Bennett, M.R. Modeling Synthetic Gene Oscillators. (2012) *Math Biosci.*, 236(1): 1-15.

2.1 Introduction

In order to perform the multitude of functions necessary to survive, cells must be able to regulate the expression pattern of their genes [16]. Often, this is accomplished through genetic networks – intricate webs of interactions between regulatory elements controlling protein production. The topological similarities between gene networks and electronic devices have lead many to draw analogies between the two, and hence gene networks are often referred to as "gene circuits" [24,82]. Coupled to this analogy is the belief that mathematical modeling of genetic networks will lead to new insights into their role in regulating cellular functions [83,84].

To better understand the regulatory mechanisms of gene networks, both quantitatively and qualitatively, synthetic biologists construct small-scale systems that can be studied in fine detail [85,86]. Synthetic gene circuits are generally comprised of a few interacting genes that can be placed into organisms with minimal interference

from the hosts' own regulatory processes. While naturally occurring gene networks are generally much more complicated, the simplicity of synthetic circuits provides a more manageable framework with which to test mathematical models. In addition, it is hoped that new, practical technologies will emerge from testing the limits of rationally engineered gene circuits. Indeed, the complexity of synthetic circuits and range of their functionality have greatly increased. For instance, synthetic gene circuits now exist with diverse behaviors, such as toggle switches [30], pulse counters [87], and image detectors [88, 89].

One of the most studied types of synthetic gene circuits, however, is the oscillator. These circuits are designed to produce periodic changes to the expression level of the target genes, in turn generating periodic changes in the concentration of the resulting proteins. To date there have been numerous synthetic gene oscillators reported in the literature, beginning with the "repressilator" over a decade ago [31]. Since then, circuits displaying oscillatory behavior have evolved greatly and have displayed a host of interesting properties [1, 3–5, 11–14, 32, 90].

There are several reasons why synthetic gene oscillators have been so heavily studied. First, oscillations are an important natural phenomena observed in cellular life. For instance, the circadian rhythms and the eukaryotic cell cycle are controlled by genetic networks that act as oscillators [91, 92]. Additionally, some stress response signaling pathways, such as the p53 and NF-κB pathways, can respond to stimuli with transient oscillations [93, 94]. While the oscillations do not persist, the dynamical properties of the fluctuations can determine the specific downstream response [95]. Therefore, it is important that we understand how genetic oscillations occur and how they are

regulated in order to understand cellular physiology.

The second reason, and perhaps the more important reason, that synthetic gene oscillators are studied is that they are simple circuits that nevertheless display rich dynamical behavior. This provides synthetic biologists with the opportunity to test mathematical theories of genetic regulation.

Mathematical models used for predicting gene network dynamics have varied greatly, ranging from Boolean networks [96, 97] to large-scale discrete stochastic simulations [98, 99]. The most common type of model consists of coupled nonlinear ordinary differential equations (ODEs) that use a combination of Hill functions, Michaelis-Menten enzyme kinetics [100, 101], and exponential decay to simulate the dynamics of gene products. Although these types of models have proved useful, they often use a quasi-steady-state assumption (QSSA) that ignores reactions that occur on fast time scales [102–105]. In addition, the inherent randomness associated with reactions comprised of a finite number of molecules is not included in these approximations. Stochastic simulations have therefore been used to address this randomness. These types of models are usually either Langevin-type stochastic differential equations [102] or discrete stochastic simulations, such as Gillespie's algorithm [98].

In this review, we provide an overview of mathematical models that have been used to describe the dynamics of experimentally constructed synthetic gene oscillators. We focus not on the analysis of these models, which often involves the use of nonlinear dynamics, but instead on how the models are derived from the underlying biological phenomena. We will pay particular attention to common assumptions that go into

model derivations, how models based on these assumptions can fail, and theoretical techniques that attempt to model circuit dynamics more accurately.

Often, models that use common approximation techniques, such as the QSSA, are qualitatively correct but are quantitatively inaccurate. While more complicated models are more quantitatively accurate, they generally suffer from being more difficult to simulate and analyze. Finding the right balance between model complexity and quantitative accuracy is one of the goals of synthetic biologists. Eventually it is hoped that the study of mathematical models of gene regulation will provide scientists not only the means to predict cellular behavior but also the tools with which to rationally design synthetic gene circuits for use in industrial and biomedical applications.

2.2 Transcriptional Regulation and Protein Dynamics

Before we can begin to discuss how to model synthetic gene oscillators, it is first necessary to understand the biochemical processes by which proteins are produced within the cell and how that production is regulated. Modeling the resulting dynamics of protein production and gene regulation can be done in many different ways [83]. It is therefore imperative to understand the underlying mechanisms that lead to the various assumptions that go into mathematical models. In this section, we examine how gene expression is regulated at the transcriptional level. We start by describing how promoters initiate transcription and how initiation is modulated by transcription factors. Next, we outline the basic processes that occur during protein production, including a description of the different ways protein production can be modeled. We then examine how proteins decay through cellular growth and proteolytic processes.

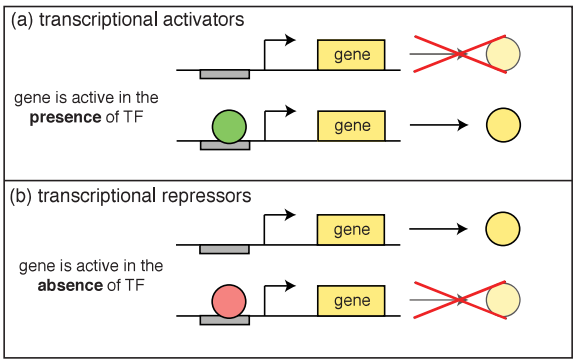


Figure 2.1 : **Regulation of promoters through transcriptional activators and repressors.** **(A)** Transcriptional activators initiate transcription of a gene once bound to the promoter. In the absence of a transcriptional activator the gene is silenced, or off. **(B)** Transcriptional repressors inhibit transcription of a gene by binding to the promoter region. The removal of the transcriptional repressor allows the promoter to become transcriptionally active.

2.2.1 Promoters and Transcription Factors

A promoter is a region of DNA just upstream of the coding region of a gene that acts as a regulator of the gene's expression. For synthetic biologists, constructing synthetic circuits starts with matching promoters with a specific type of protein called a transcription factor (TF). Transcription factors are an integral part of the transcription machinery and interpret external signals for the activation or silencing of specific genes. There are generally three types of promoters used in synthetic biology: (1) promoters that are up-regulated by transcriptional activators, as shown in Fig. 2.1(a), (2) promoters that are down-regulated by transcriptional repressors, as shown in Fig. 2.1(b), and (3) constitutive promoters which are always active.

Promoters that are regulated by transcription factors contain within them sequences of DNA, called operator sites, to which TFs can bind. Once bound to the operator, the TFs either prohibit transcription of the gene (repression) or recruit enzymes that begin the decoding process (activation).

For regulatable promoters, the activity of the gene, *i.e.* the rate at which transcription is initiated, is determined by the concentration of TF available to activate or repress the gene, and the probability that those available TFs are bound to the operator site. As an example, consider a simple promoter containing a single operator site. Further assume that the transcription factor must first dimerize before it can bind to the operator site. If the concentrations of DNA and TF are constant, then we can write the following set of chemical reactions:

$$P + P \underset{k_{-d}}{\overset{k_d}{\rightleftharpoons}} P_2 \tag{2.1}$$

$$O_u + P_2 \underset{k_{-b}}{\overset{k_b}{\rightleftharpoons}} O_b \tag{2.2}$$

where P, P_2, O_u and O_b are the concentrations of TF monomer, TF dimer, unbound operator, and bound operator, respectively, and the forward and reverse reaction rate constants are shown. If each of these reactions is in equilibrium we have

$$k_d P^2 = k_{-d} P_2 \tag{2.3}$$

$$k_b O_u P_2 = k_{-b} O_b. \tag{2.4}$$

The total concentration of DNA is constant (in this example), so that $O_u + O_b = N$. Solving for O_u gives us

$$O_u = \frac{N}{1 + (P/K)^2}, \tag{2.5}$$

where $K = \sqrt{k_{-b}k_{-d}/k_b k_d}$. Similarly, since $O_b = N - O_u$, we have

$$O_b = \frac{N\,(P/K)^2}{1 + (P/K)^2}. \tag{2.6}$$

Equations (2.5) and (2.6) are generally called Hill functions [106]. The constant K, either called the Hill constant or dissociation constant represents the concentration of TF needed for half occupancy. The exponent (here 2) is either called the Hill coefficient or the Hill exponent, and is a measure of the cooperativity of the TF. The Hill coefficient need not be 2; its value depends on, among other things, the oligomerization of the TF. For instance, with a quick derivation, one can show that the Hill coefficient for a tetrameric TF is 4. If TFs bind cooperatively to multiple binding sites, then the Hill coefficient is not restricted to being a whole number.

If the promoter fires at a rate α_u when unbound and α_b when bound, then the total rate at which initiation events occur is

$$\begin{aligned} \nu_{init}(P) &= \alpha_u O_u + \alpha_b O_b \\ &= \alpha_u \frac{N}{1 + (P/K)^2} + \alpha_b \frac{N\,(P/K)^2}{1 + (P/K)^2}. \end{aligned} \tag{2.7}$$

For repressible systems, it is generally assumed that a bound promoter does not fire, so that $\alpha_b = 0$. Likewise, for activatible promoters, it is often assumed that $\alpha_u = 0$. However, promoters can sometimes be "leaky", in that initiation can occur even when it is not favored. For instance, the initiation rate of a leaky repressible promoter might have $0 < \alpha_b \ll \alpha_u$. In this case, it is sometimes preferable to write Eq. (2.7) as

$$\nu_{init}(P) = \alpha_{leak} + \alpha_r \frac{N}{1 + (P/K)^2}, \tag{2.8}$$

where α_{leak} is the constant rate of leaky (*i.e.* unregulated) initiation, and α_r is the "regulated" initiation rate. Note that Eqs. (2.7) and (2.8) are equivalent when

$\alpha_{leak} = \alpha_b$ and $\alpha_r = \alpha_u - \alpha_b$.

In reality, the initiation rate of a promoter can be much more complicated than Eq. (2.7) suggests. For instance, many promoters contain multiple operator sites that act in tandem to regulate the promoter [10, 107]. This will change the functional form of the associated Hill function. Also, there can exist more complicated regulatory schemes involving the looping of DNA [108, 109]. In that case, DNA can be thought to have more than just two states, further complicating the initiation rate. In eukaryotic cells, TFs must enter the nucleus before they can affect the promoter. Sometimes, this shuttling in and out of the nucleus can be a complicated process [110], making the determination of the true initiation rate extremely difficult. Finally, many TFs, especially those used in synthetic biology, are dependent on small molecule ligands, often sugars, which bind to the TF and change its activity. For instance, the lactose repressor, LacI, will only bind to the operator in the absence of ligand [16], while the arabinose regulator, AraC, only activates transcription in the presence of a ligand [111]. Therefore, the true association rates of TFs to the promoters are ligand dependent, adding another layer of complexity to the process.

2.2.2 Transcription and translation

In the previous section, we were careful to state that the promoter initiates transcription of the gene. It is tempting to equate the protein production rate with the initiation rate, leading to an ODE of the form

$$\dot{x} = \alpha_u \frac{N}{1 + (P/K)^2} + \alpha_b \frac{N\,(P/K)^2}{1 + (P/K)^2} - r_{deg}(x), \tag{2.9}$$

where x is the concentration of the protein to be produced, the overdot represents differentiation with respect to time, and $r_{deg}(x)$ is the concentration dependent degra-

dation rate of the protein. Certainly, there are instances when Eq. (2.9) is a good approximation of protein dynamics. However, the creation of protein is a complicated process involving a large number of molecular complexes and reactions. Some mathematical models of synthetic gene circuits have begun to take some of these processes into consideration, and it is therefore worthwhile to examine, in some detail, the precise manner in which protein is produced within cells.

Once the transcription of a gene has been initiated, there is still a long chain of events that must be completed before the protein it encodes is active. While the following list of these reactions is not meant to be exhaustive, the basic process, outlined in Fig. 2.2, follows these basic steps [112]:

1. An enzyme called RNA polymerase binds to the DNA sequence encoding a protein and begins to create messenger RNA (mRNA). In a process called *transcription*, each base pair of the gene is read sequentially and a corresponding nucleotide is added to the growing single-stranded mRNA molecule.

2. In prokaryotic gene expression, the elongating mRNA chain can begin to create protein even before it has completely finished. In eukaryotes the process is slightly more complex. Modifications are often needed to prepare the mRNA, and the complete mRNA molecule must also be transported out of the nucleus before protein production can begin.

3. Once prepared, the mRNA must then be decoded to create protein, in a process is called *translation*. The first step in this process is the binding of ribosome to the mRNA strand.

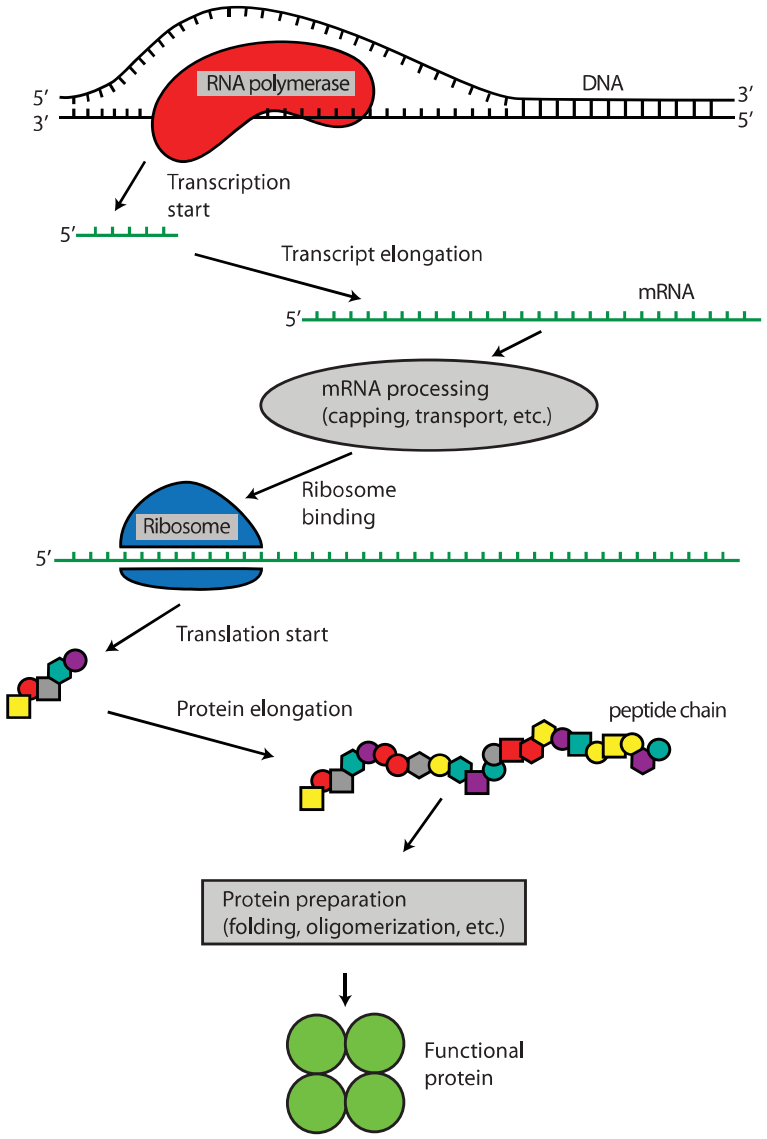


Figure 2.2 : **Schematic of steps involved in protein production.** In order for a mature protein to be produced, the gene must first be transcribed and the mRNA must then be translated. Each of these processes is comprised of hundreds or thousands of reactions that must occur as the polymerase or ribosome progress through the sequence of nucleotides. In addition, once translation has been completed, the nascent protein must then fold and, often, oligomerize.

4. Once bound to the mRNA, the ribosome reads the nucleotide sequence of the mRNA and catalyzes the reactions that build the protein one amino acid at a time.

5. Once the protein chain is complete it must fold correctly and often oligomerize before it is fully functional. Many large cellular enzymatic complexes (like the ribosome) consist of a number of different proteins that must hetero-oligomerize to function as a unit. Most transcription factors used in synthetic biology operate as either homodimers or homotetramers.

Of course, for each of the above steps there are other sub-steps. For instance, during transcription, polymerase must read from single stranded DNA. This means that the double stranded DNA must first be separated and unwound before polymerase can function. These processes in themselves are then made up of other sub-sub-steps, etc.

Due to the large number of steps required to make functional protein, Eq. (2.9) can sometimes be a poor approximation of the dynamics because it assumes that transcriptional initiation instantaneously leads to active protein. When modeling synthetic gene circuits, the first correction to Eq. (2.9) that is often made is to include the dynamics of the associated mRNA. If we assume that promoter initiation firsts leads to mRNA through transcription, and then subsequent translation of that mRNA leads to protein we are led to a two-dimensional system of the form

$$\dot{m} = \frac{\alpha_u N}{1+(P/K)^2} + \frac{\alpha_b N (P/K)^2}{1+(P/K)^2} - r_m(m) \tag{2.10}$$

$$\dot{x} = r_{tl} m - r_x(x), \tag{2.11}$$

where m is the concentration of mRNA, r_{tl} is the rate of translation, and $r_m(m)$ and

$r_x(x)$ are the degradation rates of mRNA and protein, respectively.

Systems of ODEs that include mRNA dynamics are generally more accurate than those without, especially for gene circuits that include non-trivial dynamics. Note however, that these types of ODE systems still ignore a host of reactions that occur during protein formation and transcriptional regulation. The basic assumption that is being made is that some reactions (such as oligomerization, TF-DNA binding and protein folding) occur on a much faster time scale than transcription and translation. If this is so, one can further assume that the fast reactions are in a state of local temporal equilibrium. Hence, these types of model derivations fall under the guise of the QSSA.

There do exist problems with the QSSA, however. For instance, because the QSSA is essentially a projection of the multidimensional dynamics onto a slowly evolving manifold, some aspects of the dynamics of the full system can be lost [102, 103, 105]. As an alternative, one can resort to writing differential equations for every species in the system using mass-action kinetics. Non-reduced, "complete" models that describe the dynamics of every species within a system, however, come with their own problems. First, for a given system, the complete set of biochemical species is often too large or unknown. Second, the network of associated reactions, describing which species interact with each other, is also often unknown. Third, even for known reactions, the forward and reverse kinetic rate constants are unknown or at best poorly bounded. Finally, analysis of such nonlinear systems, which can be tens or even hundreds of dimension, is extremely difficult. All of this makes large-scale unreduced models impractical, at best.

Another approach, which has recently been gaining popularity, is the use of delay differential equations (DDEs) [24, 29, 113–118]. Instead of writing down each of the reactions necessary for the creation of a protein, or just a few as in the case of the inclusion of mRNA dynamics, one can assume that each of the reactions in the pathway takes some amount of time. As Bel *et al.* showed, the completion time of linear biochemical reaction chains can be nearly deterministic and finite [119]. In that limit, then, the production of protein could be approximated by

$$\dot{x} = \frac{\alpha_u N}{1 + (P(t-\tau)/K)^2} + \frac{\alpha_b N \left(P(t-\tau)/K\right)^2}{1 + (P(t-\tau)/K)^2} - r_{deg}(x), \tag{2.12}$$

where $P(t-\tau)$ is the time-delayed concentration of the transcription factor, and τ is the delay time.

Of course, the dynamics of Eq. (2.12) are not much different from its non-delayed version, Eq. (2.9). Without feedback the dynamics of $x(t)$ are essentially unchanged, with the exception that they are delayed. However, as we will see in the case where feedback is present, delay can have dramatic consequences.

2.2.3 Protein Degradation

Thus far, we have not specified the exact form of the degradation rates of the proteins. To do this we must examine the ways in which the concentration of molecular species decline within cells. The most ubiquitous form of protein "degradation" is due to cellular growth and division, i.e. the protein concentration dilutes as the total reaction volume grows. Therefore, if we assume that cells are growing exponentially with a

rate γ, then we can write

$$r_{deg}(x) = \gamma x. \tag{2.13}$$

Cellular growth and division, however, is generally more complicated than simple exponential growth. The true growth rate of cells depends on a number of external factors. In addition, the growth of some cell types, such as *Saccharomyces cerevisiae* is biphasic at the single cell level. As *S. cerevisiae* progress through the cell cycle, the total cell volume increases at different rates. During G1 phase, cell volume grows very slowly, if at all, as the cell determines whether or not conditions are right to initiate cell division [120]. Once cell division has been initiated, the cell volume begins to increase as a new cellular bud forms and grows [121]. Furthermore, protein partitioning upon cell division can be unequal between mother and daughter, creating cellular heterogeneity [122, 123].

Often, in synthetic gene circuits, it is desirable to increase the degradation rates of the proteins in order to decrease the resulting time scales of the system. Experimentally, this has been most often accomplished with the use of degradation tags, especially the *ssrA* tagging system [124]. *ssrA* tags are C-terminal amino acid sequences that can be inserted at the end of the target gene. These short peptide tails are recognized by proteolytic enzymes within *Escherichia coli* that first denature and then degrade the tagged protein. Adding these tags to target proteins can significantly increase their degradation rates [6], thereby adding another level of control over circuit dynamics. Further, enzymatic degradation has also been shown to increase the robustness of oscillations in synthetic circuits [125]. Several synthetic gene oscillators have made use of *ssrA* tagging [4, 5, 31], and the system has also been ported into *S. cerevisiae* [126].

Because *ssrA*-mediated degradation is an enzymatic reaction, the dynamics are not exponential. Instead, they can be modeled to follow Michaelis-Menten enzyme kinetics [100], such that

$$r_{deg,enz} = \frac{V_{max}x}{K_m + x}, \tag{2.14}$$

where V_{max} and K_m are the well-known Michaelis-Menten constants.

If multiple proteins within the synthetic circuit are tagged, then each must compete for the same pool of proteases [101]. The *total* pool of tagged protein will degrade enzymatically according to Eq. (2.14), so that

$$r_{deg}(T) = \frac{V_{eff}T}{K_m + T}, \tag{2.15}$$

where V_{eff} is the effective maximal rate for the pool of proteins, $T = \sum_i x_i$ and x_i is the concentration of the i^{th} tagged protein species in the circuit. For the i^{th} protein, however, only a fraction x_i/T is being degraded by the proteases. This means that the actual degradation rate of x_i can be written as

$$r_{deg,i}(x_i, T) = \frac{V_{max,i}x_i}{K_m + \sum_j x_j} - \gamma x_i, \tag{2.16}$$

where $V_{max,i}$ is its maximum degradation rate, the binding affinity of the protease to the *ssrA* tag is identical for all protein species, and we have included the dilution term. Note that the tagging of multiple protein species has introduced degradation-mediated coupling into the dynamics. While often ignored, such coupling can affect the overall dynamics of the system [101].

We should point out that, just like *ssrA*-mediated protein degradation, transcription and translation are also enzymatic processes. It is uncommon, though, to model these

processes with Michaelis-Menten dynamics. This is because it is often assumed that the total number of *all* promoters or transcripts is much larger than that of the particular promoter or transcript being modeled. In that case, the total pool of substrate is so large that it is effectively independent of the concentration of the promoter or transcript in question. Therefore, the denominators of the Michaelis-Menten equations for transcription and translation can be assumed constant.

Like proteins, mRNA dynamics are also greatly affected by degradation. In general, mRNA are subject to similar decay mechanisms as proteins, including cellular growth, enzymatic degradation and general instability of the molecules. Synthetic biologists have engineered pathways targeting mRNA for degradation (see Ref. [127] for a good review), but these mechanisms have not been widely used to construct synthetic gene oscillators. However, there is one major exception. The mammalian oscillator constructed by Tigges *et al.* [3] uses anti-sense mRNA to target sense mRNA for degradation. We will discuss this form of regulation in more detail in Section 2.4.1.

2.2.4 Small Number Effects

So far, the equations that we have derived depend on mass-action kinetics. By this, we mean that the concentrations of each reactant are large enough that ODE models are good approximations of the true dynamics. If the number of reactants involved is small, as is often the case for gene networks, the QSSA fails for several reasons.

First, the rate of a binding event, especially homo-oligomerization, is very sensitive to the number of reactants at low concentrations. For instance, if we let P be the

number of proteins, then the rate of dimerization is actually proportional to $P \cdot (P-1)$ and not P^2. ODE models can easily be rewritten with this in mind, but care must be taken when simulating such systems in order to avoid negative estimates of oligomer concentrations. Similarly, if the copy number of a gene (and hence its promoter) is large compared to the number of available transcription factors then incorrect use of QSSA can lead to overestimation of bound operators.

The second complication that arises at low reactant numbers stems from the fact that chemical reactions are not deterministic and the relative stochasticity increases as the number of reactant molecules decreases. For genetic regulatory systems, there are often very few molecules involved, leading to significant amounts of noise in *in vivo* synthetic circuits. In addition to this "intrinsic" noise, external factors can randomly impact the dynamics of circuits [2, 128].

While we cannot go into detail into all the origins and consequences of noise in this review it should be noted that there has been a great deal of work, both theoretical and experimental, characterizing noise in genetic circuits [99, 129–132] and how circuit architecture affects noise [133, 134]. From a modeling standpoint, there are several options for dealing with noise [102]. The most common methods use some form of discrete stochastic simulation, such as Gillespie's algorithm or its variants [98,99,135,136]. However, other methods such as stochastic differential equations and master equation approaches have also been used [102, 137–139].

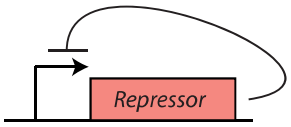


Figure 2.3 : **A genetic circuit diagram of direct negative feedback.** The promoter (arrow) regulates the expression of a gene encoding a transcriptional repressor (red box). The transcription factor acts to down-regulate its own expression (blunt-end line).

2.3 Negative Feedback Oscillators

In Section 2.2 we discussed the important processes by which protein dynamics are regulated and the various ways that they can be modeled. These biological mechanisms are present in all synthetic gene circuits, yet mathematical models do not always take each into account. To examine this issue in more detail, we now turn to oscillators that are based on negative feedback loops. The negative feedback loop is important to understand because it is essential not only to synthetic gene oscillators, but also to circadian oscillators [140]. Further, while the circuit diagram for transcriptional negative feedback often looks simple, as shown in Fig. 2.3, it can be modeled in different ways. The amount of detail that a model contains often determines the resulting dynamics.

2.3.1 Direct Negative Feedback

At its most basic, a transcriptional negative feedback loop is a system in which the transcription of a gene results in the down-regulation of that same gene (Fig. 2.3). Often this is because the gene in question encodes for a transcription factor that can bind to its own promoter and inhibit further transcription of itself. To describe this

situation, a simple ODE model based on Hill function kinetics can be written as

$$\dot{x} = \frac{\alpha}{1 + (x/K)^n} - \gamma x, \tag{2.17}$$

where x is the concentration of the repressor protein, α is the maximum production rate, K is the concentration of repressor needed for half repression, n is the Hill coefficient and γ is the degradation rate coefficient.

Of course, Eq. (2.17) cannot oscillate because there is only one dimension. In fact, there is just a single stable fixed point given by the positive real solution to

$$\gamma x = \frac{\alpha}{1 + (x/K)^n}. \tag{2.18}$$

Why then, does negative feedback form the basis for oscillations in synthetic gene circuits?

The answer comes when we begin to expand away from the simplistic, one dimensional QSSA model for negative feedback. If we also include the dynamics for the mRNA in the system, we arrive at

$$\dot{m} = \frac{\alpha}{1 + (x/K)^n} - \delta m \tag{2.19}$$

$$\dot{x} = \beta m - \gamma x, \tag{2.20}$$

where m is the concentration of mRNA, β is the translation rate and the degradation rate coefficients for mRNA and protein are δ and γ, respectively.

Ostensibly, the differences between Eq. (2.17) and the system (2.19)-(2.20) are not great. The former is just one dimensional, while the latter is two dimensional. Two dimensions is enough to support oscillations, of course, but system (2.19)-(2.20) does

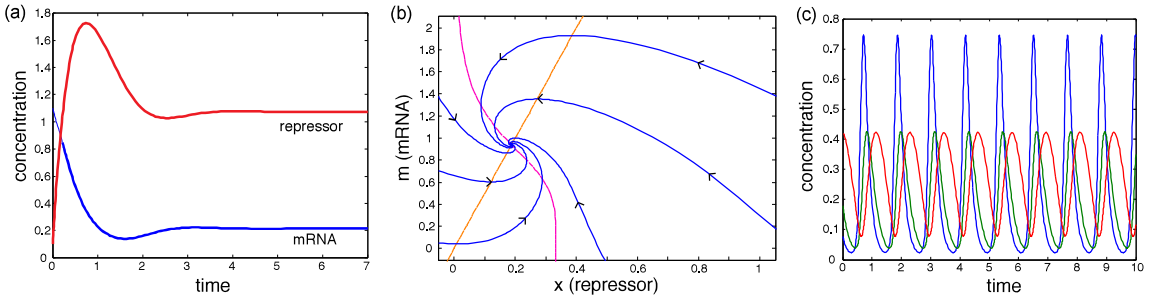


Figure 2.4 : **Behavior of a negative feedback loop. (A)** Here the mRNA dynamics are explicitly modeled, as in Eqs. (2.19) and (2.20). The protein (red curve) and mRNA (blue curve) concentrations settle onto a fixed point after a transient. **(B)** The trajectories of Eqs. (2.19) and (2.20) in m-x space. The imaginary components of the eigenvalues of the stable fixed point create damped oscillations as the trajectories spiral into the fixed point. Here we show 6 representative trajectories in blue. The orange and pink curves are the null-clines of Eqs. (2.19) and (2.20), respectively. **(C)** When a third dimension is added, as in Eqs. (2.21)-(2.23), stable oscillations are achieved. Here, the blue, green and red curves are the values of x, y and z, respectively, as a function of time.

not exhibit oscillations. Like its one dimensional analog, the system that includes mRNA has a single stable fixed point. However, the fixed point can have oscillatory-like behavior.

Figure 2.4 shows a trajectory of system (2.19)-(2.20). Note that while the fixed point is attracting, the eigenvalues have imaginary components (the real parts are both negative). The imaginary components of the eigenvalues produce the damped oscillations of the repressor concentrations, and provide a tantalizing clue as to how true oscillations are formed. The repressor "overshoots" its steady state position whereupon it begins to relax onto the fixed point.

By adding more intermediate species in the linear pathway that leads from the ini-

tiation of transcription to the down-regulation of the promoter, a "diffusive delay" is created that can lead to oscillations [26]. Even with just three dimension, oscillations are relatively easy to achieve, especially if the protein decays enzymatically. For instance, consider the following set of equation:

$$\dot{x} = \frac{\alpha}{1+(z/K)^n} - \gamma_x x \tag{2.21}$$

$$\dot{y} = \beta_y x - \gamma_y y \tag{2.22}$$

$$\dot{z} = \beta_z y - \frac{Vz}{K_m + z}, \tag{2.23}$$

where x, y and z can be thought of as the concentration of mRNA, an intermediate species, and the repressor, respectively. This set of equations produces stable limit cycle oscillations, as shown in Fig. 2.4(c). What the intermediate species is we have not specified, and there are several possibilities that present themselves. In a rough sense, y could be another transcription factor that up-regulates the production of z. Of course, the functional form of the regulation would most likely look different than above, but the principle still holds. As we will see later, this form of negative feedback loop, consisting of more than one gene, has been well studied.

Another plausible explanation is that y is some intermediate, non-active version of the mature repressor, z. Perhaps the repressor needs to be activated by an enzyme, or unfolded and immature versions of the protein can be thought of as another species altogether. This line of reasoning quickly leads to the concept of delayed negative feedback. As discussed in Section 2.2.2, delay in protein production can be thought of as the sequential assembly of the mRNA and mature protein. In fact, delayed negative feedback oscillators have been well studied and are thought to form the basis of some naturally occurring cellular oscillators, such as circadian oscilla-

tors [24, 29, 116, 117, 141–143].

2.3.2 The Delayed Negative Feedback Oscillator

Single gene negative feedback loops can, under certain circumstances, exhibit sustained oscillations. This fact was first observed experimentally by Stricker *et al.* [5] in a circuit they constructed using LacI to repress itself, as shown in Fig. 2.5. Previous theoretical work had predicted that such oscillations might exist in single gene circuits [26], or that stochasticity might induce oscillations when delay is present [29].

Essentially, one can model the delayed negative feedback oscillator by using an extension of Eqs. (2.21)-(2.23):

$$\dot{x}_0 = \frac{\alpha}{1 + (x_N/C)^\beta} - \gamma x_0 \tag{2.24}$$

$$\dot{x}_i = P_i(x_{i-1}) - P_{i+1}(x_i) - \gamma x_i \tag{2.25}$$

$$\dot{x}_N = P_N(x_{N-1}) - \frac{V x_N}{K + x_N} - \gamma x_N, \tag{2.26}$$

where $i \in \{1, N-1\}$ and there are N intermediate species. The production of the repressor is still initiated by the repressive Hill function term, but instead of creating fully formed protein the result is the initial intermediate, x_0. Subsequent intermediate species, x_i for $i = 1...(N-1)$, are then created according to some production function, $P_i(x_{i-1})$, roughly irreversibly, by the previous species in the linear chain. Eventually the fully formed repressor, x_N, emerges and is able to down-regulate the initiation process. Note that in this scheme we have allowed only the fully functional repressor to enzymatically decay, but this need not be the case.

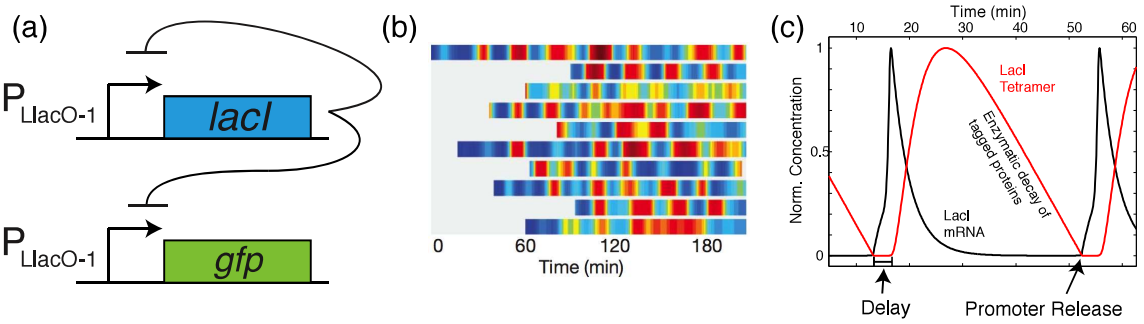


Figure 2.5 : **The delayed negative feedback oscillator.** Stricker *et al.* predicted computationally, and proved experimentally, that a circuit containing a single gene repressing itself was sufficient to produce oscillations [5]. **(A)** Circuit diagram of the delayed negative feedback oscillator. There are two copies of the LlacO-1 promoter [10], one driving *lacI* and the other driving the gene encoding the green flourescent protein (*gfp*). In the absence of IPTG, the product of the *lacI* gene, LacI, down-regulates the promoters. In addition, both genes were *ssrA*-tagged in order to increase the degradation rates of the resulting proteins [6]. **(B)** The level of flourescent protein in single cells as a function of time (red indicates high fluorescence while blue indicates low fluorescence). Each horizontal time series represents the trajectory of a single cell. **(C)** The origins of delayed negative feedback oscillations are apparent in this simulation. Here there is a small delay between the time that the mRNA begin to be formed and the time at which functional LacI tetramer are produced. Once enough LacI has been made, transcription shuts down and the mRNA levels begin to drop. The resulting burst in repressor eventually decays through proteolysis. Once the LacI concentration falls below the threshold level, another burst of mRNA occurs, starting the process anew. Figure adapted from Stricker *et al.* (2008) [5].

From a stochastic viewpoint, the time it takes from initiation to fully formed protein is determined by the number and nature of the steps in the pathway. If each reaction step is irreversible, and exponentially distributed with rate λ, then the mean delay time is given by $\tau = N/\lambda$. Further, as the number of reactions increases while keeping τ fixed, the variability in the delay time decreases such that it can be considered to be deterministic [119].

With the above in mind, we can now rewrite the system of equation, Eqs. (2.24)-(2.26), as [116]

$$\dot{x} = \frac{\alpha'}{1 + (x_\tau/K)^\beta} - \frac{Vx}{K_m + x} - \gamma x, \tag{2.27}$$

where $x_\tau = x(t-\tau)$ and α' is the effective maximum production rate. This correction to α is necessary due to the possible degradation and dilution of intermediate species in the reaction pathway. To see this, assume that the degradation of the intermediate species is only due to dilution, with rate γ. Then the probability that any one protein makes it all of the way through the reaction chain to become a functional repressor is simply given by $\exp(-\gamma\tau)$. In that case, we can rewrite Eq. (2.27) as

$$\dot{x} = \frac{e^{-\gamma\tau}\alpha}{1 + (x_\tau/K)^\beta} - \frac{Vx}{K_m + x} - \gamma x, \tag{2.28}$$

so that $\alpha' = \alpha \exp(-\gamma\tau)$. Note that if there is active degradation, either acting on the mRNA or nascent protein, the correction to α will be more complicated.

Inspired by the experimental evidence for delayed negative feedback oscillations, Mather *et al.* theoretically investigated a system similar to Eq. (2.27) [116]. They found that the period of oscillations observed in such systems can be arbitrarily long compared to the delay time. In addition, they showed that tight repressor binding

and strong enzymatic decay of the protein were essential to obtaining oscillations.

2.3.3 Multi-gene Negative Feedback Circuits

As mentioned above, one way of looking at the three equations that make up the "diffusive delay" system, Eqs. (2.21)-(2.23) is to consider them a multi-gene network, with the intermediate species, y, being a transcriptional activator that up-regulates the repressor, z. If we allow our negative feedback circuit to contain more than one gene, then there are any number of ways in which to construct it (at least theoretically). Figure 2.6 shows the complete sets of topologically distinct negative feedback circuits consisting of up to five genes. The general rule of thumb is that for circuits with a ring architecture there must be an odd number of repressors in order for the entire loop to be a negative feedback circuit.

Of course, the production of each protein in the circuit will not be linearly dependent on another, as in Eqs. (2.21)-(2.23), but instead be produced according to some Hill function. In that case, a simple ODE model of a two gene negative feedback loop, as shown in Fig. 2.6, could be written as:

$$\dot{m}_a = \alpha_a \frac{K_r^{n_r}}{K_r^{n_r} + r^{n_r}} - \delta_a m_a \tag{2.29}$$

$$\dot{a} = \beta_a m_a - \gamma_a a \tag{2.30}$$

$$\dot{m}_r = \alpha_r \frac{a^{n_a}}{K_a^{n_a} + a^{n_a}} - \delta_r m_r \tag{2.31}$$

$$\dot{r} = \beta_r m_r - \gamma_r r, \tag{2.32}$$

where r and a are the concentrations of the repressor and activator, respectively, and m_r and m_a are their respective concentrations of mRNA; α_i, K_i and n_i for $i \in \{a, r\}$

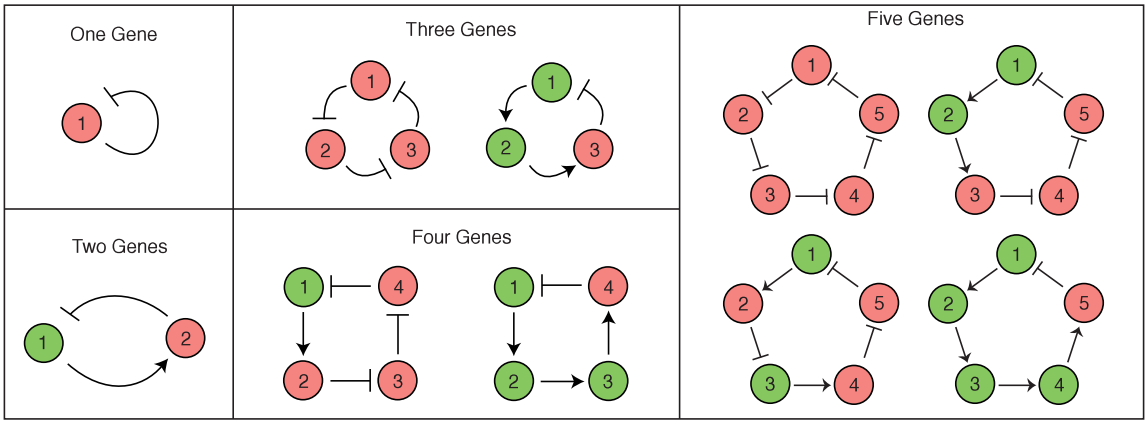


Figure 2.6 : **Negative feedback loops can be made up of any number of genes.** Shown are the topologically distinct negative cyclic feedback loops containing up to 5 transcription factors. Red and green circles indicate transcriptional repressors and activators, respectively. The blunt end lines and arrows represent repression and activation, respectively.

are the respective constants describing the Hill functions; and δ_i and γ_i are the mRNA and protein degradation rate coefficients.

While there are many types of multi-gene negative feedback circuits possible, the repressilator is by far the most well-studied [31].

2.3.4 The Repressilator

The first reported experimental realization of a synthetic gene oscillator is the well-known "repressilator" [31]. This circuit contains three genes, each of which is a transcriptional repressor, as shown in Fig. 2.7(a). The repressilator was based on a form of digital circuit, the ring oscillator. In electronics, a ring oscillator is an odd numbered set of NOT gates that are linked in a circle. These types of electronic cir-

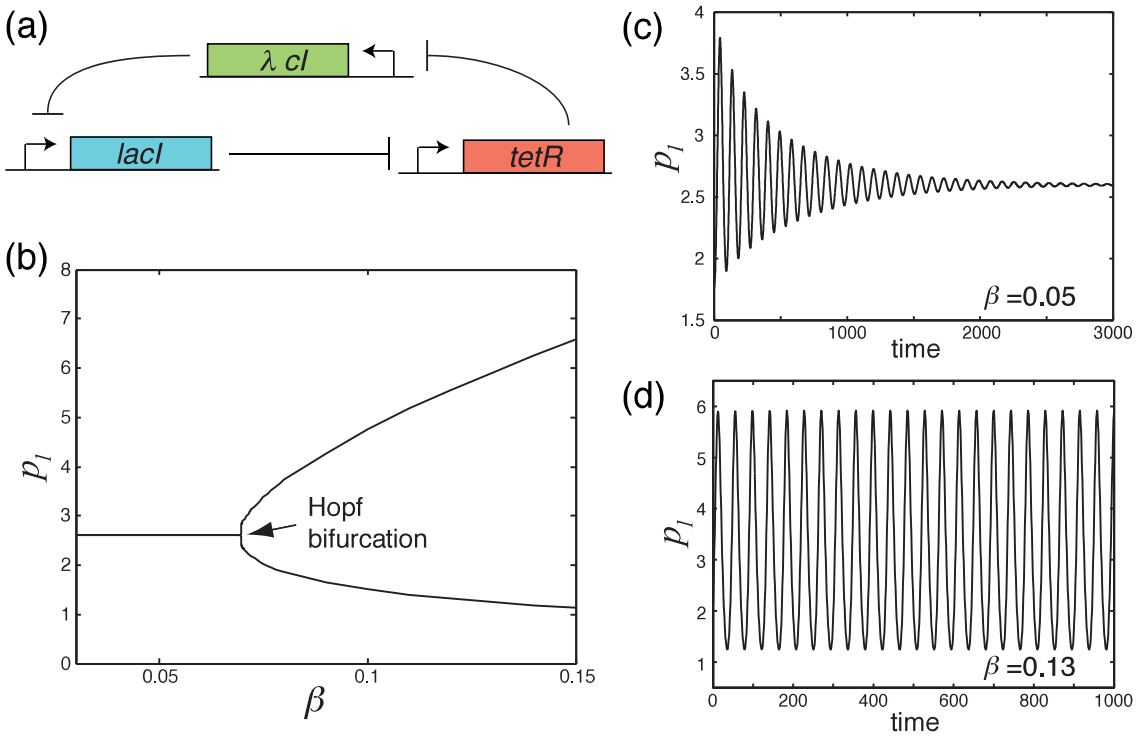


Figure 2.7 : **The model for the repressilator circuit oscillates in a specific range of parameters.** **(A)** The circuit consists of three genes each of which represses a specific promoter that controls expression of one of the other genes. **(B)** A Hopf bifurcation occurs at approximately $\beta = 0.07$. **(C)** For low values of β, *i.e.* to the left of the bifurcation point, the system undergoes damped oscillations. **(D)** The repressilator exhibits stable limit cycle oscillations when β is large.

cuits oscillate, with a period that is inversely proportional to the circuit delay (which is determined by the number of NOT gates).

To build the genetic version of the ring oscillator, Elowitz and Leibler used three genes encoding transcriptional repressors: the lactose repressor, LacI; the tetracycline repressor, TetR; and the CI repressor from bacteriophage λ (see Fig. 2.7(a)). Next, they placed these genes under the control of promoters that are regulated by one of the three transcription factors, such that each represses the transcription of the next

gene in the circle. Therefore, using Hill functions one can write the system as

$$\dot{x}_1 = \alpha_{0,1} + \frac{\alpha_1}{1 + (x_3/K_3)^{n_3}} - \gamma_1 x_1 \tag{2.33}$$

$$\dot{x}_2 = \alpha_{0,2} + \frac{\alpha_2}{1 + (x_1/K_1)^{n_1}} - \gamma_2 x_2 \tag{2.34}$$

$$\dot{x}_3 = \alpha_{0,3} + \frac{\alpha_3}{1 + (x_2/K_2)^{n_2}} - \gamma_3 x_3, \tag{2.35}$$

where x_1, x_2 and x_3 are the concentrations of LacI, TetR and cI, respectively. Note that we have included the leakiness of each promoter, represented by the constitutive production rates, $a_{0,i}$.

In their original model, Elowitz and Leibler included the mRNA dynamics for each protein. Also, by assuming that various constants are the same for each protein, i.e. $\alpha_{0,i} = \alpha_0$, $K_i = K$, $n_i = n$ and $\gamma_i = \gamma$, one can rescale the concentrations and time to arrive at the compact, dimensionless form of the equations:

$$\dot{m}_i = -m_i + \frac{\alpha}{1 + p_j^n} + \alpha_0 \tag{2.36}$$

$$\dot{p}_i = -\beta\,(p_i - m_i)\,, \tag{2.37}$$

where $i \in \{1, 2, 3\}$ represents a gene in the circuit and j represents its predecessor. The number of protein copies per cell produced from a given promoter type during continuous growth is α_0 in the presence of saturating amount of repressor (an effect of a leaky promoter), and $\alpha + \alpha_0$ in its absence; β denotes the ratio of the protein decay rate to the mRNA decay rate; and n is the Hill coefficient. Time is rescaled in units of mRNA lifetime; protein concentrations are written in units of K, the number of repressors necessary to half-maximally repress a promoter; and mRNA concentrations are rescaled by their translation efficiency, the average number of proteins produced per mRNA molecule [31].

By performing bifurcation analysis on Eqs. (2.36) and (2.37), one can determine where in parameter space the system should oscillate. For instance, Fig. 2.7(b) shows that a Hopf bifurcation occurs at approximately $\beta = 0.07$. When β is less than 0.07 oscillatory behavior is transient, whereas the the system oscillates for values of β larger than 0.07, as show in Figs. 2.7(c) and 2.7(d), respectively.

Since it was first reported, theorists have examined many models for the repressilator [105, 114, 144, 145]. For instance, several groups have examined ODE models with n repressing elements and showed that not only can periodic solutions exist, but also multiple steady states, and aperiodic heteroclinic cycles [146, 147]. Others have looked at how stochasticity [148, 149] or the copy number of each gene affects oscillations [150].

2.4 The Role of Positive Feedback

As we have seen above, a transcriptional negative feedback loop is sufficient to generate oscillations in protein concentrations in and of itself. Circuits relying solely on negative feedback generally work on the principle of delay, whether it is diffusively through a transcriptional cascade or due to the time lag necessary for protein production. Another form of synthetic gene oscillator, based on the principle of linking both negative and positive feedback to create relaxation oscillations [151], has also been widely studied. While they are harder to construct experimentally, it is predicted that genetic oscillators containing both positive and negative feedback should be more robust and more tunable than oscillators containing only negative feedback [152, 153].

2.4.1 Experimental Realizations of Linked Positive and Negative Feedback

Several synthetic gene circuits that exhibit oscillatory behavior have been built on the principle of linked positive and negative feedback. The first was by Atkinson *et al.*, in which they used a simplified model of gene regulation to predict oscillations [11]. Their mathematical model used tri-phasic (i.e. piece-wise linear) functions instead of Hill functions to represent transcriptional control and predicted that the circuit shown in Fig. 2.8(a) would oscillate. Indeed, when they constructed the circuit in *E. coli* they did observe long-period oscillations, though they damped out after several periods.

The use of a tri-phasic function might seem odd, but it is a decent approximation of a Hill function. The idea is that the Hill function can be thought of as containing three distinct regions: basal production, maximum production, and an intermediate regime. If we were to construct a tri-phasic function for the activation of a gene by an activator it might take the form (see [11]):

$$f(x) = \begin{cases} B & \text{if } x < B \\ x & \text{if } B < x < M \\ M & \text{if } x > M, \end{cases} \tag{2.38}$$

where x is a dimensionless concentration of activator, and B and M are dimensionless rates of basal and maximal production, respectively. If we compare this to a dimensionless Hill function of the form

$$f(x) = B + (M - B)\,\frac{x^n}{K^n + x^n}, \tag{2.39}$$

where $K = (B + M)/2$, we see that the fit, while not perfect, can be reasonably close,

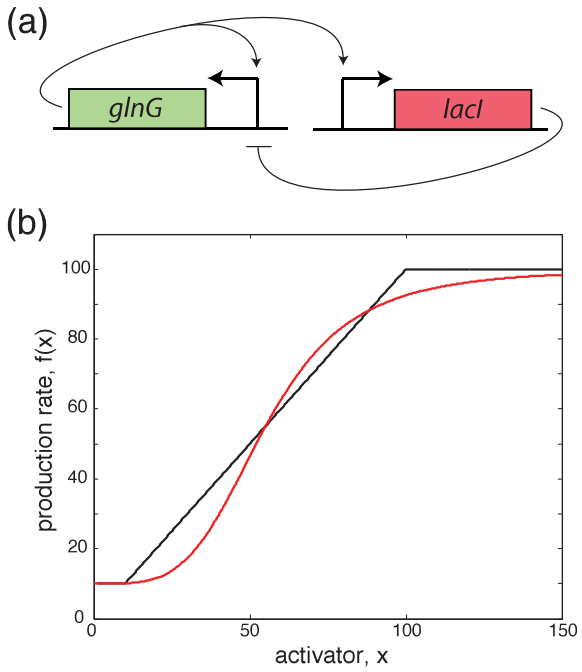


Figure 2.8 : **The synthetic oscillator construction by Atkinson *et al.*** [11]. **(A)** Circuit schematic of the circuit. The activator (encoded by the *glnG* gene) up-regulates both proteins, while the repressor (encoded by the *lacI* gene only down-regulates the activator. **(B)** The tri-phasic regulation functions used by Atkinson *et al.* are decent approximations of Hill functions. Here, an activating tri-phasic function, Eq. (2.38), is shown in black, while its Hill function counterpart, Eq. (2.39), is shown in red. Here, $B = 10$, $M = 100$ and $n = 4$.

as shown in Fig. 2.8(b).

The piece-wise linear Hill function, Eq. (2.38), while an approximation of a true Hill function, has the distinct advantage that it is much easier to analyze than its non-linear counterpart. Further, it has been shown that models of gene oscillators using Heaviside step-functions instead of Hill functions still give good quantitative agreement [118]. Therefore, piece-wise linear Hill functions are an attractive option for modeling gene networks, especially when analytical solutions are desired.

Two other synthetic gene oscillators incorporating positive and negative feedback are of importance here. Each uses standard QSSA models with Hill functions, but each also includes more than just transcriptional regulation. In the first, Fung *et al.* created a network, termed the "Metabolator," that couples transcriptional regulation to metabolism [4]. Essentially, two metabolite pools are cyclicly interconverted into one another through the action of two enzymes. One of the metabolites up-regulates the production of two proteins (one that encourages and one that inhibits the cyclic interconversion of the two metabolites) that provide the feedback. Therefore, it was necessary for the authors to use a mathematical model that not only includes transcriptional dynamics but also metabolite dynamics. The result is a system of ODEs that combines Hill function dynamics for the proteins and Michaelis-Menten and mass-action kinetics for the metabolites.

The other network, constructed by Tigges *et al.*, was created in mammalian cells [3]. This network made use of regulation at the mRNA level by creating both sense and anti-sense mRNA of a coding region, as shown in Fig. 2.9. Anti-sense mRNA (*i.e.*

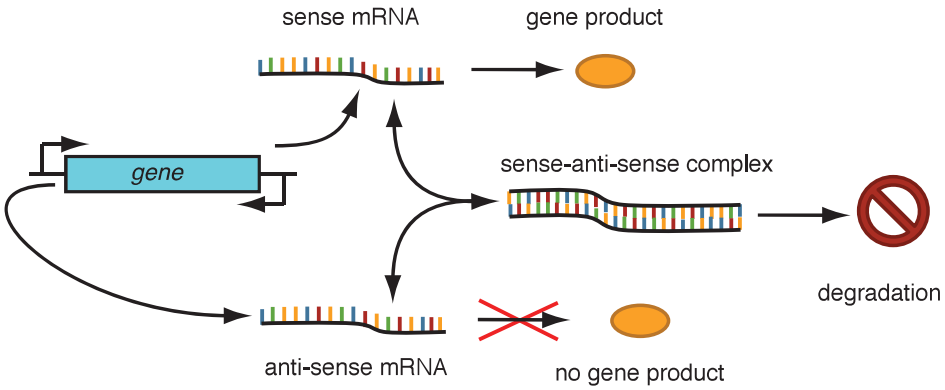


Figure 2.9 : **Sense and anti-sense mRNA interactions from the mammalian synthetic gene oscillator** [3]. A gene is flanked on either side by a promoter. The "sense" promoter initiates the transcription (reading left to right in the above figure) of sense mRNA, which is able to generate functional protein encoded by the gene. The anti-sense promoter initiates backwards transcription of the gene (here right to left), producing anti-sense mRNA. The anti-sense mRNA is unable to produce protein. The sense and anti-sense mRNA can hetero-dimerize into a degradable complex.

the reverse compliment of the sense mRNA that cannot be translated) can bind to the sense mRNA. The dimer cannot be translated and degrades faster than either strand does alone. Because the sense-anti-sense dimer provides a new sink for the sense mRNA, the creation of the anti-sense mRNA is a form of negative regulation of the protein that is translated from the sense mRNA.

While the mammalian oscillator contains transcriptional regulation, the presence of regulation at the mRNA level increases the complexity of the network and the subsequent modeling. One can extend the Hill-function based ODE models of transcriptional regulation to include terms describing the kinetics of the mRNA species and their interactions. Specifically, if one uses standard mass-action kinetics the resulting

kinetics of the mRNA species can be modeled as:

$$\dot{m}_s = k_s - k_f m_s m_a + k_r m_d - \gamma_s m_s \quad (2.40)$$

$$\dot{m}_a = k_a - k_f m_s m_a + k_r m_d - \gamma_a m_a \quad (2.41)$$

$$\dot{m}_d = k_f m_s m_a - k_r m_s - \gamma_d m_d, \quad (2.42)$$

where m_s and m_a are the concentrations of the sense and anti-sense mRNA, respectively. The sense and anti-sense mRNA are created at some rate, either k_s or k_a, which are functions of the associated transcription factor concentrations (not modeled here). In addition, the two forms of mRNA can dimerize into an sense-anti-sense dimer (m_d) with a forward rate k_f and a reverse rate k_r. All three forms degrade, but the dimer will generally degrade more quickly.

As predicted for oscillators containing both positive and negative feedback, the oscillator constructed by Tigges *et al.* exhibited tunable oscillations. The same authors also used a similar mechanism to create a synthetic low-frequency oscillator [32].

2.4.2 Large-scale Modeling of a Dual Feedback Oscillator

There exists another form of the dual feedback oscillator that was created by Stricker *et al.* [5]. While topologically it is not very different from the dual feedback oscillators described above, it deserves mention for several reasons. First, it is an extremely robust oscillator – meaning that nearly every cell in the population oscillates and the oscillations do not damp out. Second, the oscillatory period was tunable by controlling the extracellular concentrations of small molecule ligands. Finally, the modeling of this oscillator goes beyond standard Hill-function ODEs, and in doing so highlights

many of the problems that arise when using QSSA-based Hill function techniques.

The basis for the Stricker oscillator came from Hasty *et al.* [34], who proposed a two-gene circuit consisting of a transcriptional activator that up-regulates both genes (to provide positive feedback) and a repressor that represses both, as shown in Fig. 2.10(a). In this case, and provided the two promoters are identical, the ODEs might take the form

$$\dot{a} = p_a \frac{1+\alpha a^2}{(1+a^2)(1+r^4)} - \gamma_a a \tag{2.43}$$

$$\dot{r} = p_r \frac{1+\alpha a^2}{(1+a^2)(1+r^4)} - \gamma_r r, \tag{2.44}$$

where a and r are the (rescaled) concentrations of activator and repressor, respectively; α is the fold-change of activated transcription; γ_a and γ_r are the degradation rates of the activator and repressor, respectively; and p_a and p_r are the basal transcription rates.

In order for Eqs. (2.43) and (2.44) to make sense, the promoter driving both genes must be regulated by both the activator and the repressor. This can be achieved, experimentally, by placing the regulatory elements (*i.e.* the DNA sequences that correspond to the binding sites) of each transcription factor near the promoter [10]. The derivation of equations like (2.43) and (2.44) is similar to that for promoters with a single transcription factor input explained in Section 2.2.1. Instead of having just two states for the promoter, however, all combinatorial possibilities of DNA-transcription factor complexes must be taken into account. For instance, if there is one binding site for the activator and one for the repressor then number of promoter states is 4: unbound; bound by the activator only; bound by the repressor only; and bound by

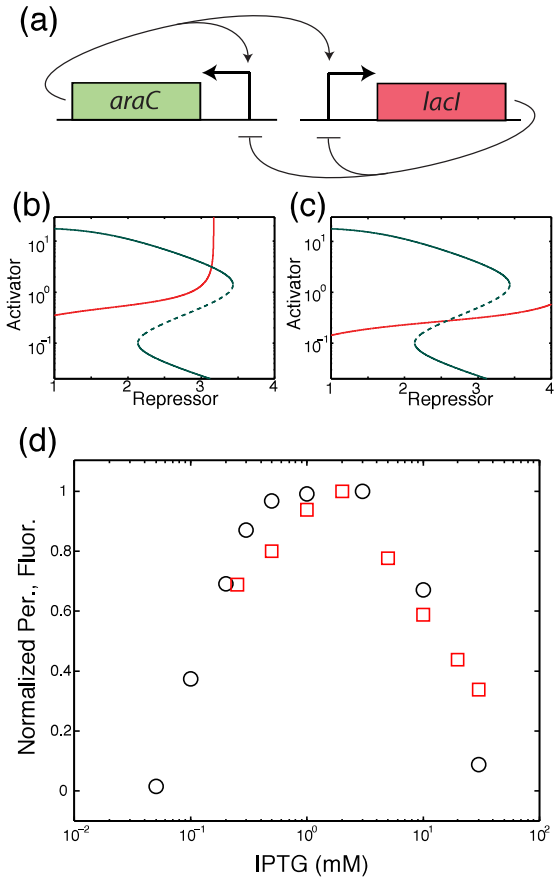


Figure 2.10 : **The dual-feedback oscillator constructed by Stricker *et al.*** [5]. **(A)** Circuit diagram of the oscillator. Negative feedback is provided by LacI while positive feedback is provided by AraC. Both genes are regulated by a hybrid promoter that responds to both transcription factors. **(B)** Null-clines for the model system, Eqs. (2.43) and (2.44). Here, the green line is the nullcline for the activator equation and the red line is the nullcline for the repressor. A solid line indicates stability while a dashed line indicates instability. When the degradation rates of the two proteins are the same the intersection of the two nullclines produces a stable fixed point. **(C)** However, if the degradation rate of the activator is 5 times greater than that of the repressor, the two nullclines intersect at an unstable fixed point – resulting in limit cycle oscillations. **(D)** Experimentally measured normalized period of the oscillator (red squares) and normalized induction strength of the promoter (in the absence of feedback, black circles) as a function of IPTG. Interestingly, the period of the oscillator appears to be proportional to the strength of the promoter at a given concentration of IPTG. Figure adapted from Stricker *et al.* (2008) [5]..

both activator and repressor. Each state of the promoter consequently has its own rate of transcriptional initiation.

Figures 2.10(b) and 2.10(c) show nullclines for Eqs. (2.43) and (2.44) in two different parameter regimes. The intersection of the two curves determines the fixed point of the whole system. When the degradation rates of the two proteins is the same (Fig. 2.10(b)), the intersection indicates a stable fixed point. However, if the degradation rate of the activator is roughly 5 times greater than that of the repressor (Fig. 2.10(c)), the intersection of the two nullclines occurs in the region of the unstable fixed point in the activator's nullcline. In this case, the result is a stable limit cycle.

Experimentally, this synthetic gene oscillator consisted of two genes: the repressor *lacI* and the activator *araC*. Each gene was driven by a synthetic hybrid $P_{\mathrm{lac/ara-1}}$ promoter, which is simultaneously repressed by LacI and activated by AraC [10]. In addition, the authors used *ssrA* tags to increase the degradation rates of the proteins [6]. The circuit diagram for this network is shown in Fig. 2.10(a).

While the ODE model predicts relaxation oscillations, as described above, it was found that the oscillations predicted by the QSSA model had little in common with the oscillations actually observed in the experiments. In particular, the tunability of the oscillator as a function of the ligand concentrations and the period of the oscillations were difficult to achieve in the the simple model when realistic parameters were used. For instance, to achieve the fast periods (as low as ~ 15 mins), unrealistically fast protein degradation rates were required with half-lives on the order of a few minutes. This is faster than any measured degradation rate mediated by *ssrA* tags [6].

One clue as to the origin of the discrepancy was found in the correspondence between the period of the oscillator and the strength of the promoter (in the absence of feedback). As shown in Fig. 2.10(d), the period of the oscillator is roughly proportional to the promoter strength, as a function of the concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG, the ligand for LacI). One possible explanation for this behavior is the presence of dynamical delay in the system. As the promoters begin to transcribe, they do so at a rate proportional to the induction level. Therefore, before the repressor shuts off production, the amount of protein produced is also proportional to the induction level. Then, enzymatic decay reduces the levels of both transcription factors until the repressor reaches a low enough level that the promoters can begin transcription again. If the time it takes for the burst of transcription to happen is small compared to the decay time, then the total period of the oscillator is controlled by the roughly zeroth-order enzymatic degradation of the protein. Hence, the period will be proportional to the induction strength of the promoters [116].

Instead of introducing delay directly into their QSSA model, however, the authors decided to expand the QSSA model by incorporating all known (or supposedly known) reactions in the system, without reducing the model onto a slow manifold. The resulting system of equations (which is too large to reproduce here) is considerably more vast than the original QSSA model – going from 2 dimensions to a total of 27. By expanding the original 2 dimensional QSSA model, the authors pay the price of increasing the number of unknown parameters. Still, the expanded model does a much better job predicting the experimentally observed behavior of their oscillator when compared to the original QSSA model [5].

If delay were the cause of the oscillations, then the network might not need the positive feedback loop in order to oscillate. By manipulating their large-scale ODE model to disable the positive feedback loop, they did indeed obtain fast, low amplitude oscillations. Informed by this computational prediction, the authors then went back to their experimental system to create a related strain with only negative feedback. Just as they had predicted, the strain did exhibit low amplitude oscillations. This is the origin of the single gene delayed negative feedback oscillator discussed is Section 2.3.2. Further, the experimental findings of Stricker *et al.* showed that the system containing just negative feedback was much less tunable than its dual-feedback counterpart. This is in agreement with theoretical predictions made by Tsai *et al.* [152]. So, at least in this case, the addition of positive feedback to a negative feedback loop produces oscillations that are more robust and tunable than those obtained from negative feedback alone.

2.5 Synthetic Multi-cellular Oscillators

The natural next step in the development of synthetic gene oscillators is to engineer into them cellular signaling mechanisms that create population-wide oscillatory dynamics. The most common method for coupling gene expression among cells in a population is to use quorum sensing systems from bacteria [154]. These systems use an enzyme to metabolize S-adenosyl-methionine (SAM) into an N-acyl-homoserine lactone (AHL), which can diffuse into and out of cells within the population and act as a ligand for transcription factors. Quorum sensing systems have been used successfully in a number of applications, including synthetic gene circuits that are capable

of cellular population control [155], spatial pattern formation [156], cellular photography [88], programmed invasion of cancer cells [157], image edge detection [89], and multicellular logic gating [158].

2.5.1 Synthetic Population Control Circuits

The first experimental realization of a synthetic circuit containing intercellular communication that oscillates was reported by Balagaddé *et al.* [12]. To do this, the authors used a combination of an existing synthetic circuit that was designed to control population size [155] and a microfluidic device [159, 160] designed to act as a micro-chemostat for microbial populations [12]. A schematic of the regulatory process is shown in Fig. 2.11(a).

The cells in the population create a signaling molecule (AHL) that induces a cell death pathway within each cell. As the population size increases, the concentration of AHL also increases. The rise in AHL subsequently activates LuxR, which then up-regulates the enzyme that causes cell death. Finally, the cell death enzyme down-regulates the population size by killing cells. The topology of the whole system, which can be thought of as a diffusive delay negative feedback loop, is depicted in Fig. 2.11(b).

The model used by Balagaddé *et al.* introduces us to two new aspects of modeling synthetic gene circuits: cellular signaling and population dynamics. The growth dynamics of the population can be written as

$$\dot{N} = kN\left(1 - \frac{N}{N_m}\right) - dEN - DN, \tag{2.45}$$

where N is the number of cells in the chemostat. The first term on the r.h.s. of Eq.

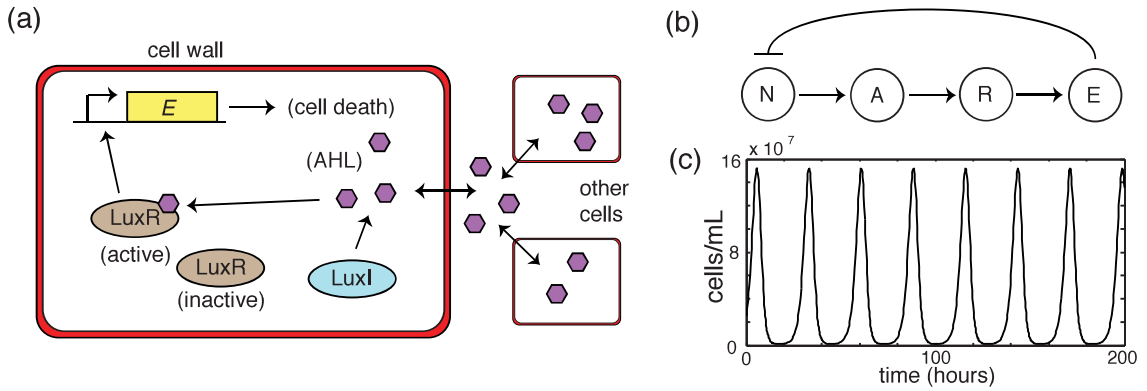


Figure 2.11 : **The population control oscillator** [12]. **(A)** Circuit schematic for the population control circuit. The proteins LuxR and LuxI are constitutively produced. LuxI creates AHL, which diffuses into and out of the cell. Cellular AHL activates LuxR, which is then able to up-regulate the enzyme that causes cell death. **(B)** Circuit topology of the model used to describe the strain in a microchemostat. Note that it constitutes a negative feedback loop with diffusive delay. **(C)** The simple model will oscillate under the right conditions. Shown is the cell density as a function of time, obtained by integrating Eqs. (2.45)-(2.48).

(2.45) represents logarithmic growth of the population, with maximum rate k and a maximum population size N_m. The second term is the death rate of the cells, which is proportional to a constant, d, and the amount of the enzyme, E, responsible for fatality. Finally, the third term represents the efflux of the cells out of the chemostat at a dilution rate D.

The enzyme that creates the AHL, LuxI, and the transcriptional activator that responds to AHL are both produced constitutively within each cell and hence their cellular concentrations can be considered constant. Therefore, the production of the AHL in the whole culture is proportional to the number of cells, and the amount of

activated LuxR within each cell is proportional to the amount of AHL. This gives us

$$\dot{A} = \nu_A N - (d_A + D)\, A \tag{2.46}$$

$$\dot{R} = k_R A - d_R R, \tag{2.47}$$

where A is the concentration AHL in the chemostat and R is the concentration of activated LuxR in each cell. Finally, the enzyme that causes cell death is up-regulated by activated LuxR, and therefore its dynamics can be approximated by

$$\dot{E} = k_E R - d_E E, \tag{2.48}$$

where k_E is the coefficient for production, and d_E is the degradation rate. If the dilution rate of the chemostat, D, is tuned properly, the systems exhibits distinct oscillations (Fig. 2.11(c)), which were confirmed experimentally.

Interestingly, Balagaddé *et al.* used the same microchemostat and similar gene circuits to create a synthetic ecosystem containing two distinct strains – a "predator" strain and a "prey" strain [161]. Modeled after the classic predator-prey system [162], the predator strain induced death in the prey strain, while the prey strain repressed death in the predator strain. If conditions are right, theoretical models predict that the population sizes of the two strains will oscillate. However, unlike the population size control circuit which displayed robust stable oscillations, the predator-prey system was much harder to tune experimentally.

2.5.2 The Synthetic Coupled Oscillator System

The study of coupled oscillators began when the Dutch physicist Christiaan Huygens discovered that two pendulum clocks could synchronize their oscillations when hung

from a common support [163]. Since that time, physicists and mathematicians have come a long way in understanding how systems of coupled oscillators coordinate their behavior [164–166]. Synthetic biologists, too, have examined how oscillations in single cells can be coupled to create population-level synchrony [145, 167–170]. While the population control circuit constructed by Balagaddé *et al.* described above shows oscillations at the population level in a chemostat, it does not actually constitute a system of coupled oscillators. A system that does couple oscillations between cells was constructed by Danino *et al.* [13].

Danino *et al.* used cellular signaling proteins that create and degrade AHL (LuxI and AiiA, respectively). By placing the genes encoding these two proteins downstream of a promoter regulated by LuxR (which was constitutively expressed), a dual feedback loop is created. LuxI creates AHL, which up-regulates transcription, while AiiA destroys AHL, down-regulating transcription. The circuit diagram for this system is shown in Fig. 2.12.

Like its non-coupled predecessor [5], the coupled system displays robust oscillations. Danino *et al.* observed well-synchronized behavior of the oscillators once the cell density was high enough. Also, because the AHL had to diffuse through the population, propagating waves of oscillations were apparent. Instead of the usual system of ODEs, the authors had to use a system of partial differential equations (PDEs) to describe the spatio-temporal dynamics.

The dynamics for the two proteins, AiiA (A) and LuxI (I), can be written as the

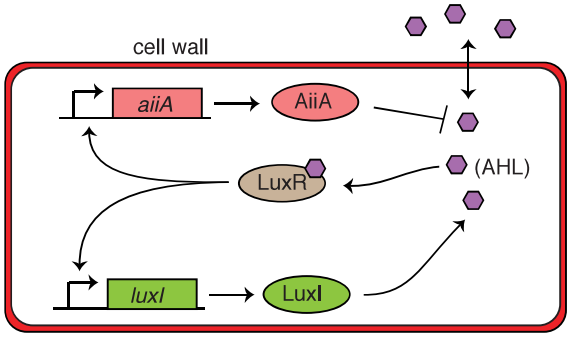


Figure 2.12 : **The coupled synthetic gene oscillator** [13]. LuxI provides the positive feedback by creating AHL, which up-regulates the promoters via constitutively produced LuxR. The enzyme AiiA provides the negative feedback by degrading internal AHL. Dynamical coupling of the cells is created by the diffusion of AHL out of each cell, where it may spatially diffuse in the medium and diffuse into other cells.

following set of PDEs:

$$\frac{\partial A}{\partial t} = \frac{\alpha_{0A} + \alpha_A H_{i,\tau}^2}{K^2 + H_{i,\tau}^2} - \frac{V_A A}{K_m + (A + I)} \tag{2.49}$$

$$\frac{\partial I}{\partial t} = \frac{\alpha_{0I} + \alpha_I H_{i,\tau}^2}{K^2 + H_{i,\tau}^2} - \frac{V_I I}{K_m + (A + I)} \tag{2.50}$$

where the production of both proteins is a time-delayed function of the internal concentration of AHL (H_i, where the τ denotes the time delay), which up-regulates the promoters via an interaction with the constitutively produced LuxR (which is not modeled). Here, α_{0i} ($i \in \{A, I\}$) is the basal and α_i the activated rate of protein synthesis. In addition, the two proteins were tagged for degradation by the same protease, and hence decay with a coupled enzymatic process, with Michaelis-Menten constants V_i and K_m. Note that while Eqs. (2.49) and (2.50) are similar to the ODE models previously described, they are still functions of space.

The spatial dependence of the system is more evident in the dynamical equations for

the internal and external concentrations of AHL (H_i and H_e, respectively). These equations are

$$\frac{\partial H_i}{\partial t} = \frac{V_c I}{K_I + I} - \frac{k A H_i}{K_H + H_i} + D\left(H_e - H_i\right) \tag{2.51}$$

$$\frac{\partial H_e}{\partial t} = -D_e\left(H_e - H_i\right) - \mu H_e + D_1 \frac{\partial^2 H_e}{\partial x^2}, \tag{2.52}$$

where internal AHL is created enzymatically by LuxI, with constants V_c and K_i and degraded enzymatically by AiiA with constants k and K_H (because AiiA is a variable, its concentration cannot be wrapped up into the normal Michaelis-Menten constant in the numerator). The diffusion of AHL into and out of the cell is modeled by a two-state diffusion process, with rates D (internally) and D_e (externally). The difference in these two rates is due to mass conservation (see [13]). Externally, AHL can be lost due to fluid flow through the chamber, at a rate μ, and is free to spatially diffuse (here in one spatial dimension) with diffusion constant D_1. In addition, Danino *et al.* were also able to model the effect of cellular density on the system. High cellular densities reduce protein production rates, but also change the effective external rate diffusion of AHL into and out of the cells, D_e [13].

2.6 Synthetic *in vitro* Gene Oscillators

One reason to create synthetic gene circuits is that they can be partially decoupled from the hosts' own regulatory processes. Of course, decoupling is far from complete, as synthetic networks still rely on the hosts' machinery to drive the reactions. Some researchers have taken the concept of decoupling the circuit from the cell even further in an attempt to analyze, in detail, every reaction that occurs in a circuit. This is done by using *in vitro* transcriptional systems that mimic *in vivo* regulatory

processes [14, 90, 171, 172]. For instance, Kim *et al.* demonstrated that ultrasensitive bi-stable switches could be constructed *in vitro* using DNA, RNA, and two enzymes: bacteriophage T7 RNA polymerase and *E. coli* ribonuclease H [172]. These networks have been extended to create cell-free oscillators [14, 90].

Of particular interest is the work of Kim *et al.* in which they construct several different circuits that exhibit oscillations: negative feedback, coupled negative and positive feedback, and a three ring repressive circuit akin to the repressilator [14]. The basic components of all of these designs are the same: 1) DNA switches that are partially double-stranded (so they cannot be transcribed); 2) single stranded DNA oligos that complete the partially double stranded strands (and hence allow transcription to proceed); 3) RNA strands transcribed from the DNA that form DNA-RNA hybrids with the single-stranded DNA oligos; and 4) enzymes that make and degrade RNA (RNA polymerase and ribonuclease H, respectively). By carefully designing the sequences of the DNA and oligos, the RNA products can either up- or down-regulate downstream components within the system.

As an example, let us examine the negative feedback circuit depicted in Fig. 2.13. The circuit is comprised of two DNA switches (Sw21 and Sw12) that, when in the "ON" state, can be transcribed to generate the RNAs rI2 and rA1, respectively. These two RNAs act to regulate the state of the switches: rI2 inhibits Sw12 and rA1 activates Sw21.

To model these networks the authors used a combination of methods. First, they analyzed "simple" models. For instance, the simple model for the negative feedback

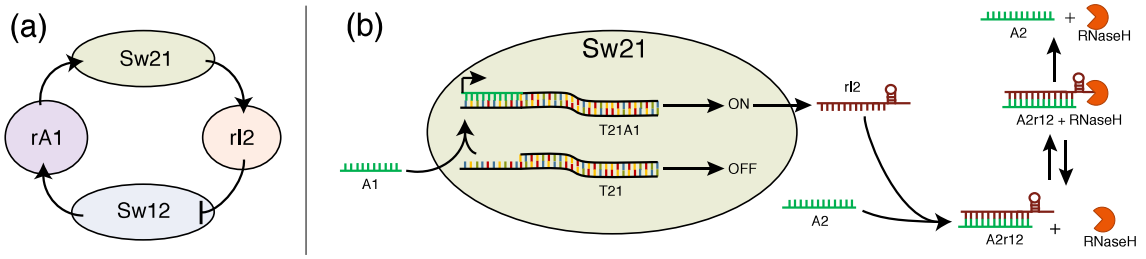


Figure 2.13 : **An *in vitro* negative feedback oscillator** [14]. **(A)** The design uses two "genes," to form a two-step negative feedback loop. Here, the genes are DNA switches (labeled Sw21 and Sw12). The products of these two genes (rI2 and rA1) act to either inhibit or activate their downstream targets. **(B)** Each DNA switch is controlled by partially double-stranded DNA. Here, we show Sw21 as an example. Sw21 is controlled by the DNA strand T21. When unbound, T21 is in the "OFF" state and cannot be transcribed. However, if the olgio A1 binds to form the T21A1 complex, the switch is turned "ON" and is free to create RNA rI2. The sequence of rI2 is such that it can bind to the oligo A2. When bound to rI2, A2 cannot bind to T12, the DNA controlling the switch Sw12. Hence rI2 acts to repress Sw12. However, RNaseH can degrade rI2 to free up A2. The switch Sw12 works in a similar fashion.

circuit can be written as

$$\frac{\mathrm{d}[rA1]}{\mathrm{d}t} = k_p[T12A2] - k_d[rA1] \tag{2.53}$$

$$\frac{\mathrm{d}[rI2]}{\mathrm{d}t} = k_p[T21A1] - k_d[rI2] \tag{2.54}$$

$$\tau\frac{\mathrm{d}[T12A2]}{\mathrm{d}t} = \frac{[T12^{tot}]}{1+([rI2]/K_I)^n} - [T12A2] \tag{2.55}$$

$$\tau\frac{\mathrm{d}[T21A1]}{\mathrm{d}t} = \frac{[T21^{tot}][rA1]^m}{K_A^m + [rA1]^m} - [T21A1], \tag{2.56}$$

where $[T12A2]$ and $[T21A1]$ are the concentrations of the "ON" state DNA switches and $[rA1]$ and $[rI2]$ are the concentrations of their respective RNA products. Each RNA is transcribed from its respective DNA at a rate k_p and is degraded at a rate k_d. Earlier work by the authors had shown that steady-state concentrations of the ON-state switches is well-modeled by Hill functions [172]. Here, this is represented

in Eqs. 2.55 and 2.56. Note that, in the steady state, these two equations give

$$[T12A2] = [T12^{tot}]\frac{1}{1+([rI2]/K_I)^n} \quad (2.57)$$

$$[T21A1] = [T21^{tot}]\frac{[rA1]^m}{K_A^m+[rA1]^m}, \quad (2.58)$$

where n and m are Hill coefficients, K_I and K_A are Hill constants, and $[T12^{tot}]$ and $[T21^{tot}]$ are the total concentrations of switch DNA (for instance $[T12^{tot}] = [T12] + [T12A2]$). The constant τ, in Eqs. 2.55 and 2.56 then gives the time scale of relaxation of transient perturbations away from these fixed points.

While simple models, like that shown above, helped the authors determine where in the design space they should expect oscillations, it was found that they did a poor job quantitatively describing the experimentally determined dynamics. In an attempt to gain quantitative agreement, they developed more detailed models that took into account all known reactions in each system. While these models are too large to reproduce here, it should be noted that they, too, had difficulty predicting the measured dynamics of the networks. The authors suggest that the presence of waste products and finite levels of materials and energy were primarily responsible for the discrepancies and lack of robustness. Still, this study exemplifies the difficulties inherent to modeling the dynamics of complex biochemical regulatory networks.

2.7 Discussion

One of the most compelling reasons to construct synthetic gene oscillators is that they provide a simple, yet highly dynamic, genetic regulatory process and thus provide more data than do circuits exhibiting only steady-state behavior. Synthetic gene

oscillators are therefore a valuable platform with which to test mathematical models of gene regulation. Thanks to the experimental construction of the various synthetic oscillators, mathematical models describing gene oscillations have evolved. We now have a better handle on what processes are important to the dynamics of gene regulation.

Still, we have a long way to go before we can say that we truly have a grasp of the dynamics of synthetic oscillators. There exist many factors influencing cellular and genetic regulatory processes not mentioned above that are usually ignored in most mathematical models. To make matters worse, the production of protein itself can influence cellular growth rates, creating a feedback loop [173]. Another major problem is that the number of non-specific reactions between the biochemical species is unknown.

If full-scale models of *in vitro* systems still have difficulty, however, what hope is there for mathematical models of synthetic gene oscillators *in vivo*? Certainly some large scale modeling of *in vivo* synthetic gene oscillators have, at least to some extent, accurately predicted experimental results – such as the model of the dual-feedback oscillator by Stricker *et al.* [5]. But the model created by Stricker *et al.* had an inordinate number of unknown parameters that had to be fit and many reactions within the circuit that had to surmised. This makes such modeling a tiresome and inexact affair.

At what level of detail, then, does one need in order to accurately model synthetic gene circuits? As of yet there is no easy answer to this question, and the question itself is one reason why synthetic gene oscillators are such an attractive model system for theorists and synthetic biologists.

Chapter 3

Experimental Materials and Methods

3.1 Microfluidics and Synthetic Biology

Microfuidic devices are an important microscopy technological advancement for synthetic biology and integral part of synthetic gene circuit development. It is critical to collect data at the single cell level to make conclusive statements about the real time dynamics of synthetic gene circuits. Microfluidic devices offer a balance of continuous real time data collection with population level statistics over long periods of time. This type of data is integral to the mathematical analysis needed to measure the dynamic response to changes or perturbations by the the synthetic gene circuit. The fluorescence data generated from these experiments can be accurately measured over time using tracking algorithms that follow cells through the cell trap area over time.

3.2 Bacterial Microfluidic Design

For the ssrA degradation with the dual feedback oscillator experiments, the four port microfluidic device designed by the Hasty lab was used for all experiments [5]. The single substrate ssrA degradation experiments needed a microfluidic device that contained a "dail-a-wave" (DAW) design which allows for the rapid and accurate switching between two distinct medias. The Hasty Biodynamics lab had previously developed a DAW device for tailored for imaging bacteria. This DAW device offered more than binary switching, allowing for incremental mixing of the two distinct

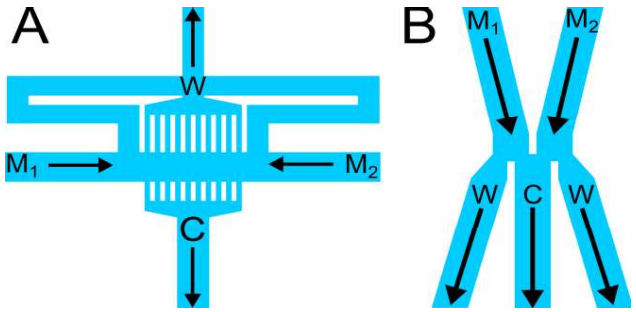


Figure 3.1 : **Comparison of DAW Designs.** (A) The incremental mixing DAW on the *E.coli* DAW microfluidic device design. (B) The simpler *S. ceresiae* DAW microfluidic device design.

medias (Fig. 3.1(A)). Additionally, they had also developed a simpler binary DAW device tailored for *Saccromyces ceresiae* (Fig. 3.1(B)).

For the single substrate degradation experiments, binary control would be sufficient, simplify workflow, and calibration of the microfluidic device considerably for each experiment. With the clean room facilities at Rice, it was possible to construct a simpler *E. coli* DAW microfluidic device. The trapping region of the yeast microfluidic device is show in Fig. 3.2 (A), where the green region traps the cells and the blue region acts as channels for media to enter and waste to leave. To accommodate for the fluid dynamic constraints of making a trap region small enough for a bacteria cell, four new designs were proposed (Fig. 3.2 (B-E)). Ultimately, none of these designs were operational and the more complicated incremental DAW was fabricated for this work.

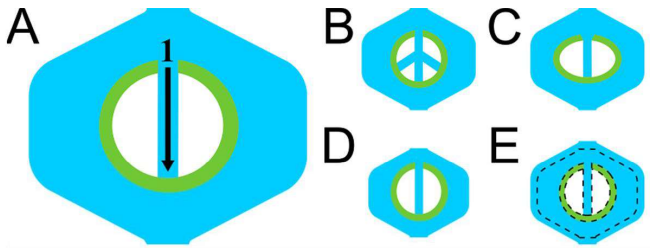


Figure 3.2 : **Reduced complexity DAW design options for bacteria.** (A) The yeast DAW trap design, where green indicated the cell trapping area and the blue are channels that bring fresh media to cells and remove waste. To accommodate for the reduced trap height for bacteria and comply with fluid dynamic constraints, the in coming channel needed to be modified. The following modifications were proposed (B) Shunts off the main channel that would aid pushing waste out of the trap area. (C) Making the trap area more oval to shorten the main incoming channel. (D) Decreasing the size of the whole trapping area. (E) Keeping the in coming channel and trap area the same, but reduce the size of exit channels.

3.3 Construction of Plasmids and Molecular Biology Protocols

3.3.1 Construction of Single Substrate Degradation Plasmids

To quantify the degradation rates of each ssrA tag variant, a set of plasmids was constructed using a pET28 plasmid backbone (Table 3.3.1). Each plasmid was consisted of the IPTG inducible promoter $p_{L-lacO1}$ promoter, yemGFP (monomeric yeast-enhanced green fluorescent protein) with the ssrA LAA degradation tag variant sequence added to the C-terminus. Once this plasmid was constructed, PCR mutagensis was used to change the last three amino acids into the LVA, AAV, and ASV variant. To create a plasmid with no degradation tag on GFP, to measure the degradation rate due to cell growth and division, the "T-S" linker between the was changed into a Spe1 restriction enzyme site using PCR mutagenesis. With Spe1 sites

Plasmid Name	Opening Reading Frame	Backbone, Resistance
pET28-GFP-LAA	$p_{L-lacO1}$:*yemGFP*:TS:LAA	ColE1, kan^R
pET28-GFP-LVA	$p_{L-lacO1}$:*yemGFP*:TS:LVA	ColE1, kan^R
pET28-GFP-AAV	$p_{L-lacO1}$:*yemGFP*:TS:AAV	ColE1, kan^R
pET28-GFP-ASV	$p_{L-lacO1}$:*yemGFP*:TS:ASV	ColE1, kan^R
pET28-GFP-NT	$p_{L-lacO1}$:*yemGFP*	ColE1, kan^R

Table 3.1 : Catalog of Single Substrate ssrA Degradation Plasmids

flanking the degradation sequence, the ssrA tag could then be removed using Spe1 digestion, and self-ligated to reform a functional plasmid.

3.3.2 Construction of Dual Feedback Oscillator Plasmids

The dual feedback oscillator ssrA variant plasmids were constructed from strains supplied from the Hasty lab [5] (Table 3.3.2). AraC-LAA and GFP-LAA were constructed on the same ColE1 plasmid. LacI-LAA was constructed on a p15A plasmid. To modify the ssrA degradation tag on AraC, a segment of the plasmid containing GFP-LAA was removed using the restriction enzymes NheI and ApaLI and self-ligated back together. Removing the GFP-LAA segment guaranteed that only the AraC ssrA tag was modified. Once removed, site directed mutagenesis was used to change the last three amino acids to LVA, AAV, and ASV. Once the variations of the ssrA degradation tag were made, GFP-LAA was reinserted back into the plasmid. To create a variant of the plasmid containing AraC with no degradation tag, the last six nucleotides of the degradation tag sequence were mutagenized to a SpeI

Plasmid Name	Opening Reading Frame	Backbone, Resistance
pZA14-LacI-LAA	$p_{Lac/Ara}$:*lac*:ssrA-LAA	p15A, amp^R
pZA14-LacI-LVA	$p_{Lac/Ara}$:*lacI*:ssrA-LVA	p15A, amp^R
pZA14-LacI-AAV	$p_{Lac/Ara}$:*lacI*:ssrA-AAV	p15A, amp^R
pZA14-LacI-ASV	$p_{Lac/Ara}$:*lacI*:ssrA-ASV	p15A, amp^R
pZA14-LacI-NT	$p_{Lac/Ara}$:*lacI*	p15A, amp^R
pJS167-AraC-LAA	$p_{Lac/Ara}$:*yemGFP*:ssrA-LAA, $p_{Lac/Ara}$:*yemGFP*:ssrA-LAA	pBR322, kan^R
pZA14-AraC-LVA	$p_{Lac/Ara}$:*yemGFP*:ssrA-LVA, $p_{Lac/Ara}$:*yemGFP*:ssrA-LAA	pBR322, kan^R
pZA14-AraC-AAV	$p_{Lac/Ara}$:*yemGFP*:ssrA-AAV, $p_{Lac/Ara}$:*yemGFP*:ssrA-LAA	pBR322, kan^R
pZA14-AraC-ASV	$p_{Lac/Ara}$:*yemGFP*:ssrA-ASV, $p_{Lac/Ara}$:*yemGFP*:ssrA-LAA	pBR322, kan^R
pZA14-AraC-NT	$p_{Lac/Ara}$:*yemGFP*, $p_{Lac/Ara}$:*yemGFP*:ssrA-LAA	pBR322, kan^R

Table 3.2 : Catalog of Oscillator ssrA Degradation Plasmids

site. The plasmid was originally constructed with a "TS" linker between the protein and degradation sequence is also a Spe1 site. Using the Spe1 restriction enzyme, the degradation tag was removed and the plasmid was self ligated. To change the degradation sequences on LacI site directed mutagenesis was used to change the last three amino acids from LAA to LVA, AAV, and ASV.

3.3.3 Construction of Orthogonal Degradation Plasmids

Orthogonal degradation tag plasmids were built using insertion mutagenesis, except RepAm, which was inserted with isothermal assembly. The pET28-GFP-NT (No Tag) was used a plasmid template to construct each orthogonal degradation plasmid (Table 3.3.3).

Plasmid Name	Opening Reading Frame	Backbone, Resistance
pET28-RepA-GFP	$p_{L-lacO1}$:RepA:TS:*yemGFP*	ColE1, kan^R
pET28-GFP-SulA	$p_{L-lacO1}$:*yemGFP*:TS:SulA	ColE1, kan^R
pET28-GFP-GGSPG-SulA	$p_{L-lacO1}$:*yemGFP*:GGSPG:SulA	ColE1, kan^R
pET28-GFP-PAPAP-SulA	$p_{L-lacO1}$:*yemGFP*:PAPAP:SulA	ColE1, kan^R
pET28-GFP-SoxS	$p_{L-lacO1}$:*yemGFP*:TS:SoxS	ColE1, kan^R
pET28-GFP-GGSPG-SoxS	$p_{L-lacO1}$:*yemGFP*:GGSPG:SoxS	ColE1, kan^R
pET28-GFP-PAPAP-SoxS	$p_{L-lacO1}$:*yemGFP*:PAPAP:SoxS	ColE1, kan^R
pET28-GFP-108	$p_{L-lacO1}$:*yemGFP*:TS:cI108	ColE1, kan^R
pET28-GFP-104	$p_{L-lacO1}$:*yemGFP*:TS:cI104	ColE1, kan^R
pET28-GFP-βGal	$p_{L-lacO1}$:*yemGFP*:TS:βGal	ColE1, kan^R
pET28-βGal-GFP	$p_{L-lacO1}$:βGal:TS:*yemGFP*	ColE1, kan^R

Table 3.3 : Catalog of Orthogonal Degradation Plasmids

3.4 Single Cell Time Lapsed Microscopy using Microfluidic Devices

Cell Culture and Media

Single substrate plasmids were transformed into JS006 LT cells and plated on LB agar + 50 μg/mL of kanamycin. A single colony was selected to inoculate an overnight culture that did not exceed 18 hours. To prepare cells for the microfluidic device, 25 μL of the overnight culture was diluted into 25 mL of LB + 2 mM of IPTG + 50 μg/mL of kanamycin and allowed to grow until an OD_{600} of 0.15 was reached. In the microfluidic device, cells grow in the presence of inducer for an additional two hours, after which the inducer was removed and the decay of green fluorescence observed by images recorded every three minutes. Red fluorescent dye was added to the inducing

media to ensure the complete removal of the inducer.

Oscillator plasmids were transformed into JS006 LB agar + 50 μg/mL of kanamycin + 50 μg/mL of ampicillin. A single colony was selected to inoculate an overnight culture that did not exceed 18 hours and 25 μL of the overnight culture was diluted into 25 mL of LB + 50 μg/mL of kanamycin + 50 μg/mL of ampicillin and allowed to grow until an OD_{600} of 0.15 was reached. On the microfluidic device, circuit dynamics are activated by the addition of 2 mM IPTG and 0.7% arabinose in LB. The dynamics were observed at single-cell resolution using microfluidic devices and time-lapse microscopy, with images recorded every three minutes [37].

3.5 Image Analysis of Single Cell Time Lapsed Microscopy Experiments

3.5.1 Segmentation

Segmentation of gray scale images is done manually. All gray scale images must be segmented before running through the MatLab algorithm (Fig. 3.3). Segmentation is the designation of an area of pixels that represents a cell. Segmenting each cell over the course of an experiment removes the need for the algorithm to (1) differentiate the area of a cell from the image background and (2) decide when a cell division event has occurred.

3.5.2 MatLab Tracking Algorithm

The tracking of cell lineage across images was done using a custom cell-tracking algorithm written in Matlab (available at github.com/alanavc/rodtracker). For each cell,

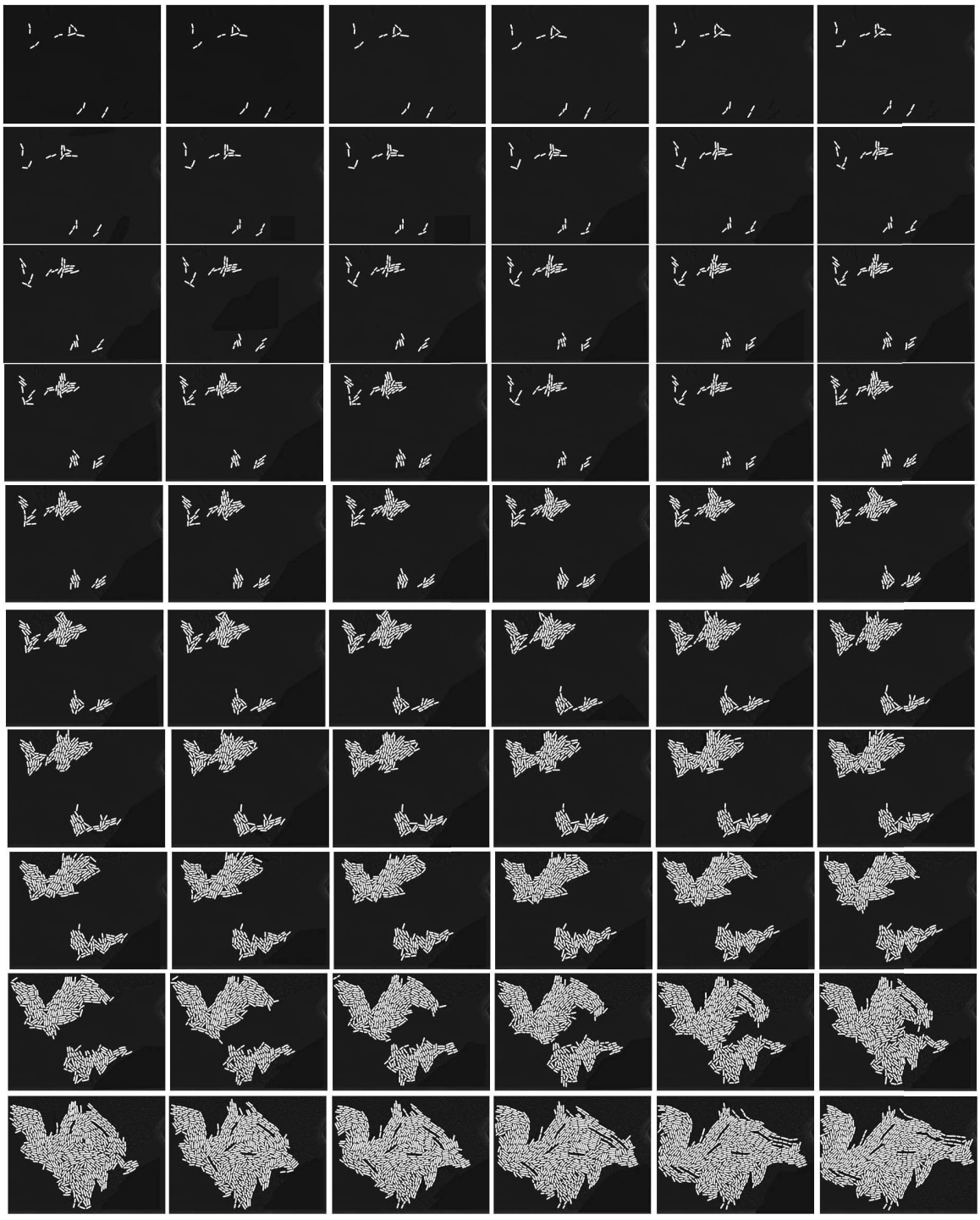


Figure 3.3 : **Segmentation files for a Completed Scope Run.** Every cells on each image of a microfluidic experiment is segmented, which allows for easy discrimination of the pixel cell area from the background by the MatLab algorithm.

C, in an image the position and length is determined, $P_C = (x, y)$ and L_C, respectively. Then, all cells were found in the next image whose position P_{next} was near P, that is $|P_C - P_{next}| < d_{move}$. The parameter d_{move} equals the maximal movement of a cell from one frame to the next. From the cells satisfying this criterion, cells were selected with length L_{next} similar to L_C, that is $|L_C - L_{next}| < d_{growth}$. The parameter d_{growth} equals the maximal expected growth between frames. All pairs of cells whose length $L_{next,1}$ and $L_{next,2}$ approximately added up to L_C were also found, that is $|L_C - (L_{next,1} + L_{next,2})| < d_{growth}$. With this criteria, a 'lineage graph' was created where each cell in an image had a set of possible transitions from one image to the next. Each transition corresponded to either movement between frames connecting the cell to a single descendant, or division and movement, connecting a cell to two descendants.

This graph was then reduced by removing inconsistent transitions (e.g., a cell can only have one possible location in the next image or two locations if it divided). The reduced graph was further reduced by only selecting transitions that minimized $\sum_{\text{cell } C}(|L_C - L_{next}| + |P_C - P_{next}|)$. The final graph consisted of transitions where each cell is associated with a unique cell (if the cell moved) or two cells (if the cell moved and divided). The lineage trajectories were then computed using the lineage graph and the fluorescence data.

MatLab Tracking Algorithm

Workflow

Input: Files with segmented cells (Fig. 3.4(A)).

Output: CSV file with lineage trajectory (Fig. 3.4(C)).

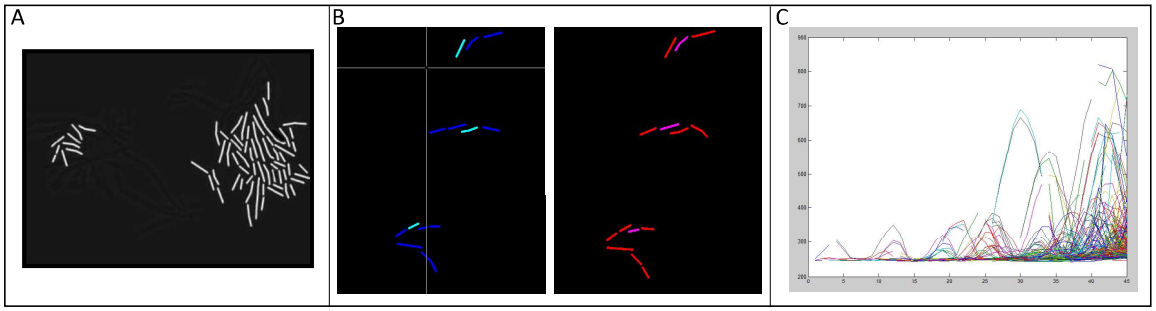


Figure 3.4 : **Simplified workflow for tracking cells in microfluidic experiments.** **(A)** Segment cells on gray scale images to easily distinguish cells from background and clearly indicate a division event. **(B)** Run segmented gray scale images through the MatLab tracking algorithm. This algorithm takes statistical guess to mapping cells from the previous image to the subsequent one. If it is confident in the guess, the cell with be dark blue on the previous image and red on the subsequent image. If it is unsure the cell will be cyan on the previous image and magenta on the subsequent image. Lastly, if it has no clue, the cell will be white. If it is a new cell emerging into the image, it will appear yellow. **(C)** Once cell are tracked through the experimental run, MatLab maps the pixels on the gray scale images onto the fluorescent images and measures the fluorescent intensity at those locations. The final output of the algorithm is a graph of fluorescence over time.

Requirement: Matlab with image processing toolbox (parallel computing toolbox is optional) (Fig. 3.4(B))

Tracking Cells from Frame to Frame

1. Set up the correct parameter values.
2. Run read_data_fun to load data in MatLab.
3. Run graph_manualcorrection_fun to do manual correction of cells yellow cells. You can switch between the current and next frame with "right click". To manually declare the lineage of a cell click on a cell in the "current frame"; then click on the two daughter cells (or twice in the same cell if it did not divide) in the "next frame". After a manual correction press ENTER to save it. If there are no more yellow cells in the current frame press ENTER to continue to next frame (Table 3.5.2).
4. Run graph_red_fun to do trivial corrections.
5. Run graph_manualcorrection2_fun to check that there are no yellow cells in any current frame. Manually correct (as in step 3) if necessary.
6. Run check_for_errors_fun to look for errors in step 5 (repeat step 5 if necessary).
7. Run graph_optimization_fun to get the most likely movement of cells.
8. Run getdata_fun to get fluorescence data in mat file.
9. Run getdata_csvfile_fun to save data in CSV format. The CSV file will have the lineage trajectories in the following tree format.

Cell Color	**Color Code Interpretation**	Action Required
Blue	Cell in current frame with exactly 1 candidate (or exactly 2 daughter cells) in the next frame.	None, correct tracking event.
Cyan	Cell in current frame with more than 1 candidate (or more than 2 daughter cells) in the next frame.	Select correct cell in next frame.
White	Cell that apparently disappeared from current frame.	Find correct cell or if left image field, select background.
Red	Cell in next frame with exactly 1 candidate in the current frame.	None, correct tracking event.
Magenta	Cell in next frame with more than 1 candidate in the current frame	Select correct cell in previous frame.
Yellow	Cell that apparently appeared in next frame	Occurs when a cell divides or was incorrectly segmented. Select daughter cells or fix segmentation.

Table 3.4 : Matlab Tracking Algorithm Color Code and Necessary Actions

Chapter 4

Understanding the Role of Protein Degradation on Synthetic Gene Circuits

Synthetic biology seeks to understand endogenous gene networks by breaking down complex gene networks into simple components. Synthetic gene circuits are built from mathematical predictions and are characterized experimentally. First-generation synthetic biology studies aimed to create a library of orthogonal parts that could be mixed and matched to facilitate circuit construction [174]. Extensive research has been conducted on the construction of new promoters, activators, and repressors; however, protein degradation in synthetic gene circuits has been restricted to the ssrA degradation system. Most synthetic circuits utilize degradation tags to normalize and speed up the rate of degradation for circuit components. Protein degradation tags are essential for synthetic gene circuit dynamics as they accelerate protein dynamics and preventing the build-up of circuit proteins that could impact circuit behavior. The ability to vary protein degradation rates allows greater control of the dynamics of synthetic gene circuits.

Currently, ssrA C-terminal degradation tags are still widely, but blindly, used in synthetic gene circuits to control the degradation rate of circuit proteins. *E. coli* natively uses the ssrA degradation system as a way to tag protein fragments for prompt degradation by the carboxy terminal-specific proteases ClpXP and ClpAP [38, 175]. This ssrA degradation tag is an 11-amino acid peptide tag that is placed on the C-

terminus of a protein fragment on a stalled ribosome [39, 53]. Previously published data showed that changing the last three amino acids of the consensus sequence (AANDENYLAA) alters the ability of the protein to bind ClpX, and thereby varying the amount of time required to degrade a tagged protein [6]. In Andersen *et al.* [6], relative degradation rates in bulk culture were established for each of these ssrA degradation tag mutants. The wild-type LAA exhibited a similar degradation rate to the mutant LVA [6]. Changing the last three amino acids to AAV and ASV slowed the bulk degradation rate significantly. Although relative bulk culture degradation rates were examined, these ssrA degradation tags have not been quantified at single-cell resolution. Assigning a degradation rate to each of these tags and understanding the influence of degradation on synthetic gene circuit dynamics can allow for better design and construction of circuits. Degradation rates are often computationally modeled using exponential decay equations. However, proteins degraded through the ClpXP machinery would more accurately be modeled through nonlinear kinetics. This work uses a simple inducible single substrate system to quantify theA degradation variants. Single cell analysis of this system provides more appropriate data for more extensive computational modeling of these degradation rates.

Once degradation rates are assigned to each ssrA degradation tag mutant, we can isolate the effects of varying this parameter on synthetic gene circuit behavior. To quantify these effects, we used the well-studied (experimentally and computationally) dual feedback oscillator as a "model" synthetic gene circuit. The dual feedback two gene oscillator was built using the lactose repressor protein, LacI, and a regulator of the L-arbinose operon, AraC (Fig. 4.1 (A)) [5]. A hybrid promoter ($p_{lac/ara-1}$) [10] was engineered to allow both AraC and LacI to competitively bind, to regulate the

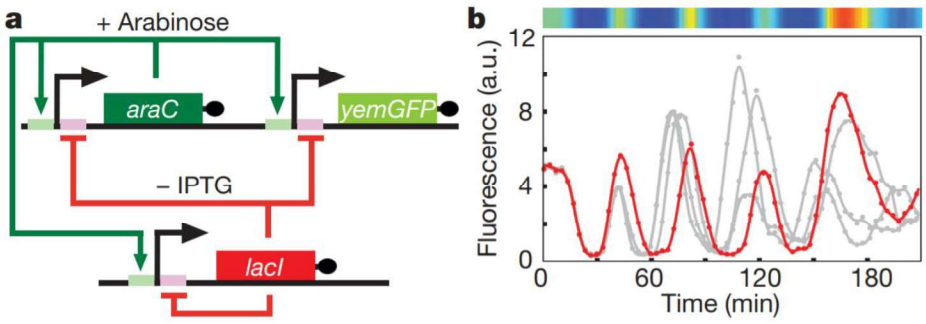


Figure 4.1 : **The Dual Feedback Oscillator. (A)** The dual feedback oscillator was constructed of an activator (AraC), a repressor (LacI), a fluorescent reporter (GFP). **(B)** The interconnected circuit topology yields sustain oscillatory behavior. Figured adapted from [5].

activity of this promoter. In the dual feedback oscillator, this hybrid promoter controls the expression of *araC*, *lacI*, and *yemGFP*. Activation of the synthetic oscillator begins with externally added arabinose, inducing the expression of AraC, and repression is modulatedbwith the addition of isopropyl β-D-1-thiogalactopryanoside (IPTG), which induces LacI dissociation and allows expression of the genes under its control. As AraC accumulates, it induces its own expression and that of LacI. As the LacI protein concentration increases, the repressor protein shuts down the expression of GFP, AraC, and itself. The activation and repression of GFP expression yields measurable fluorescencet oscillations (Fig. 4.1 (B)). All proteins are enzymatically removed from the circuit through the ssrA LAA degradation tag variant by ClpXP.

Utilizing the dual feedback oscillator as a model synthetic circuit allows for the analysis of varying degradation rates on overall circuit dynamics, both experimentally and computationally. The goal of this work is to determine a single substrate degradation rate for each ssrA degradation tag variant with single cell resolution, and determine

the effect of varying degradation rate in a synthetic gene circuit on the overall circuit dynamics. Understanding the role of degradation on dynamics allows for the ability to further tune synthetic gene circuits.

4.1 Gene Network and Experimental Design

4.1.1 Single Substrate ssrA Degradation Quantification

In order to understand the role of degradation in synthetic gene circuits, it was necessary to first quantify the degradation rate of each ssrA degradation tag variant on a single substrate. In our single substrate system, protein degradation due to dilution by cell division and enzymatic degradation of ClpXP combines to contributed to the overall protein degradation rate. Generally, as a cell divides, the overall number of proteins in a cell is divided between the two daughter cells. This degradation due to dilution rate was accurately quantified in our experimental system by measuring the protein disappearance rate of GFP with no ssrA degradation tag attached to the C-terminus of the protein. When the ssrA degradation tag was present, enzymatic degradation also occurred. The ssrA degradation tag directed the GFP to be enzymatically degraded, specifically by the ClpXP degradation machinery. Changing the last three amino acids of the 11-amino acid tag caused instability in the binding to the ClpX subunit of the ClpXP degradation machinery, thereby slowing recognition of the tag and ultimately decreasing the enzymatic degradation rate.

A set of plasmids was constructed using a pET28 plasmid backbone to analyze the degradation rates of each ssrA tag variant. Each plasmid was constructed with the

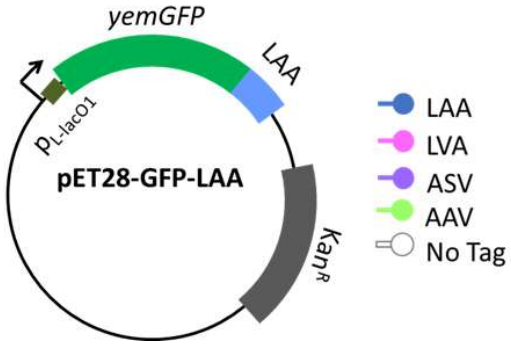


Figure 4.2 : **General plasmid design for single substrate degradation experiments.** Single substrate degradation plasmids have a pET28 Kan^R backbone with the IPTG inducible promoter $\text{p}_{L-lac01}$ regulating the expression of *gfp-ssrA*.

IPTG inducible promoter $\text{p}_{L-lacO1}$ promoter, *yemGFP* (monomeric yeast-enhanced green fluorescent protein), with the ssrA LAA degradation tag variant sequence added to the C-terminus (Fig. 4.2). Once this plasmid was constructed, PCR mutagenesis was used to alter the last three amino acids, to yield the LVA, AAV, and ASV variants. In order to create a plasmid with no degradation tag on GFP (to measure the degradation due to dilution rate by cell growth and division),the T-S linker between the protein and the tag was altered to a SpeI restriction enzyme site using PCR mutagenesis. With SpeI sites flanking the degradation sequence and, te ssrA tag wereas tn be reved using SpeI digestion. Then the plasmid was self-ligated to reform a functional plasmid.

To create a measurable single substrate system, the IPTG-inducible promoter pLac01 was used to drive the expression of GFP. The ssrA degradation tag was added to the C-terminus of GFP by insertion mutagenesis. Site-directed mutagenesis was used to change the last three amino acids for each variant. After this system was built on a

pET28 high-copy number expression plasmid, it was transformed into JS006 competent cells [5] with LacI integrated back into the genome. Cells transformed with this plasmid system were induced with 2 mM IPTG to activate GFP expression prior to loading on the microfluidic device to ensure a saturating steady state level of fluorescence. Once loaded, these cells grew in the presence of inducer for an additional two hours, after which the inducer was removed and the decay of green fluorescence was recorded over time. Red fluorescent dye was added to the inducing media to ensure the complete removal of the inducer (Fig. 4.3).

Time-lapse single cell microscopy was used to accurately measure the single substrate fluorescence degradation rate of each of the ssrA degradation tag variants. A custom microfluidic device that wllows rapid switching between inducing media and non-inducing media was used. The custom microfluidic device used a dial-a-wave junction to rapidly and accuratly switch between two distinct medias (Fig. 4.4(A-Red Circle, B)). This transition from inducing media to non-inducing media was nearly instantaneous and took less than three minutes for the red dye to no longer be detected (Fig. 4.4(C)). The single substrate plasmid (pET28-GFP-ssrA) was transformed into JS006 LT cells and plated on LB agar + 50 μg/mL of kanamycin. A single colony was selected to inoculate an overnight culture that did not exceed 18 hours. To prepare cells for the microfluidic device, 25 μL of the overnight culture was diluted into 25 mL of LB + 2 mM of IPTG + 50 μg/mL of kanamycin and allowed to grow until an OD_{600} of 0.15 was reached.

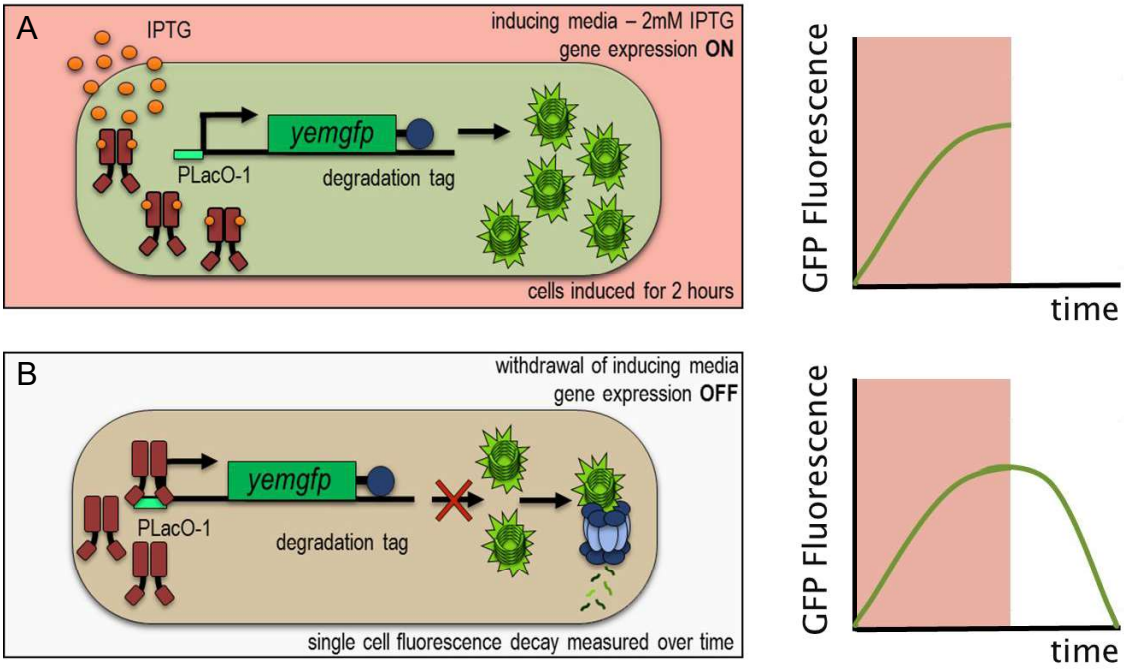


Figure 4.3 : **Experimental design for quantifying single substrate degradation rates. (A)** JSOO6 LT cells were transformed with the single substrate pET28 plasmid were induced with 2 mM IPTG (indicated by the presence of red fluorescence) for two hours. GFP fluorescence increased during this time. Then, the inducing medium was rapidly and accurately removed from the cells and replaced with non-inducing medium. **(B)** Fluorescence decay was measured over time.

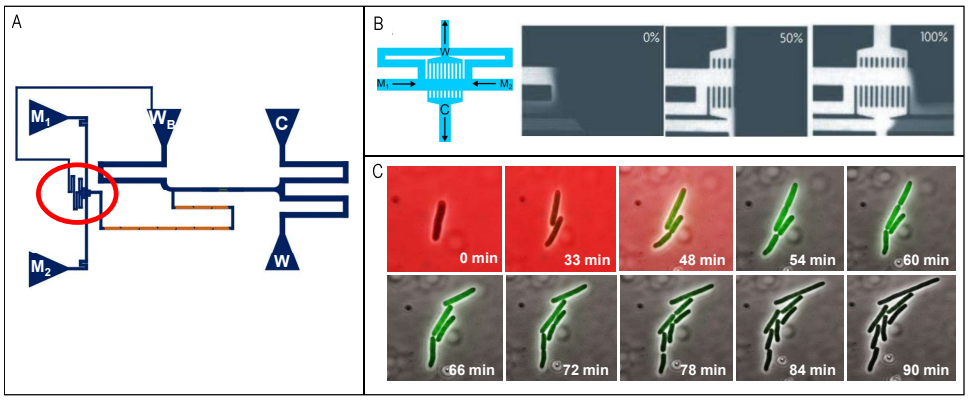


Figure 4.4 : **Microfluidic setup for single substrate degradation experiments.** **(A)** Bacterial DAW microfluidic design, where green indicates the cell trapping area, orange indicates the choatic mixers that thoroughly mix the ratios of media coming from the DAW (circled in red). M_1 and M_2 are media ports, W_B is a water balance for the DAW, C is the cell port, where cells are loaded into the device, and W_W is the waster port. **(B)** Appearance of the DAW at 0% inducing media, 50% inducing media, and 100% inducing media. **(C)** Images at 0 and 33 min of cells growing in 100% inducer (2 mM IPTG). Image at 48 min of the trapping regions, as media is being removed from the system. Images at 54 through 90 min showing cells growing in the absence of inducers; fluorescence decay is observed.

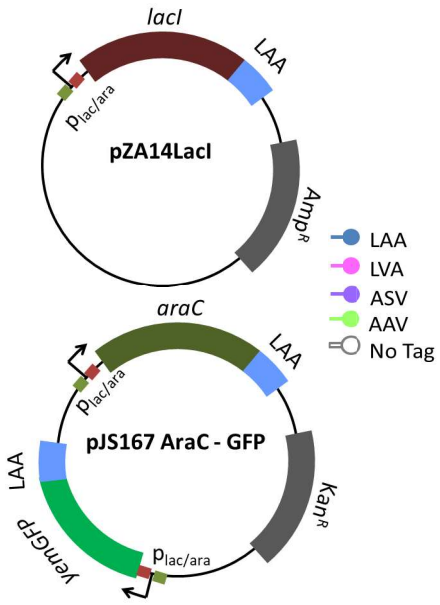


Figure 4.5 : **General plasmid design for the dual feedback oscillator two-plasmid system.** The repressor plasmid (pZA14LacI) has a p15A Amp^R backbone with the hybrid promoter $p_{lac/ara}$ regulating the expression of *lacI-ssrA*. The activator plasmid (pJS167 AraC-GFP) has a pBR322 Kan^R backbone with the hybrid promoter $p_{lac/ara}$ regulating the expression of *araC-ssrA* and *gfp-laa*.

4.1.2 ssrA Degradation Variants for Dual Feedback Oscillator Components

The dual feedback oscillator strains were supplied from the Hasty lab [5]. AraC-LAA and GFP-LAA were constructed on the same ColE1 plasmid. LacI-LAA was constructed on a p15A plasmid (Fig. 4.5). To modify the ssrA degradation tag on AraC, the segment of the plasmid that contained GFP was removed using the restriction enzymes Nhe1 and ApaLI, followed by self-ligation of the plasmid. Once the GFP-LAA section was removed, site-directed mutagenesis was used to change the last three amino acids to LVA, AAV, and ASV. After the variations of the ssrA degradation tag were made, GFP was reinserted back into the plasmid. To create a variant

of the plasmid containing AraC with no degradation tag, the last six nucleotides of the degradation tag sequence were mutagenized to an SpeI site. The plasmid was originally constructed with a "TS" linker between the protein and degradation sequence is also a SpeI site. Using the SpeI restriction enzyme, the degradation tag was removed and the plasmid was self-ligated. After all modifications to the AraC degradation sequences were made, GFP was re-ligated back into the plasmid. To alter the degradation sequences on LacI, site-directed mutagenesis was used to change the last three amino acids from LAA to LVA, AAV, and ASV.

After individual degradation rates were measured for each ssrA variant constructed, we wanted to understand how changing the degradation rate on the activator, AraC and on the repressor, LacI, in the dual feedback oscillator would alter the observed circuit dynamics. Experimentally, the two plasmids were transformed into JS006 competent cells. To activate circuit dynamics, transformed cells were grown in 2 mM IPTG and 0.7% arabinose. The dynamics were observed at single-cell resolution using microfluidic devices and time-lapse microscopy, with images recorded every three minutes (Fig. 4.6) [37] .

4.2 Results and Discussion

Fluorescence degradation was assumed to begin when the inducer was removed indicated by the absence of red fluorescence. The maximum fluorescence values varied for each degradation tag, ranging from 1,500-12,000 A.U. for all of the tags. GFP-LAA had the lowest fluorescence and both GFP-ASV and GFP-NT had the highest fluorescence values that were four times higher than GFP-LAA. The maximum fluo-

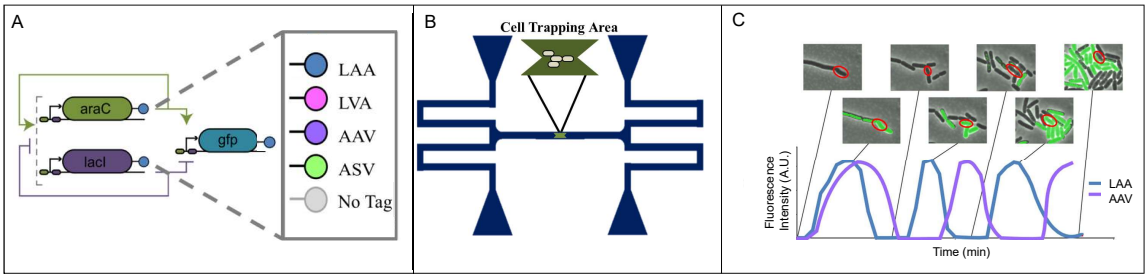


Figure 4.6 : **Experimental design for quantifying the effect of degradation on the dual feedback oscillator.** **(A)** Degradation of circuit components, AraC and LacI, were varied by altering the ssrA degradation tag variant located at their C-terminus. **(B)** The Hasty four-port microfluidic device was used to image all dual feedback oscillator microfluidic runs. **(C)** Cells were tracked over time, and their fluorescence measured and plotted to calculate changes in oscillation behavior.

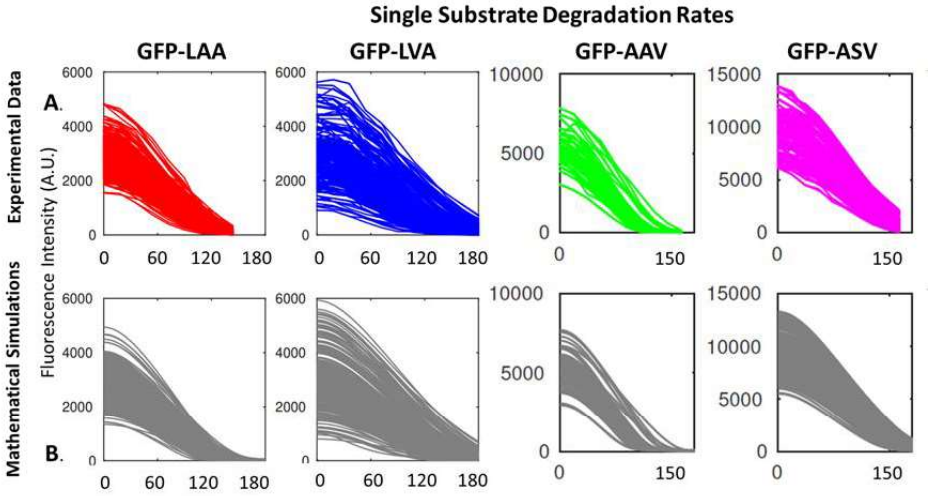


Figure 4.7 : **Enzymatic degradation mathematical models accurately predict experimental results.** **(A)** Experimental single-cell, single-substrate degradation data for each ssrA variant. From these experimental data, a model was built to calculate degradation rates for each ssrA variant. **(B)** Mathematical simulation of the experimental data. The model accurately fits the experimental data.

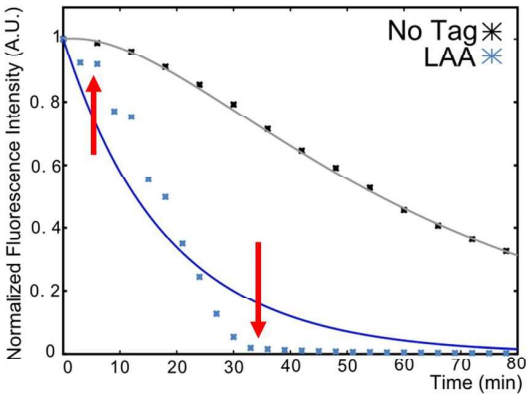


Figure 4.8 : **Degradation models that do not use nonlinear dynamics fail to accurately model enzymatic degradation data.** Red arrows indicate gaps in exponential model fit to experimental data. This is model does not capture enzymatic degradation dynamics.

rescence value can be related to the overall speed of the degradation tag (Fig. 4.7(A)). For example, because the ClpXP degradation machinery can quickly and efficiently degrade LAA tagged proteins, the steady state expression level is much lower than the steady state expression level of GFP-ASV which is more slowly degraded.

A short lag in the initiation of decay was evident after the removal of the inducer, presumably due to residual mRNA transcripts. This lag was typically around 10 minutes and could be measured from the initial non-linear section of the fluorescence decay data. The linear portion of the fluorescence decay data was used to extract the enzymatic degradation rate for each ssrA tag variant after the dilution rate was subtracted. Computational modeling was used to quantify the degradation dynamics. The following model was considered:

$$\dot{x} = \alpha \exp(-\delta t) - \beta x - h(x) \tag{4.1}$$

where $\alpha > 0$ represents the amount of RNA still present and δ represents its half-life; β is the growth rate of cells, and $h(x)$ is the degradation rate that depends on the tag. The form of $h(x)$ depends on the dynamics of degradation. Also, t is the time since IPTG is absent and decay is observed. Assuming first-order reaction for the degradation, the following model is obtained

$$\dot{x} = \alpha \exp(-\delta t) - \beta x - \beta_{\text{tag}} x, \tag{4.2}$$

where β_{tag} depends on the tag used. This model failed to accurately describe the dynamics of proteins that are ssrA-tagged and are being enzymatically degraded (Fig. 4.8 - Red Arrows.) .

A more appropriate model for enzymatically degraded proteins is the model below, in which degradation is nonlinear:

$$\dot{x} = \alpha \exp(-\delta t) - \beta x - \frac{\gamma x}{R_{\text{tag}} + x}, \tag{4.3}$$

where γ is the maximal degradation rate (when x is large), and R_{tag} is the protein concentration at which degradation is half the maximal rate. This model was able to fit degradation dynamics very well Fig. 4.7(B). Using this model, protein degradation rates were assigned for each of the ssrA variants with units of percentage of total protein concentration degradation per min. GFP-LAA had a degradation rate of 8.2%/min, that of GFP-LVA was 7.2%/min, that of GFP-AAV was 5%/min, and that of GFP-ASV was 2.2%/min (Table 4.2).

Single Substrate ssrA Variant	Degradation Rate
GFP-LAA	8.2%
GFP-LVA	7.2%
GFP-AAV	5%
GFP-ASV	2.2%

Table 4.1 : Quantified ssrA Degradation Rates.

4.2.1 Mathematical Analysis for Quantifying Single Substrate Degradation Rates

The dynamics of the concentration of a single protein can be modeled by

$$\dot{x} = \alpha - \beta x - h(x) \tag{4.4}$$

where α represents the rate of production due to constitutive expression, β is the growth rate of cells, and $h(x)$ is the degradation rate.

The simplest model of degradation is first-order decay, $h(x) = \beta_d x$, where the decay rate β_d depends on the particular properties of degradation. In the case of degradation by proteolysis, a more accurate model can be obtained by using $h(x) = \frac{\gamma x}{R_0 + x}$, which corresponds to the Michaelis-Menten kinetics. Here γ is the maximal degradation rate (when x is large), and R_0 is the protein concentration at which degradation is half the maximal rate.

The reactions that describe how the protein and protease interact also describes the protein degradation :

$$X + P \underset{k_f}{\overset{[k_r]}{\rightleftharpoons}} XP \rightarrow [k_d]P, \tag{4.5}$$

where X denotes the protein, P denotes the protease, and $k_{f/r/d}$ are the rates of the reactions. Using a Michaelis-Menten or a quasi-steady-state approximation, the degradation rate is of the form

$$\text{degradation rate of X} = -\frac{\gamma[X]}{R_0 + [X]}.$$

Here R_0 depends on k_r, k_f, and, depending on which approximation is used, may depend on k_d. Importantly, γ depends on k_d and the concentration of the degradation machinery, but not on the rates k_r, k_f. Changing the tag of a protein only changes the rates k_r, k_f, which only affects R_0 and not γ. Thus, regardless of the tag used, the value of the parameter γ will be the same.

To estimate the values of γ and R_0 from experimental data, consider the differential equation

$$\dot{x} = \alpha \exp(-\delta t) - \beta x - \frac{\gamma x}{R_0 + x}, \tag{4.6}$$

where x is the concentration of α is the protein production rate, δ corresponds to the degradation of mRNA and maturation of GFP, β is the growth rate, γ and R_0 are the Michaelis-Menten constants from enzymatic degradation, and t is the time after removal of the inducer. The term $\alpha \exp(-\delta t)$ corresponds to the fact that after the inducer is removed, no new mRNA is transcribed, but existing mRNA is still being translated and degraded.

The difference between experimental and simulated trajectories were minimized to estimate the value of the set of parameters $p = (\alpha, \delta, \beta, \gamma, R_0)$, as follows. First, Eq. 4.6 was simulated for a given set of parameters p and the trajectory $x_{\text{sim}}(t, p)$, for $0 \leq t \leq T_{\text{final}}$ was obtained. Then, the error was computed between simulations and experimental data as

$$E(p) = \sum_{0 \leq t \leq T_{\text{final}}} \left(x_{\text{exp}}(t) - x_{\text{sim}}(t, p)\right)^2,$$

where $x_{\text{exp}}(t)$ is a single-cell trajectory. A gradient descent was used to identify the set of parameters p that minimized $E(p)$. The parameters obtained are listed in Table 4.2.1. Fig. 4.7(B) shows how the simulations fit the experiments.

4.2.2 Influence of Protein Degradation on Dual Feedback Oscillator Dynamics

Initially, degradation rate parameters for AraC and LacI were increase to yield slower degradation in the previously published model [5] (Fig. 4.9). This model predicted that as component protein degradation rates slowed, oscillation periods would increase. Slowing the rate of AraC degradation slightly longer oscillations than those seen for the corresponding ssrA variant on LacI (Table 4.2.2). When the degradation tag is completely removed from AraC, the model predicts that GFP fluorescence will steadily increase, whereas removing the degradation tag from LacI, predicts the shut down of GFP fluorescence.

Experimentally, decreasing the degradation rate of circuit proteins in the dual feedback oscillator does disrupt normal oscillator dynamics (Table 4.2.2). Slowing the degradation rate of LacI corresponds with increased oscillator periods. Slowing the

parameter	value	units
β	ln(2)/30	min^{-1}
δ	ln(2)/15	min^{-1}
γ	90	(molecules/cell) min^{-1}
α_{LAA}	107	(molecules/cell) min^{-1}
α_{LVA}	118	(molecules/cell) min^{-1}
α_{AAV}	146	(molecules/cell) min^{-1}
α_{ASV}	159	(molecules/cell) min^{-1}
α_{notag}	151	(molecules/cell) min^{-1}
R_{LAA}	73	molecules/cell
R_{LVA}	278	molecules/cell
R_{AAV}	173	molecules/cell
R_{ASV}	367	molecules/cell

Table 4.2 : Estimated parameters for Eq. (4.6)

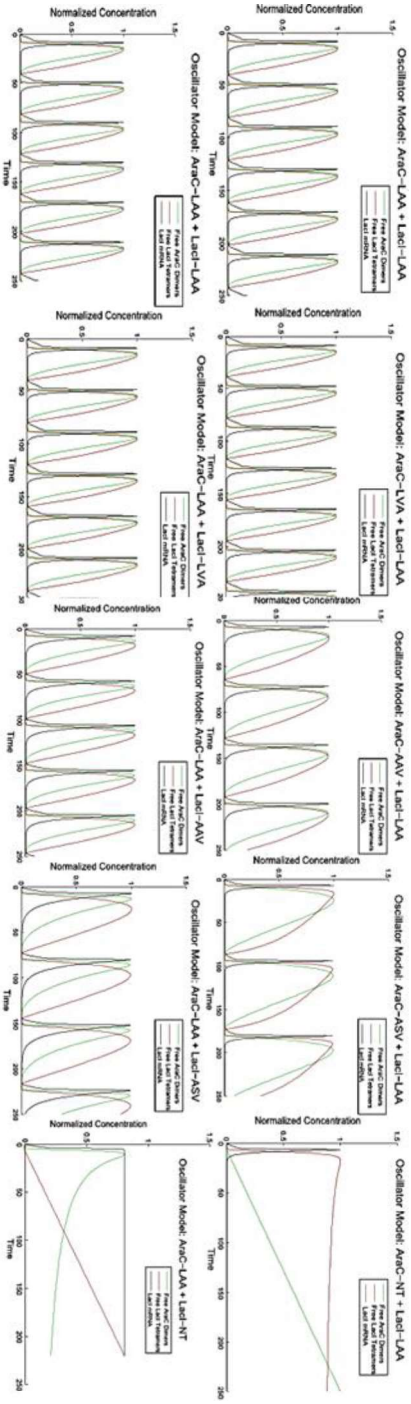


Figure 4.9 : **Predicted and experimental period lengths.** Model prediction of increased oscillation periods which correlated with decreasing degradation rates. The green line represents free AraC dimers, the red line represents free LacI tetramers, and the black line represents LacI mRNA.

AraC-LAA +	Predicted Period Length (min)	Experimental Period Length (min)
LacI-LAA	40.5	37 ±10
LacI-LVA	39.9	42 ±9
LacI-AAV	49.3	85 ±25
LacI-ASV	72.0	Cell Cycle Dependent Fluctuations
LacI-No Tag	0	No Fluorescence Observed
LacI-LAA +	**Predicted Period Length (min)**	**Experimental Period Length (min)**
AraC-LAA	40.5	37 ±10
AraC-LVA	39.0	35 ±7
AraC-AAV	55.7	Stochastic Fluctuations
AraC-ASV	86.1	Stochastic Fluctuations
AraC-No Tag	0	Stochastic Fluctuations

Table 4.3 : The predicted [5, 6] and experimental oscillator period lengths for dual feedback oscillator ssrA variants.

degradation rate of AraC causes circuit dynamics to collapse. Changing the degradation rate for the activator, AraC, led to drastic differences in circuit dynamics than what was predicted from the mathematical model. The model predicted that the oscillatory period would increase as the degradation of AraC also increased. However, oscillations were only observed in AraC-LAA and AraC-LVA. For AraC-AAV, AraC-ASV, and AraC-No Tag, stochastic fluctuations, or random noisy flashes, of fluorescence were observed. It should be noted that the fluorescence levels in these fluctuations were not much more than background, but were still measurable. An explanation for these dynamics is if the increased concentration of AraC (through slowed degradation) allows AraC to quickly bind and activate transcription as LacI is quickly enzymatically degraded.

The most interesting ssrA degradation tag variant result was that of LacI-ASV. Mathematically, the model predicts that the oscillatory period length should double the period length of LacI-LAA. Experimentally, slowing the degradation rate with LacI-ASV causes the circuit to break, and cell cycle independent oscillations were no longer observed. Initially, no fluorescence was observed. However, after 5 hours, short bursts of fluctuations were observed as the cell begins to divide (Fig. 4.10). A possible explanation for these cell cycle-dependent fluctuations is that the slow degradation of LacI-ASV allows LacI-ASV to accumulate and shut down the expression of the dual feedback oscillator. However, due to leaky expression, AraC-LAA slowly accumulates, but not enough to overpower LacI-ASV. After a long time, AraC-LAA reaches a critical protein concentration. When the cell divides and the protein concentration is halved, AraC-LAA can occupy free promoter sites and activate the expression of GFP-LAA. However, AraC-LAA and GFP-LAA are quickly degraded and LacI-ASV

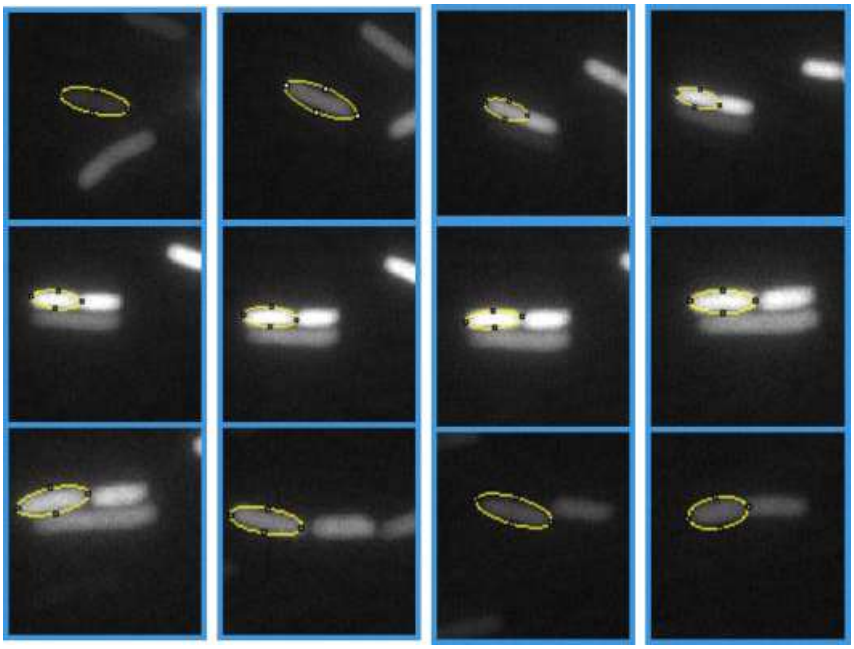


Figure 4.10 : **LacI-ASV degradation is slow enough to break the oscillator circuit dynamics and produce cell cycle-dependent fluctuations.** As the encircled cell begins to divide, GFP fluorescence increases. As the cell finishes septation, fluorescence begins to degrade.

regains control of the circuit dynamics. Another theory is that due to the overwhelming amount of GFP, and the stress of a crowded, high-density growing environment, additional protein quality control pathways are activated. For example, SulA, a Lon protease adapter/chaperon protein, is activated as an SOS signal when DNA damage is detected, resulting in cell division ceasing and cell growth cannot continue until the damaged proteins and SulA are degraded.

Computational modeling incorporating ssrA degradation rates determined from single substrate experiments was used to quantify the effect of different degradation rates on the dual feedback oscillator. As with the single-substrate case, nonlinear dynamics

were considered for protein degradation due to proteolysis (Eq. 4.3). The model describes the dynamics of LacI (r), AraC (a), immature GFP (g), and mature GFP (G)

$$\dot{r} = \alpha_r h\left[r_{\tau_r}, a_{\tau_r}\right] - g_r(r,a,g,G) - \beta r, \tag{4.7}$$

$$\dot{a} = \alpha_a h\left[r_{\tau_a}, a_{\tau_a}\right] - g_a(r,a,g,G) - \beta a, \tag{4.8}$$

$$\dot{g} = \alpha_g h\left[r_{\tau_g}, a_{\tau_g}\right] - g_g(r,a,g,G) - \lambda g - \beta g, \tag{4.9}$$

$$\dot{G} = \lambda g - g_G(r,a,g,G) - \beta G, \tag{4.10}$$

where $x_\tau = x(t-\tau)$ for x in $\{r,a,g\}$, and

$$h(r,a) = \frac{f^{-1} + \left(\frac{a}{C_a}\right)^2}{\left(1+\left(\frac{a}{C_a}\right)^2\right)\left(1+\left(\frac{r}{C_r}\right)^4\right)^2}, \tag{4.11}$$

is the composite Hill function describing the activity of the hybrid promoters as a function of activator and repressor concentrations, and

$$g_x(r,a,g,G) = \frac{\gamma_x x / R_{x,\mathrm{tag}_x}}{1 + r/R_{r,\mathrm{tag}_r} + a/R_{a,\mathrm{tag}_a} + g/R_{g,\mathrm{tag}_g} + G/R_{G,\mathrm{tag}_G}} \tag{4.12}$$

is the nonlinear function describing enzymatic degradation of x in $\{r,a,g,G\}$.

Here, α_r, α_a and α_g are the maximal production rates, and τ_r, τ_a, and τ_g are the transcriptional delay times. For x in $\{r,a,g,G\}$, γ_x and R_{x,tag_x} are the Michaelis-Menten parameters for ClpXP-mediated proteolysis, and C_a and C_r are the concentrations needed for half-maximal induction and repression. In addition, f is a unitless measure of the strength of the activation by a compared to basal production, λ is the maturation rate of GFP, and β is the dilution rate due to cell growth.

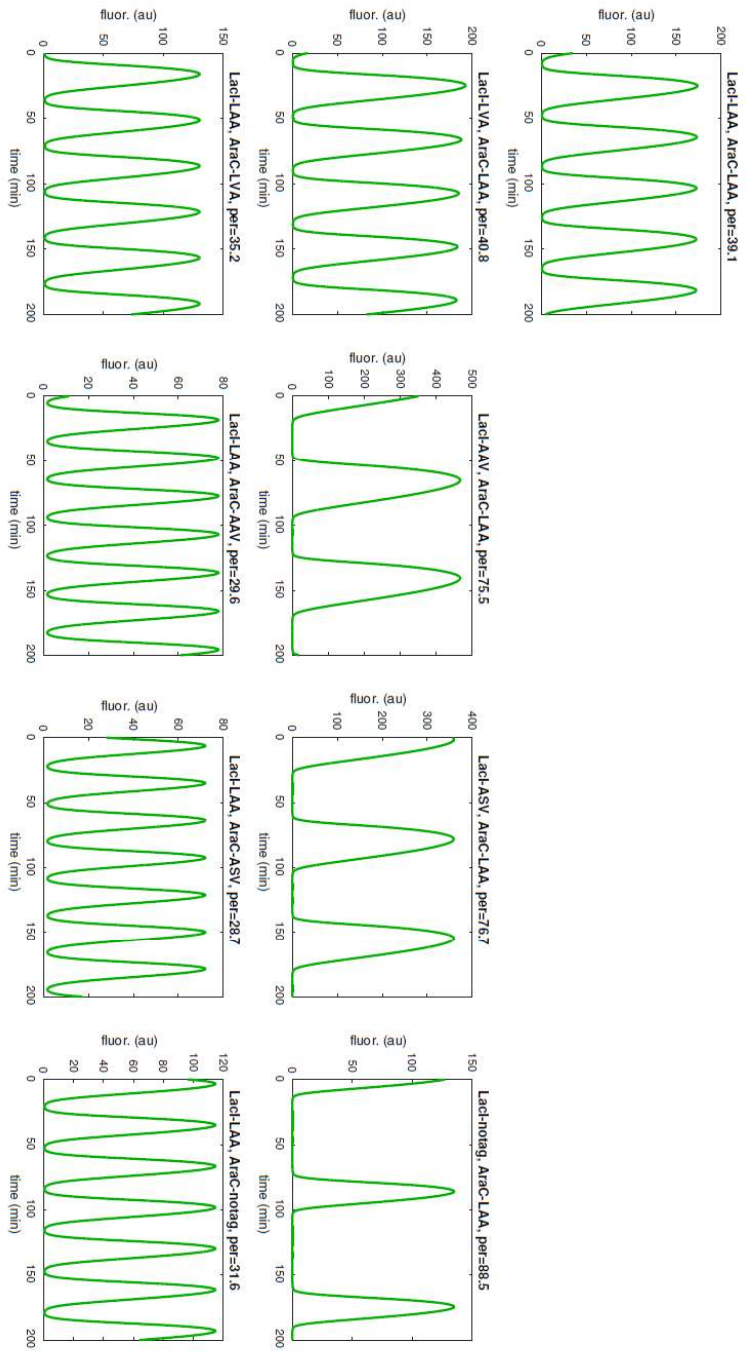


Figure 4.11 : **Incorporation of single substrate degradation rates was insufficient to predict experimental dual feedback oscillator data degradation data.** These predictions incorporated a delay for the degradation of LacI due to its increased stability, but very poorly predict circuit behavior.

Initially, a predictive model could not be achieved (Fig. 4.11). It was only possible to generate parameter values from experimental data fits, and this approach caters to the initial conditions to each degradation tag variation, providing little insight about circuit dynamics. However, recently it has been shown that the overall stability of a protein greatly impacts the ability of ClpXP to degrade it [75]. The ClpXP degradation machinery is composed of a homohexamer ring of ClpX that binds to a homoheptameric ClpP peptidase [7]. Once a protein has been tagged for degradation, the adapter protein SspB binds to the AADENY peptide sequence of the ssrA degradation tag [38]. SspB guides the protein to the ClpX ATPase subunit and anchors the protein to the ClpX subunit. The last three amino acids are responsible for the binding affinity of the ATPase and the tagged protein. Changing the last three amino acids disrupts the ability of the tag to properly dock to ClpX, causing a delay in translocation into ClpP. Previous research has shown that this process is not highly specific, and the tagged protein can go through multiple rounds of binding to and dissociation from ClpX before it properly docks to the degradation machinery [75]. When this docking is complete, ClpX exerts a mechanical force to physically break the tertiary structure of the protein. ClpX uses ATP to generate a power stroke that pulls on the proteins, attempting to strain the existing protein structure [40, 75, 81]. This effort is necessary to unfold the protein into a peptide chain that can be successfully fed through the ClpX pore to be subsequently degraded by the ClpP peptidase in 5 to 15 amino acid increments [7]. Not all of the ClpX at one time can bind substrates and act as a rigid body for this mechanical force. Depending on the stability of the tagged protein, one power stroke may not be sufficient to achieve global denaturation. This requirement for another ATP hydrolysis event to generate a power stroke can cause the target protein to lose affinity and be released. SspB remains attached to

the strained protein and attempts to bind again to ClpX until the protein is unfolded and degraded in its entirety [40, 75, 81].

From a mathematical model perspective, it is evident that the previous dual feedback oscillator model cannot accurately capture all of the biophysical and mechanical dynamics that occur when a protein is being degraded. In addition, it would be difficult to correctly model, assuming classic Michaelis-Menten enzyme kinetics. Previous mathematical models for the dual feedback oscillator did not allow for different Michaelis-Menten parameter values (R_0) for LacI, AraC, and GFP. The biochemical evidence of the ClpX power stroke, the binding parameter R_0 was allowed to change with each substrate. Prior to these data, there was no experimental evidence of a change in the value of R_0(AraC) and R_0(LacI). However, given that LacI is a tetramer and one of the most stable proteins, it should take many more rounds of mechanical effort to strain the protein enough to destroy its protein structure and be degraded as a peptide chain, compared to AraC. With this change, it was possible to generate a predictive model that fits well with the experimental data for LacI ssrA variants. Additionally, capturing the dynamics competition of tagged protein targets to the binding of ClpXP gives the most accurate enzymatic degradation term for the dual feedback oscillator. The model even predicted a short burst in fluorescence that occurs when LacI is tagged with ASV (Fig. 4.12). For AraC, we could easily predict the dynamics for LAA and LVA, however, predicting the correct amplitude (fluorescence intensity) is challenging, but the fast, short oscillation periods were captured by the model.

The results of this work are important to the area of synthetic biology for several

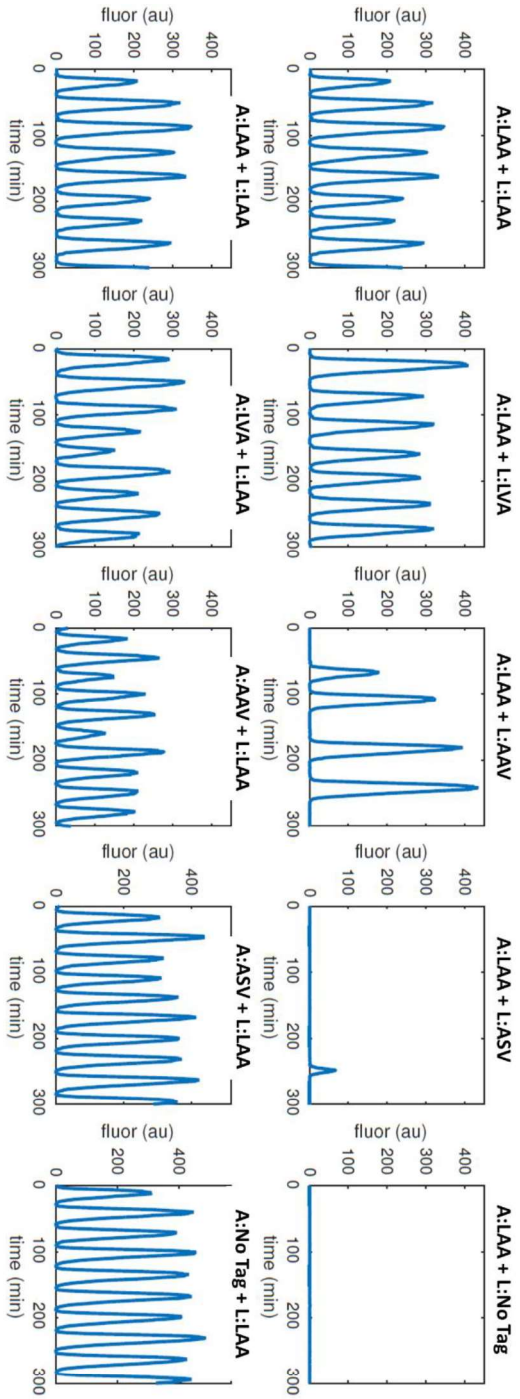


Figure 4.12 : **Incorporation of substrate stability and binding competition to ClpXP into a degradation term accurately predicts experimental data.** With a enzymatic degradation term that comprehensively accounts for the degradation delay to (1) substrate stability and (2) binding competition to ClpXP, it is possible to accurately predict experimental data–even the cell cycle dependent fluctuations for LacI-ASV and AraC-LAA.

reasons. First, it adds to the knowledge of parts available to build novel synthetic networks. Quantifying each degradation tag rates provides synthetic biologists new parameters that can be used to fine-tune circuit dynamics. In addition, it provides accurate parameter values for building descriptive mathematical models. This point is crucially important, because a mathematical understanding of the parameters that are critical to the functioning of a novel circuit reduces the amount of time needed to experimentally fine-tune the circuit.

Chapter 5

Engineering Orthogonal Degradation Tags

The ClpXP machinery is an efficient machine that requires ATP to function. To conserve energy, bacterial cells have evolved to maintain critical amounts of these enzymes [176] [177] [178]. Previous research has shown that exclusive use of ssrA degradation tags can saturate the degradation machinery, thereby generating coupled network behavior [15]. Additional ssrA-tagged proteins, allow for the possibility to overload the ClpXP degradation machinery. With this unexpected burden on these proteases, tagged proteins must wait in line to be degraded (Fig. 5.1); directly coupling the tagged proteins and leading to a strong correlated observed behavior [15]. Mathematically, this phenomenon is a biological application of queuing theory.

Previous studies have demonstrated queuing experimentally by transforming *E. coli* cells with the dual feedback oscillator and an inducible promoter driving the expression of CFP-LAA (Fig. 5.2(A)). Knowing that the dual feedback oscillator does not overload the degradation machinery while operational, CFP expression can gradually be increased until oscillations are no longer detectable and correlative fluorescent patterns are observed (Fig. 5.2(B)). Transition from an underloaded to an overloaded ClpXP state reveals itself in significant crosstalk between these two independent networks.

Therefore, not only can slower degradation rates alter network dynamics, but copi-

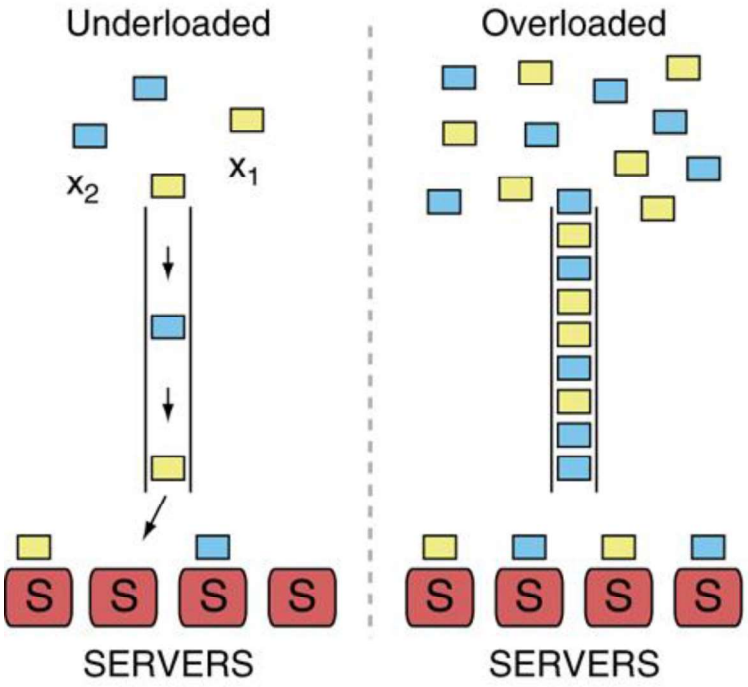


Figure 5.1 : **Correlation of overwhelmed protein degradation machinery to queuing theory.** The yellow and blue boxes represent two different types of jobs that the servers can accomplish. When the system is underloaded, there are more servers than there are jobs. When the system becomes overloaded, the excess number of jobs piles up, requiring a wait time to be processed by the limited amount of servers. Experimentally, each colored box represents a different protein. Figure adapted from [15] with permission.

ous amounts of tagged protein can overwhelm the degradation machinery and halt network dynamics. Thus, expansion of the library of degradation tags to include orthogonal proteases that minimize circuit crosstalk is needed. Additionally, not all proteins or engineered split proteins have a free, non-functional C-terminus to devote to a degradation tag (e.g. split T7 polymerase). An N-terminal degradation tag would allow finer control of these types of circuit proteins. To build larger, more complex synthetic gene circuits, it is necessary to look for orthogonal degradation machinery in the native degradation pathways of *E. coli*.

In addition to ClpXP, there are other AAA+ proteases native to *E. coli* that aid in regulating protein quality control. ClpAP, HslUV (ClpYQ), Lon, and FtsH are energy-dependent enzyme complexes that often require chaperones to degrade specific substrates [7]. All of these proteases are ATP-dependent proteases, with ATPase domains of these complexes are responsible for substrate recognition [124]. These chaperone proteins or the substrates themselves tightly bind to specific sites on the proteases. Using the amino acid binding sequences as inspiration, degradation tags were engineered to target this additional degradation machinery. Prospective peptide tags that target alternative degradation proteases include: RepA (ClpAP protease), SulA (Lon protease), c1-104/108 (HflB protease), and SoxS (Lon protease) (Table 5).

Each of the degradation tags listed above was cloned into the IPTG-inducible pET28-GFP plasmid using insertion mutagenesis (with the exception of RepA, which was inserted using isothermal assembly). Degradation tags were crudely evaluated using a fluorescent plate reader. Liquid cultures were induced with 2 mM IPTG for 4 hours. An aliquot of the culture was then spun down, washed several times with LB, and

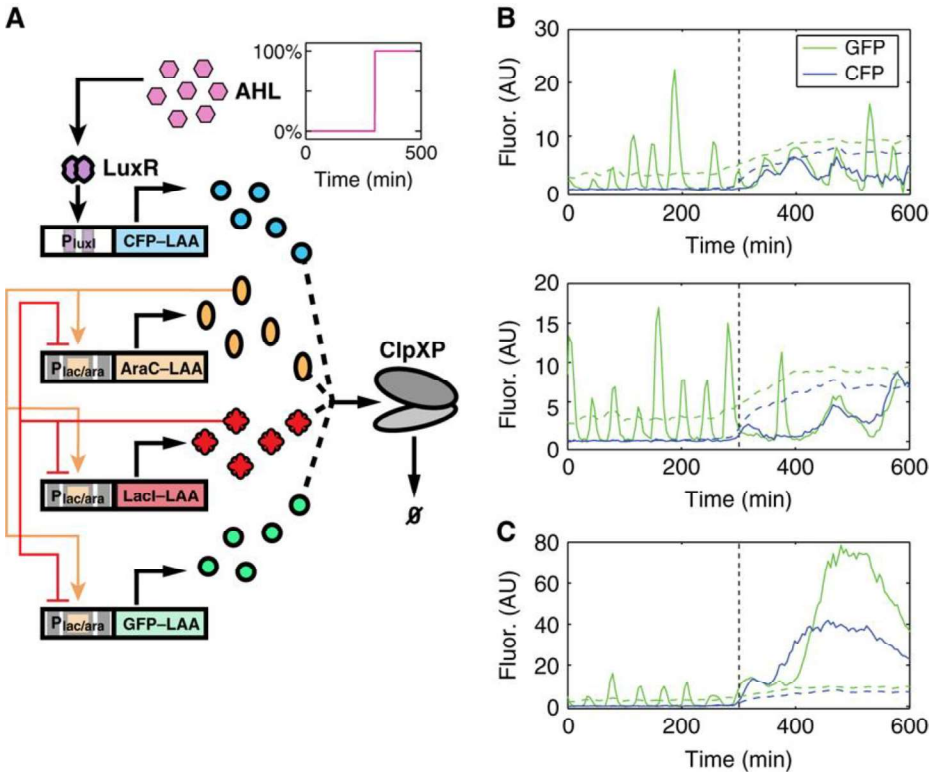


Figure 5.2 : **It is possible to overwhelm the ClpXP degradation machinery resulting in coupled circuit dynamics.** (A) The network is designed to exhibit enzymatic coupling. AraC-LAA, LacI-LAA, and GFP-LAA are components of the dual feedback oscillator. CFP-LAA is produced with the addition of the inducer, AHL. (B) Whole-field fluorescence traces show CFP and GFP oscillations, indicating enzymatic coupling. Figure adapted from [15] with permission.

then resuspended in LB without IPTG. The induced cells and the newly suspended cells were loaded into a 96-well plate. The OD_{600} and the green fluorescence were measured every 10 min for 8 to 12 hours. Promising degradation tags were then more accurately quantified by single-cell time-lapse microscopy. GFP expression was induced with 2 mM IPTG for 2 hours and then withdrawn. The amount of fluorescence was recorded every 3 minutes for 8 to 12 hours.

Peptide Tag	Tag Placement	Amino Acid Sequences	Target Protease	Published Half Life
ssrA	C-Terminal	AANDENYALAA	ClpXP	16 min
RepA	N-Terminal	MNQSRSDILYADIE	ClpAP	unknown
SulA	C-Terminal	KIHSNLYH	HslVU	1.2 min
c1-104	C-Terminal	IHTVR	HflB	unknown
c1-108	C-Terminal	SLLWS	HflB	45 min
SoxS	C-Terminal	PSDYRHRL	Lon	2 min

Table 5.1 : Summary of Degradation Tag Catalog

5.0.1 ClpAP Degradation Machinery: RepA Degradation Tag

The proteolytic component, ClpP, can also bind with another ATPase, ClpA [179]. ClpA functions similarly to ClpX, degrading substrates specifically, starting from a variety of degradation signals [180], including RepA [181]. RepA is a pSC101 plasmid protein that natively controls cell division through the SOS response in *E. coli* [181]. Several research groups have demonstrated that that last 15 amino acid residues of the protein can be synthetically added to GFP, and this is sufficient for ClpAP binding and subsequent degradation [179]. Most importantly, these data show that this 15-amino acid RepA tag can be recognized by the ClpAp protease whether it is attached to the C- or N-terminus of a targeted protein. RepA was selected as the first orthogonal degradation tag variation, because it would be incredibly useful in split-protein systems. Due to its large insertion size, the RepA degradation tag was added to the N-terminus of GFP by isothermal assembly.

5.1 Targeting the Lon Protease: SulA and SoxS.

The Lon protease was the first AAA+ protease to be discovered in bacteria, although ClpXP is the best characterized [182]. The Lon protein does not need an additional protein subunit; rather, it is a single peptide in which the AAA+ subunit is fused to the peptidase subunit [7]. Although it maintains a barrel-like structure, the Lon protease is composed of a homohexameric ring that is transcribed from a single gene [182]. Two of these hexamers combine to form a functional dodecamer complex. [7]. Interestingly, the Lon protease is the only AAA+ protease that can bind DNA, although the purpose of this binding remains a mystery [182].

Lon degrades a broad range of proteins and is responsible for the majority of ATP-dependent degradation of misfolded proteins [182]. For general degradation, Lon recognizes exposed hydrophobic amino acid sequences (about 15 residues) that are rarely exposed [182]. The lack of constraints on the sequence of the target hydrophobic residues allow the Lon protease to degrade a large range of substrates. However, Lon is also capable of rapid targeted degradation of specific substrates, including SulA, SoxS, and MarA [175].

5.1.1 SulA C-terminal Degradation Tag

Natively, SulA is a 18-kDa protein comprised of 169 amino acids that is produced when DNA damage is detected [183]. SulA interacts with the cell division protein FtsZ to prevent the progression of cell division. This "hold" on cell division is lifted when transcriptional regulation of the SOS signal ceases or when SulA is degraded by Lon [184–188]. SulA levels are tightly regulated and, this protein is rapidly degraded;

as excessive accumulation of SulA is toxic, making it an attractive novel degradation tag.

In the absence of an existing synthetic SulA degradation tag, the logic of the RepA degradation tag was applied to create a new synthetic degradation tag using the last 15 amino acids of these proteins, which are responsible for recognition and binding of the Lon protease, to create this degradation tag. The SulA degradation tag was cloned using insertion mutagenesis to the C-terminus of GFP (maintaining the TS linker).

5.1.2 SoxS C-terminal Degradation Tag

SoxS is a transcriptional activator that is expressed as a defense against reactive oxygen species, specifically superoxide anion [189]. The synthesis of SoxS then triggers the transcription of genes of the defense regulon [190]. Once the oxidative stress is mitigated, SoxS is rapidly degraded by the Lon protease. Similar to SulA, this rapid removal of SoxS makes it an ideal candidate for a novel degradation tag.

Previous research used western blots to demonstrate that the last several amino acids of SoxS fused to GFP were enough to confer rapid (half-life of 2 min) degradation by the Lon protease. The SoxS degradation tag was cloned using insertion mutagenesis to the C-terminus of GFP (maintaining the TS linker). An extended SoxS degradation tag comprised of 15 amino acids was built, but never evaluated.

5.2 Targeting the HslVU Protease: cI104 and cI108

Lastly, the cI degradation tag is a pentapeptide nonpolar tail that has been shown to cause the degradation of the stable phage λ cI repressor. Although ClpXP, ClpAP, and Lon account for 70-80% of the proteolytic degradation in *E. coli*, the membrane bound ATP-dependent zinc protease HflB (FtsH) is the only essential protease [191]. Natively, this protein is responsible for degrading the shock sigma factor σ^{32} and the regulatory proteins cII and cIII of phage λ [192,193]. Previous research demonstrated that adding the nonpolar pentapeptide tail of the cI variants cI-104 and cI-108 to the C-terminus of GFP leads to target degradation by HflB. cI-104 and cI-108 were selected as potential orthogonal degradation tags because the tags are considerably shorter than other potential tags (5 amino acids vs 815 amino acids). Additionally, a membrane-bound targeted degradation system could be an elegant way to spatially control degradation in synthetic gene circuits. Both cI synthetic degradation tags were cloned to the C-terminal of GFP by insertion mutatagenesis.

5.3 The Effect of Linkers on Orthogonal Tag Degradation Rates

The single-cell degradation trajectories resemble nonspecific degradation due to cell dilution. A review of literature brought up the possibility of adding flexible linkers to increase the degradation rate [56]. Subsequently, a talk at the Winter qBio conference inspired a strategy to insert a proline into the linker (GGSP) to make the linker more rigid, and also to test the more rigid PAPAP repeated linker. After adding proline to the linkers, degradation of GFP expression for SoxS-PAPAP, SoxS-GGSPG,

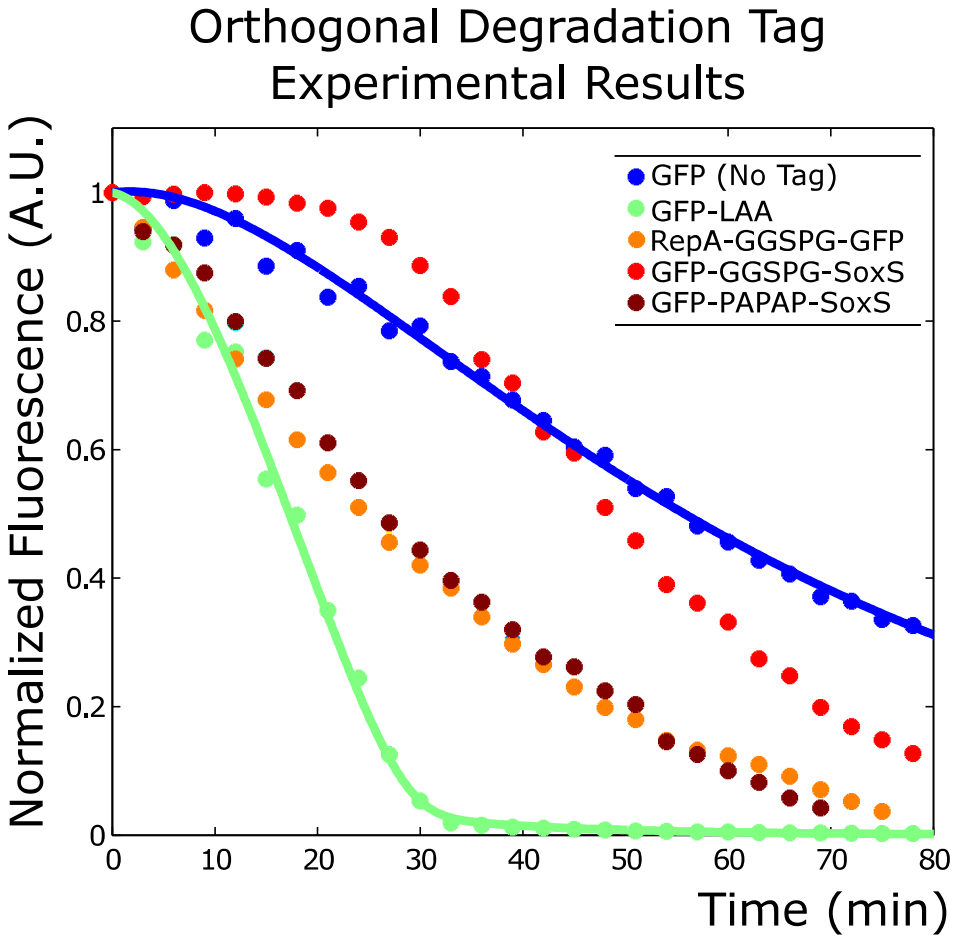


Figure 5.3 : **Using single-cell time-lapse microscopy, each of the orthogonal degradation tags was quantitated and compared to ssrA degradation tag variants.** It became apparent that SoxS-GGSPG (red dots) was not enzymatically degraded and was being degraded through non-specific degradation and cell dilution. However, SoxS-PAPAP (maroon dots) and RepA-GGSPG (orange dots) appear to be enzymatically degraded. The difference in decay shape could be due to differences in the targeted degradation machinery. RepA targets ClpAP and SoxS targets the Lon protease.

and RepA-GGSPG was observed. Each of these orthogonal degradation tags was quantified using single-cell time-lapse microscopy in a similar experimental setup to the single substrate degradation experiments. It became apparent that SoxS-GGSPS ((Fig. 5.1:red) was not being enzymatically degraded, and rather was being degraded at a rate similar to non-specific degradation. RepA-GGSPG ((Fig. 5.1:orange) and SoxS-PAPAP ((Fig. 5.1:maroon) appeared promising, even though they did not have the same decay shape as the ssrA degradation tag variants. This difference in shape could be due to differences in the ClpAP (targeted by RepA) and the Lon (targeted by SoxS) machinery. For example, the ClpA hexamer is at a much lower concentration than that of the ClpX hexamer. Due to these unclear results, this project to engineer orthogonal degradation tags was discontinued.

References

[1] O. Purcell, N. J. Savery, C. S. Grierson, and M. di Bernardo, "A comparative analysis of synthetic genetic oscillators," *J Roy Soc Inter*, vol. 7, no. 52, pp. 1503–1524, 2010.

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

[3] M. Tigges, T. T. Marquez-Lago, J. Stelling, and M. Fussenegger, "A tunable synthetic mammalian oscillator," *Nature*, vol. 457, no. 7227, pp. 309–312, 2009.

[4] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S. G. Lee, and J. C. Liao, "A synthetic gene-metabolic oscillator," *Nature*, vol. 435, no. 7038, pp. 118–122, 2005.

[5] 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*, vol. 456, no. 7221, pp. 516–519, 2008.

[6] J. B. Andersen, C. Sternberg, L. K. Poulsen, S. P. Bjorn, M. Givskov, and S. Molin, "New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria," *Appl Environ Microbiol*, vol. 64, pp. 2240–2246, 1998.

[7] A. O. Olivares, T. A. Baker, and R. T. Sauer, "Mechansistic insights into bac-

terial aaa+ proteases and protein-remodelling machines," *Nature Reviews Microbiology*, vol. 14, pp. 33–44, 2016.

[8] S. Glynn, A. Martin, A. Nager, T. Baker, and R. Sauer, "Crystal structures of asymmetric clpx hexamers reveal nucleotide-dependent motions in a aaa+ protein-unfolding machine.," *Cell*, vol. 139, pp. 744–756, 2009.

[9] D. Bolon, D. Wah, G. Hersch, T. Baker, and R. Sauer, "Bivalent tethering of sspb to clpxp is required for efficient substrate delivery: a protein-design study.," *Mol Cell.*, vol. 13, pp. 443–449, 2004.

[10] R. Lutz and H. Bujard, "Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements," *Nucl Acids Res*, vol. 25, pp. 1203–1210, 1997.

[11] M. R. Atkinson, M. A. Savageau, J. T. Myers, and A. J. Ninfa, "Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in escherichia coli," *Cell*, vol. 113, no. 5, pp. 597–607, 2003.

[12] F. K. Balagaddé, 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*, vol. 309, pp. 137–140, 2005.

[13] T. Danino, O. Mondragon-Palomino, L. Tsimring, and J. Hasty, "A synchronized quorum of genetic clocks," *Nature*, vol. 463, no. 7279, pp. 326–330, 2010.

[14] J. Kim and E. Winfree, "Synthetic in vitro transcriptional oscillators," *Mol Syst Biol*, vol. 7, p. 465, 2011.

[15] N. A. Cookson, W. H. Mather, T. Danino, O. Mongragon-Palomino, R. J. Williams, L. S. Tsimiring, and J. Hasty, "Queueing up for enzymatic processing: correlated signaling through coupled degradation," *Mol Syst Biol*, vol. 7, pp. 1–9, 2011.

[16] F. Jacob and J. Monod, "Genetic regulatory mechanisms in synthesis of proteins," *J Mol Biol*, vol. 3, pp. 318–356, 1961.

[17] D. E. Cameron, C. J. Bashor, and J. J. Collins, "A brief history of synthetic biology," *Nature Reviews Microbiology*, vol. 12, pp. 381–390, 2014.

[18] M. Ptashne, A. D. Johson, and C. O. Pabo, "A genetic switch in a bacterial virus," *Sci Am*, vol. 247, pp. 128–130, 1982.

[19] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. Eng, R. Bumgarner, D. Goodlett, R. Aebersold, and L. Hood, "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network," *Science*, vol. 292, pp. 929–934, 2001.

[20] H. V. Westerhoff and B. O. Palsson, "The evolution of molecular biology into systems biology," *Nature Biotech*, vol. 22, pp. 1249–1252, 2004.

[21] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi, "The large-scale organization of metabolic networks," *Nature*, vol. 407, pp. 651–654, 2000.

[22] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology," *Nature*, vol. 402, pp. C47–C52, 1999.

[23] D. Bray, "Protein molecules as computational elements in living cells," *Nature*, vol. 376, pp. 307–312, 1995.

[24] H. H. McAdams and L. Shapiro, "Circuit simulation of genetic networks," *Science*, vol. 269, pp. 650–656, 1995.

[25] H. H. McAdams and L. Shapiro, "Towards a circuit engineering discipline," *Curr Biol*, vol. 10, pp. R318–R320, 2000.

[26] B. C. Goodwin, "Oscillatory behavior in enzymatic control processes," *Adv Enzyme Regul*, vol. 3, pp. 425–438, 1965.

[27] A. F. J. Tiwari and R. Beckman, "Genetical feedback represson i, single locus models," *J Theor Biol*, vol. 45, pp. 311–326, 1974.

[28] A. Fraser and J. Tiwari, "Genetical feedback-repression ii, cyclic genetic systems," *J Theor Biol*, vol. 47, pp. 397–412, 1974.

[29] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, "Delay-induced stochastic oscillations in gene regulation," *Proc Natl Acad Sci USA*, vol. 102, pp. 14593–14598, 2005.

[30] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in escherichia coli," *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[31] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.

[32] M. Tigges, N. Denervaud, D. Greber, J. Stelling, and M. Fussenegger, "A synthetic low-frequency mammalian oscillator," *Nucl Acids Res*, vol. 38, no. 8, pp. 2702–2711, 2010.

[33] P. Smolen, D. A. Baxter, and J. H. Byrne, "Requency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems," *Am J*

Physiol, vol. 274, pp. C531–C542, 1998.

[34] J. Hasty, M. Dolnik, V. Rottschafer, and J. J. Collins, "Synthetic gene network for entraining and amplifying cellular oscillations," *Phys Rev Lett*, vol. 88, no. 14, p. 148101, 2002.

[35] I. H. Segel, *Enzyme kinetics: Behavior and analysis of rapid equilibrium and steady state enzyme systems.* New York, NY: Wiley, 1st ed., 1975.

[36] C. S. Pittendrigh and P. C. Caldarola, "General homeostasis of the frequency of circadian oscillations," *Proc Natl Acad Sci USA*, vol. 70, pp. 2697–2701, 1973.

[37] F. Hussain, C. Gupta, A. J. Hirning, W. Ott, K. S. Matthews, K. Josic, and M. R. Bennett, "Engineered temperature compensation in a synthetic gene clock," *Proc Natl Acad Sci USA*, vol. 111, pp. 972–977, 2014.

[38] K. Keiler, P. Waller, and R. Sauer, "Role of a peptide-tagging system in degradation of proteins translated from damaged mrna," *Science*, vol. 271, pp. 990–993, 1996.

[39] Y. Kim, R. Burton, B. Burton, and T. Sauer, R.T.Baker, "Dynamics of substrate denaturation and translocation by the clpxp degradation machine.," *Mol Cell*, vol. 5, pp. 639–648, 2000.

[40] T. A. Baker and R. T. Sauer, "Clpxp, an atp-powered unfolding and protein-degradation machine," *Biochim Biophys Acta*, vol. 1, no. 1823, pp. 15–28, 2012.

[41] I. Levchenko, M. Seidel, R. T. Sauer, and T. A. Baker, "A specificity-enhancing factor for the clpxp degradation machine," *Science*, vol. 289, pp. 2354–2356, 2000.

[42] J. M. Flynn, S. B. Neher, a. S. R. T. Kim, Y. I., and T. A. Baker, "Proteomic discovery of cellular substrates of the clpxp protease reveals five classes of clpx-recognition signals," *Mol Cell*, vol. 11, pp. 671–683, 2003.

[43] P. I. Hanson and S. W. Whiteheart, "Aaa+ proteins: have engine, will work," *Nature Reviews Molecular Cell Biology*, vol. 6, pp. 519–529, 2005.

[44] U. Wojtyra, G. Thibault, A. Tuite, and W. Houry, "The n-teriminal zinc binding domain of clpx is a dimerization domain that modulates the chaperone function.," *J Biol Chem*, vol. 278, pp. 48981–48990, 2003.

[45] U. Donaldson, L.W. Wojtrya and W. Houry, "A Solution structure of the dimeric zinc binding domain of the chaperone ClpX," *J Biol Chem.*, vol. 278, pp. 48991–48996, July 2003.

[46] E. Park, B. Lee, S. Hong, and H. Kim, "Structural basis of sspb-tail recognition by the zinc binding domain of clpx.," *J Biol Chem.*, vol. 367, pp. 514–526, 2007.

[47] S. Singh, J. Rozychki, J. Ortega, T. Ishikawa, J. Lo, A. Steven, and M. Maurizi, "Functional domains of the clpa and clpx molecular chaperones identified by limited proteolysis and deletion analysis," *J Biol Chem*, vol. 276, pp. 29420–29429, 2001.

[48] A. Martin, T. Baker, and R. Sauer, "Rebuilt aaa+ motors reveal operating principles for atp-fueled machines," *Nature*, vol. 437, pp. 1115–1120, 2005.

[49] A. Martin, T. Baker, and R. Sauer, "Distinct static and dynamic interactions control atpase-peptidase communication in a aaa+ protease.," *Mol Cell*, vol. 27, pp. 41–52, 2007.

[50] D. Kim and K. Kim, "Crystal structure of clpx molecular chaperone from helicobacter pylori," *J Biol Chem*, vol. 278, pp. 50664–50670, 2003.

[51] R. Grimaud, M. Kessel, F. Beuron, A. Steven, and M. Maurizi, "Enzymatic and structural simularities between the escherichia coli atp-dependent proteases, clpxp and clpap.," *J Biol Chem*, vol. 273, pp. 12476–12481, 1998.

[52] S. Neher, J. Villen, E. Oakes, C. Bakalarski, R. Sauer, S. Gygi, and T. Baker, "Proteomic profiling of clpxp substrates following dna damage reveals extensive instability within sos regulon.," *Mol Cell*, vol. 22, pp. 193–204, 2006.

[53] J. M. Flynn, I. Levchenko, M. Seidel, S. H. Wickner, R. T. Sauer, and T. A. Baker, "Overlapping recognition determinants within the ssra degradation tag allow modulation of proteolysis," *Proc. Natl. Acad. Sci. USA*, vol. 98, no. 19, pp. 10584–10589, 2001.

[54] S. Joshi, G. Hersch, T. Baker, and R. Sauer, "Communication between clpx and clpp during substrate processing and degradation.," *Nat Struct Mol Biol*, vol. 11, pp. 404–411, 2004.

[55] J. Flynn, I. Levchenko, R. Sauer, and T. Baker, "Modulating substrate choice: the sspb adaptor delivers a regulator of the extracytoplasmid-stress response to the aaa+ protease clpxp for degradation.," *Genes Dev.*, vol. 18, pp. 2292–2301, 2004.

[56] D. A. Wah, I. Levchenko, G. E. Rieckhof, D. N. Bolon, T. A. Baker, and R. T. Sauer, "Flexible linkers leash the substrate binding domain of sspb to a peptide module that stabilizes delivery complexes with the aaa+ clpxp protease," *Mol Cell*, vol. 12, pp. 355–363, 2003.

[57] I. Levchenko, R. Grant, D. Wah, R. Sauer, and T. Baker, "Structure of a delivery protein for a aaa+ protease in complex with a peptide degradation tag.," *Mol Cell.*, vol. 12, pp. 365–372, 2003.

[58] H. Song and M. Eck, "Structural basis of degradation signal recognition by sspb, a specificity-enhancing factor for the clpxp proteolytic machine.," *Mol Cell.*, vol. 12, pp. 75–86, 2003.

[59] D. Dougan, I. Levchenko, G. Rieckhof, D. Bolon, T. Baker, and R. Sauer, "Targeted delivery of an ssra-tagged substrate by the adaptor protein sspb to its cognate aaa+ protein clpx.," *Mol Cell.*, vol. 12, pp. 373–380, 2003.

[60] D. Bolon, R. Grant, T. Baker, and R. Sauer, "Nucleotide-dependent substrate handoff from the sspb adaptor to the aaa+ clpxp protease.," *Mol Cell.*, vol. 16, pp. 343–350, 2004.

[61] G. Hersch, T. Baker, and R. Sauer, "Sspb delivery of substrates for clpxp proteolysis probed by the design of improved degradation tags.," *Proc Natl Acad Sci USA.*, vol. 101, pp. 12136–12141, 2004.

[62] R. Burton, S. Siddiqui, Y. Kim, T. Baker, and R. Sauer, "Effects of protein stability and structure on substrate processing by the clpxp unfolding and degradation machine.," *EMBO J*, vol. 20, pp. 3092–3100, 2001.

[63] S. R. Barkow, I. Levchenk, and S. R. T. Baker, T. A., "Polypeptide translocation by the aaa+ clpxp protease machine," *Chem Biol*, vol. 16, pp. 605–612, 2009.

[64] M. Maurizi, W. Clark, Y. Katayama, S. Rudikoff, J. Pumphrey, B. Bowers, and S. Gottesman, "Sequence and structure of clp p, the proteolytic component of

the atp-dependent clp proteases of escherichia coli.," *J Biol Chem.*, vol. 265, pp. 12536–12545, 1990.

[65] M. Maurizi, S. Singh, M. Thompson, M. Kessel, and A. Ginsburg, "Molecular properties of clpap protease of escherichia coli: Atp-dependent association of clpa and clpp.," *Biochemistry*, vol. 37, pp. 7778–7786, 1998.

[66] J. Flynn, I. Levchenko, R. Sauer, and T. Baker, "Modulating substrate choice: the sspb adaptor delivers a regulator of the extracytoplasmid-stress response to the aaa+ protease clpxp for degradation.," *Genes Dev.*, vol. 18, pp. 2292–2301, 2004.

[67] A. Yu and W. Houry, "Clpp: a distinctive family of cylindrical energy-dependent serine proteases.," *FEBS Lett.*, vol. 581, pp. 3749–3757, 2007.

[68] H. Ingvarsson, M. Mate, M. Hogbom, D. Portnoi, N. Benaroudj, P. Alzari, M. Ortiz-Lombardi, and T. Unge, "Insights into the inter-ring plasticity of caseinolytic proteases form the x-ray structure of mycobacterium tuberculosis clpp1.," *Acta Crystallogr D Biol Crystallogr.*, vol. 63, pp. 249–259, 2007.

[69] M. Bewley, V. Graziano, K. Griffin, and J. Flanagan, "The asymmetry in the mature amino-terminus of clpp facilitates a local symmetry match in clpap and clpxp complexes.," *J Struct Biol.*, vol. 153, pp. 16185–16196, 2005.

[70] A. Szyk and M. Maurizi, "Crystal structure at 1.9 a of e. coli clpp with a peptide covalently bound at the active site.," *J Struct Biol.*, vol. 156, pp. 165–174, 2006.

[71] J. Wang, J. Hartling, and J. Flanagan, "The structure of clpp at 2.3 a resolution suggests a model for atp-dependent proteolysis.," *Cell*, vol. 91, pp. 447–456, 1997.

[72] H. Brotz-Oesterhelt, D. Beyer, H. P. Kroll, R. Endermann, C. Ladel, W. Schroeder, B. Hinzen, S. Raddatz, H. Paulsen, K. Henninger, J. E. Bandow, H. G. Sahl, and H. Labischinski, "Dysregulation of bacterial proteolytic machinery by a new class of antibiotics," *Nat Med*, vol. 11, pp. 1082–1087, 2005.

[73] J. Kirstein, A. Hoffman, H. Lilie, R. Schmidt, H. Rubsamen-Waigmann, H. Brotz-Oesterhelt, A. Mogk, and K. Turgay, "The antibiotic adep reprogrammes clpp, switching it from a regulated to an uncontrolled protease," *EMBO Mol Med*, vol. 1, pp. 37–49, 2009.

[74] J. Flynn, I. Levchenko, R. Sauer, and T. Baker, "Activity and specificity of escherichia coli clpap protease in cleaving model peptide substrates.," *J Biol Chem.*, vol. 269, pp. 18201–18208, 1994.

[75] M. Aubin-Tam, A. Olivares, R. Sauer, T. Baker, and M. Lang, "Single-molecule protein unfolding and translocation by an atp-fueled proteolytic machine.," *Cell*, vol. 145, pp. 257–267, 2011.

[76] J. Ortega, H. S. Lee, M. R. Maurizi, and A. C. Steven, "Alternating translocation of protein substrates from both ends of clpxp protease," *EMBO J*, vol. 21, pp. 4938–4949, 2002.

[77] J. M. Jones, D. J. Welty, and H. Nakai, "Versatile action of escherichia coli clpxp as protease or molecular chaperone for bacteriophage mu transposition," *J Biol Chem*, vol. 273, pp. 459–465, 1998.

[78] R. E. Burton, T. A. Baker, and R. T. Sauer, "Energy-dependent degradation: linkage between clpxp-catalyzed nucleotide hydrolysis and protein-substrate processing," *Protein Sci*, vol. 12, pp. 893–902, 2003.

[79] S. K. Singh, R. Grimaud, J. R. Hoskin, W. Wickner, and M. R. Maurizi, "Unfolding and internalization of proteins by the atp-dependent proteases clpxp and clpap," *Proc Natl Acad Sci USA*, vol. 97, pp. 8898–8903, 2000.

[80] G. Hersch, R. Burton, D. Bolon, T. Baker, and R. Sauer, "Asymmetric interactions of atp with the aaa+ clpx6 unfoldase: allosteric control of a protein machine.," *Cell*, vol. 121, pp. 1017–1027, 2005.

[81] J. Kenniston, T. Baker, J. Fernandez, and R. Sauer, "Linkage between atp consumption and mechanical unfolding during the protein processing reactions of an aaa+ degradation machine.," *Cell*, vol. 114, pp. 511–520, 2003.

[82] J. Hasty, D. McMillen, and J. J. Collins, "Engineered gene circuits," *Nature*, vol. 420, pp. 224–230, 2002.

[83] P. Smolen, D. A. Baxter, and J. H. Byrne, "Modeling transcriptional control in gene networks - methods, recent results, and future directions," *Bull Math Biol*, vol. 62, no. 2, pp. 247–292, 2000.

[84] T. Koide, W. L. Pang, and N. S. Baliga, "The role of predictive modeling in rationally re-engineering biological systems," *Nat Rev Microbiol*, vol. 7, pp. 297–305, 2009.

[85] D. Sprinzak and M. B. Elowitz, "Reconstruction of genetic circuits," *Nature*, vol. 438, pp. 443–448, 2005.

[86] Y. T. Zheng and G. Sriram, "Mathematical modeling: Bridging the gap between concept and realization in synthetic biology," *J Biomed Biotech*, pp. 541–609, 2010.

[87] A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins, "Synthetic gene networks that count," *Science*, vol. 324, no. 5931, pp. 1199–1202, 2009.

[88] A. Levskaya, A. A. Chevalier, J. J. Tabor, Z. B. Simpson, L. A. Lavery, M. Levy, E. A. Davidson, A. Scouras, A. D. Ellington, E. M. Marcotte, and C. A. Voigt, "Engineering escherichia coli to see light," *Nature*, vol. 438, pp. 441–442, 2005.

[89] J. J. Tabor, H. M. Salis, Z. B. Simpson, A. A. Chevalier, A. Levskaya, E. M. Marcotte, C. A. Voigt, and A. D. Ellington, "A synthetic genetic edge detection program," *Cell*, vol. 137, pp. 1272–1281, 2009.

[90] K. Montagne, R. Plasson, Y. Sakai, T. Fujii, and Y. Rondelez, "Programming an in vitro dna oscillator using a molecular networking strategy," *Mol Syst Biol*, vol. 7, p. 466, 2011.

[91] K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J. J. Tyson, "Kinetic analysis of a molecular model of the budding yeast cell cycle," *Molec Biol Cell*, vol. 11, pp. 369–391, 2000.

[92] A. Goldbeter, "A model for circadian oscillations in the drosophilla period protein (per)," *Proc Roy Soc Lond B*, vol. 261, pp. 319–324, 1995.

[93] I. Kemler and A. Fontana, "Role of IκBα and IκBβ in the biphasic nuclear translocation of NF-κB in TNFα-stimulated astrocytes and in neuroblastoma cells," *Glia*, vol. 26, pp. 212–220, 1999.

[94] N. A. Monk, "Oscillatory expression of Hes1, p53, and NF-κB driven by transcriptional time delays," *Curr Biol*, vol. 13, no. 16, pp. 1409–1413, 2003.

[95] A. Hoffmann, A. Levchenko, M. L. Scott, and D. Baltimore, "The IκB-NF-κB signaling module: Temporal control and selective gene activation," *Science*, vol. 298, no. 5596, pp. 1241–1245, 2002.

[96] H. De Jong, "Modeling and simulation of genetic regulatory systems: A literature review," *J Comp Biol*, vol. 9, pp. 67–103, 2002.

[97] V. Sevim, X. W. Gong, and J. E. S. Socolar, "Reliability of transcriptional cycles and the yeast cell-cycle oscillator," *PLoS Comp Biol*, vol. 6, no. 7, p. e1000842, 2010.

[98] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J Phys Chem*, vol. 81, pp. 2340–2361, 1977.

[99] C. V. Rao and A. P. Arkin, "Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm," *J Chem Phys*, vol. 118, pp. 4999–5010, 2003.

[100] L. Michaelis and M. L. Menten, "Die kinetik der invertinwirkung," *Biochem Z*, vol. 49, pp. 333–369, 1913.

[101] W. Mather, N. A. Cookson, J. Hasty, L. S. Tsimring, and R. J. Williams, "Correlation resonance generated by coupled enzymatic processing," *Biophys J*, vol. 99, pp. 3172–3181, 2020.

[102] T. B. Kepler and T. C. Elston, "Stochasticity in transcriptional regulation: origins consequences, and mathematical representations," *Biophys J*, vol. 81, pp. 3116–3136, 2001.

[103] R. Bundschuh, F. Hayot, and C. Jayaprakash, "Fluctuations and slow variables in genetic networks," *Biophys J*, vol. 84, pp. 1606–1615, 2003.

[104] I. Stoleriu, F. A. Davidson, and J. L. Liu, "Quasi-steady state assumptions for non-isolated enzyme-catalysed reactions," *J Math Biol*, vol. 48, pp. 82–104, 2004.

[105] M. R. Bennett, D. Volfson, L. Tsimring, and J. Hasty, "Transient dynamics of genetic regulatory networks," *Biophys J*, vol. 92, no. 10, pp. 3501–3512, 2007.

[106] A. V. Hill, "The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves," *J Physiol*, vol. 40, pp. iv–vii, 1910.

[107] B. P. Kramer, C. Fischer, and M. Fussenegger, "Biological gates enable logical transcription control in mammalian cells," *Biotech Bioeng*, vol. 87, pp. 478–484, 2004.

[108] K. S. Matthews, "DNA looping," *Microbiol Rev*, vol. 56, pp. 123–136, 1992.

[109] R. Schleif, "DNA looping," *Annu Rev Biochem*, vol. 61, pp. 199–223, 1992.

[110] L. Cai, C. K. Dalal, and M. B. Elowitz, "Frequency-modulated nuclear localization bursts coordinate gene regulation," *Nature*, vol. 455, pp. 485–490, 2008.

[111] R. B. Lobell and R. F. Schlief, "DNA looping and unlooping by AraC protein," *Science*, vol. 250, pp. 528–532, 1990.

[112] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. New York, NY: Garland Science, 5th ed., 2002.

[113] P. Smolen, D. A. Baxter, and J. H. Byrne, "Effects of macromolecular transport and stochastic fluctuations on dynamics of genetic regulatory systems," *Am J Physiol*, vol. 277, no. 4 Pt 1, pp. C777–C790, 1999.

[114] R. Zhu, A. S. Ribeiro, D. Salahub, and S. A. Kauffman, "Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models," *J Theor Biol*, vol. 246, pp. 725–745, 2007.

[115] R. Zhu and D. Salahub, "Delay stochastic simulation of single-gene expression reveals detailed relationship between protein noise and mean abundance," *FEBS Lett*, vol. 582, pp. 2905–2910, 2008.

[116] W. Mather, M. R. Bennett, J. Hasty, and L. S. Tsimring, "Delay-induced degrade-and-fire oscillations in small genetic circuits," *Phys Rev Lett*, vol. 102, no. 6, p. 068105, 2009.

[117] A. Gronlund, P. Lotstedt, and J. Elf, "Costs and constraints from time-delayed feedback in small gene regulatory motifs," *Proc Natl Acad Sci USA*, vol. 107, pp. 8171–8176, 2010.

[118] K. Josić, W. Ott, J. M. Lopez, L.-J. Shiau, and M. R. Bennett, "Stochastic delay accelerates signaling in gene networks," *PLoS Comp Biol*, vol. 7, p. e1002264, 2011.

[119] G. Bel, M. Munsky, and I. Nemenman, "The simplicity of completion time distributions for common complex biochemical processes," *Phys Biol*, vol. 7, p. 016003, 2010.

[120] G. C. Johnston, R. A. Singer, S. O. Sharrow, and M. L. Slater, "Cell division

in the yeast saccharomyces cerevisiae growing at different rates," *Microbiology*, vol. 118, pp. 479–484, 1980.

[121] N. A. Cookson, S. W. Cookson, L. S. Tsimring, and J. Hasty, "Cell cycle-dependent variations in protein concentration," *Nucl Acids Res*, vol. 38, pp. 2676–2681, 2009.

[122] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty, "Origins of extrinsic variability in eukaryotic gene expression," *Nature*, vol. 439, pp. 861–864, 2006.

[123] D. Huh and J. Paulsson, "Non-genetic heterogeneity from stochastic partitioning at cell division," *Nat Genet*, vol. 43, pp. 95–100, 2011.

[124] S. Gottesman, E. Roche, Y. Zhou, and R. T. Sauer, "The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added to by the SsrA-tagging system," *Genes Dev*, vol. 12, pp. 1138–1347, 1998.

[125] W. W. Wong, T. Y. Tsai, and J. C. Liao, "Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator," *Mol Syst Biol*, vol. 3, p. 130, 2007.

[126] C. Grilly, S. Stricker, W. L. Pang, M. R. Bennett, and J. Hasty, "A synthetic gene network for tuning protein degradation in saccharomyces cerevisiae," *Mol Syst Biol*, vol. 3, p. 127, 2007.

[127] F. J. Isaacs, D. J. Dwyer, and J. J. Collins, "RNA synthetic biology," *Nat Biotech*, vol. 24, pp. 545–554, 2006.

[128] P. S. Swain, M. B. Elowitz, and E. D. Siggia, "Intrinsic and extrinsic contributions to stochasticity in gene expression," *Proc Natl Acad Sci USA*, vol. 99, pp. 12795–12800, 2002.

[129] M. Thattai and A. van Oudenaarden, "Intrinsic noise in gene regulatory networks," *Proc Natl Acad Sci USA*, vol. 98, pp. 8614–8619, 2001.

[130] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, "Noise in eukaryotic gene expression," *Nature*, vol. 422, pp. 633–637, 2003.

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

[132] J. M. Pedraza and A. van Oudenaarden, "Noise propagation in gene networks," *Science*, vol. 307, pp. 1965–1969, 2005.

[133] D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, and M. L. Simpson, "Gene network shaping of inherent noise spectra," *Nature*, vol. 439, pp. 608–611, 2006.

[134] D. Nevozhay, R. M. Adams, K. F. Murphy, K. Josić, and G. Balázsi, "Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression," *Proc Natl Acad Sci USA*, vol. 106, pp. 5123–5128, 2009.

[135] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *J Chem Phys*, vol. 115, pp. 1716–1733, 2001.

[136] X. D. Cai, "Exact stochastic simulation of coupled chemical reactions with delays," *J Chem Phys*, vol. 126, p. 124108, 2007.

[137] S. Peleš, B. Munsky, and M. Khammash, "Reduction and solution of the chemical master equation using time scale separation and finite state projection," *J Chem Phys*, vol. 125, p. 204104, 2006.

[138] T. Tian, K. Burrage, P. M. Burrage, and M. Carletti, "Stochastic delay differential equations for genetic regulatory networks," *J Comp Appl Math*, vol. 20, pp. 696–707, 2007.

[139] Y. Lan, T. C. Elston, and G. A. Papoian, "Elimination of fast variables in chemical Langevin equations," *J Chem Phys*, vol. 129, p. 214115, 2008.

[140] J. E. Ferrell, T. Y. Tsai, and Q. Yang, "Modeling the cell cycle: why do certain circuits oscillate?," *Cell*, vol. 144, pp. 874–885, 2011.

[141] P. Smolen, D. A. Baxter, and J. H. Byrne, "A reduced model clarifies the role of feedback loops and time delays in the drosophila circadian oscillator," *Biophys J*, vol. 83, no. 5, pp. 2349–2359, 2002.

[142] K. Sriram and M. S. Gopinathan, "A two variable delay model for the circadian rhythm of neurospora crassa," *J Theor Biol*, vol. 231, pp. 23–38, 2004.

[143] L. Pujo-Menjouet, S. Bernard, and M. C. Mackey, "Long period oscillations in a G(0) model of hematopoietic stem cells," *SIAM J Appl Dyn Syst*, vol. 4, pp. 312–332, 2005.

[144] L. Chen, R. Wang, T. J. Kobayashi, and K. Aihara, "Dynamics of gene regulatory networks with cell division cycle," *Phys Rev E*, vol. 70, p. 011909, 2004.

[145] O. Buse, R. Pérez, and A. Kuznetsov, "Dynamical properties of the repressilator model," *Phys Rev E*, vol. 81, p. 066206, 2010.

[146] S. Müller, J. Hofbauer, L. Endler, C. Flamm, S. Widder, and P. Schuster, "A generalized model of the repressilator," *J Math Biol*, vol. 53, pp. 905–937, 2006.

[147] N. Strelkowa and M. Barahona, "Transient dynamics around unstable periodic orbits in the generalized repressilator model," *Chaos*, vol. 21, p. 023104, 2011.

[148] K. Kim, D. Lepzelater, and J. Wang, "Single molecule dynamics and statistical fluctuations of gene regulatory networks: A repressilator," *J Chem Phys*, vol. 126, p. 034702, 2007.

[149] A. Loinger and O. Biham, "Stocahstic simulations of the repressilator circuit," *Physical Review E*, vol. 76, p. 051917, 2007.

[150] Y. Mileyko, R. I. Joh, and J. S. Weitz, "Small-scale copy number variation and large-scale changes in gene expression," *Proc Natl Acad Sci USA*, vol. 105, pp. 16659–16664, 2008.

[151] J. Hasty, F. Isaacs, M. Dolnik, D. McMillen, and J. J. Collins, "Designer gene networks: Towards fundamental cellular control," *Chaos*, vol. 11, no. 1, pp. 207–220, 2001.

[152] T. Y. Tsai, Y. S. Choi, W. Ma, J. R. Pomerening, C. Tang, and J. Ferrell, J. E., "Robust, tunable biological oscillations from interlinked positive and negative feedback loops," *Science*, vol. 321, no. 5885, pp. 126–129, 2008.

[153] J. Wang, L. Xu, and E. K. Wang, "Potential landscpe and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations," *Proc Natl Acad Sci USA*, vol. 105, no. 34, pp. 12271–12276, 2008.

[154] M. B. Miller and B. L. Bassler, "Quorum sensing in bacteria," *Annu Rev Microbiol*, vol. 55, pp. 165–199, 2001.

[155] L. You, R. S. Cox III, R. Weiss, and F. H. Arnold, "Programmed population control by cell-cell communication and regulated killing," *Nature*, vol. 428, pp. 868–871, 2004.

[156] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130–1134, 2005.

[157] J. Anderson, E. Clarke, A. Arkin, and C. Voigt, "Environmentally controlled invasion of cancer cells by engineered bacteria," *J Mol Biol*, vol. 355, pp. 619–627, 2006.

[158] A. Tamsir, J. J. Tabor, and C. A. Voigt, "Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'," *Nature*, vol. 469, pp. 212–215, 2011.

[159] D. C. Duffy, J. C. McDonald, J. A. Schueller, and G. M. Whitesides, "Rapid prototyping of microfluidic systems in poly(dimethylsiloxane)," *Anal Chem*, vol. 70, pp. 4974–4964, 1998.

[160] M. R. Bennett and J. Hasty, "Microfluidic devices for measuring gene network dynamics in single cells," *Nat Rev Genet*, vol. 10, pp. 628–638, 2009.

[161] 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," *Mol Syst Biol*, vol. 4, p. 187, 2008.

[162] J. D. Murray, *Mathematical Biology.* New York, NY: Springer, 3rd ed., 2002.

[163] M. Bennett, M. F. Schatz, H. Rockwood, and K. Wiesenfeld, "Huygens's clocks," *Proc Roy Soc Lond A*, vol. 458, pp. 563–579, 2002.

[164] A. T. Winfree, "Biological rhythms and behavior of populations of coupled oscillators," *J Theor Biol*, vol. 16, pp. 15–42, 1967.

[165] S. H. Strogatz, "From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators," *Physica D*, vol. 143, pp. 1–20, 2000.

[166] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences.* Cambridge, 2001.

[167] D. McMillen, N. Kopell, J. Hasty, and J. J. Collins, "Synchronizing genetic relaxation oscillators by intercell signaling," *Proc Natl Acad Sci USA*, vol. 99, no. 2, pp. 679–684, 2002.

[168] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, "Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing," *Proc Natl Sci USA*, vol. 101, pp. 10955–10960, 2004.

[169] A. Wagemakers, J. M. Buldu, J. Garcia-Ojalvo, and M. A. F. Sanjuan, "Synchronization of electronic genetic networks," *Chaos*, vol. 16, no. 1, p. 013127, 2006.

[170] R. Q. Wang, C. G. Li, L. N. Chen, and K. Aihara, "Modeling and analyzing biological oscillations in molecular networks," *Proc IEEE*, vol. 96, no. 8, pp. 1361–1385, 2008.

[171] J. Ackermann, B. Wloztka, and J. S. McCaskill, "In vitro dna-based predator-prey system with oscillatory kinetics," *Bull Math Biol*, vol. 60, pp. 329–353, 1998.

[172] J. Kim, K. S. White, and E. Winfree, "Construction of an in vitro bistable circuit from synthetic transcriptional circuits," *Mol Syst Biol*, vol. 2, p. 68, 2006.

[173] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, "Interdependence of cell growth and gene expression: origins and consequences," *Science*, vol. 330, pp. 1099–1102, 2010.

[174] T. Lu, A. S. Khalil, and J. J. Collins, "Next-generation synthetic gene networks," *Nat Biotechnology*, vol. 27, pp. 1129–1150, 2009.

[175] S. Gottesman, "Proteases and their targets in escherichia coli," *Annu. Rev. Genet.*, vol. 30, pp. 465–506, 1996.

[176] J. Warner, J. Vilardell, and J. H. Sohn, "Economics of ribosome biosynthesis," *Cold Spring Har Symp Quant Biol*, vol. 66, pp. 567–574, 2001.

[177] E. Vilapriny, R. Alves, and A. Sorribas, "Minimization of biosynthetic costs in adaptive gene expression responses of yeast to environmental changes," *PLoS Comput Biol*, vol. 6, pp. 184–1037, 2010.

[178] M. Barton, D. Delneri, S. Oliver, M. Rattray, C. Bergman, and J. Bahlar, "Evolutionary systems biology of amino acid biosynthetic cost in yeast," *PLoS One*, vol. 5, p. ell935, 2010.

[179] J. R. Hoskins, K. Yanagihara, K. Mizuuchi, and S. Wickner, "Clpap and clpxp degrade proteins with tags located in the interior of the primary sequence," *Proc Natl Acad Sci USA*, vol. 99, pp. 11037–11042, 2002.

[180] B. G. Reid, W. A. Fenton, A. L. Horwich, and E. U. Weber-Ban, "Clpa mediates directional translocation of substrate proteins into the clpp protease," *Proc Natl Acad Sci USA*, vol. 98, pp. 3768–3772, 2001.

[181] J. R. Hoskins, S. Y. Kim, and S. Wickner, "Substrate recognition by the clpa chaperone component of clpap protease," *J Biol Chem*, vol. 275, pp. 35361–35367, 2002.

[182] E. Gur, *The Lon AAA Protease*. Dordrecht, Netherlands: Springer Netherlands, 1st ed., 2013.

[183] R. Freudl, G. Braun, N. Honore, and S. T. Cole, "Evolution of the enterobacterial sula gene : a component of the sos system encoding an inhibitor of cell division," *Gene*, vol. 52, pp. 31–40, 1987.

[184] S. Gottesman and M. Gottesman, "Protein degradation in e. coli: The lon mutation and bacteriophage lambda n and cii protein stability," *Nature Reviews Microbiology*, vol. 24, pp. 225–233, 1981.

[185] S. Mizusawa and S. Gottesman, "Protein degradation in escherichia coli : The lon gene controls the stability of sula protein," *Proc Natl Acad Sci USA*, vol. 80, pp. 358–362, 1983.

[186] S. Sonezaki, Y. Ishii, K. Okita, T. Sugino, A. Kondo, and Y. Kato, "Overproduction and purification of sula fusion protein in escherichia coli and its degra-

dation by lon protease in vitro," *Appl Microbiol Biotechnol*, vol. 43, pp. 304–309, 1995.

[187] I. S. Seong, J. Y. Oh, S. J. Yoo, J. H. Seol, and C. H. Chung, "Atp dependent degradation of sula, a cell division inibitor, by the hslvu protease in escherichia coli," *FEBS Lett*, vol. 456, pp. 211–214, 1999.

[188] W. Wu, Y. Zhou, and S. Gottesman, "Redundant in vivo proteolytic activities of escherichia coli lon and the clpyq (hsluv) protease," *J Bacteriol*, vol. 181, pp. 681–3687, 1999.

[189] J. T. Greenberg, P. Monach, J. T. Chou, and B. Demple, "Positive control of a global antioxidant defense regulon activated by superoxide-generating agents in escherichia coli," *Proc Natl Acad Sci USA*, vol. 87, pp. 6181–6185, 1990.

[190] B. Demple, "Redox signaling and gene control in the escherichia coli soxrs oxidative stress regulon, a review," *Gene*, vol. 179, pp. 53–57, 1996.

[191] C. Herman, D. Thevenet, P. Bouloc, G. Walker, and R. DAri, "Degradation of carboxy-terminal-tagged cytoplasmic proteins by the escherichia coli protease hflb (ftsh)," *Gene Dev*, vol. 12, pp. 247–255, 1998.

[192] C. Herman, S. Lecat, R. DAri, and P. Bouloc, "Regulation of the heat-shock response depends on divalent metal ions in an hflb mutant of escherichia coli," *Mol Microbiol*, vol. 18, pp. 1348–1355, 1995.

[193] C. Herman, S. Lecat, R. DAri, and P. Bouloc, "The hflb protease of escherichia coli degrades its inhibitor lambda ciii," *J Bacteriol*, vol. 179, pp. 358–363, 1997.