University of West Bohemia
Faculty of Applied Sciences
Department of Cybernetics

BACHELOR THESIS

Effects of dynamic riboregulators on signal processing in prokaryotic gene expression

Pilsen, 2014 Hynek Kasl

Prohlášení

Předkládám tímto k posouzení a obhajobě bakalářskou práci zpracovanou na závěr studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

V Plzni dne 16. května 2014

....

Declaration

I hereby declare that this bachelor thesis is completely my own work and that I used only the cited sources.

Acknowledgment

I would like to express my sincere gratitude to my advisor M.Sc. Daniel Georgiev, PhD for always being there for me, no matter what.

Abstrakt

Je známe, že nekódující RNA hrají významnou roli v regulaci transkripce a translace v genetických sítích. Jednoduché jednokrokové mechanismy aktivace/inhibice jsou dobře zdokumentované v literatuře. Cílem tohoto projektu je prozkoumat dynamické vlastnosti regulace translace používající opakované hybridizace RNA s oblastí 5'UTR mRNA. Je předpokládáno, že takovýto typ regulace vykazuje dynamiku vhodnou v proofreading mechanismech a podobnou mechanismům kooperativity.

Klíčová slova: regulace translace, matematické modelování, numerická optimalizace, small RNA, syntetická biologie, buněčná kybernetika

Abstract

It is well known that non-coding RNA sequences play an important role in transcriptional and translational regulatory networks. Simple one-step activation/inhibition methods are well documented in literature. The purpose of this project is to explore dynamic properties of translational regulation based on recurrent RNA hybridization with the 5' untranslated mRNA regions. It is hypothesized that RNA-based regulation exhibits dynamics with important proofreading functions and mechanisms of cooperativity.

Keywords: translational regulation, mathematical modelling, numerical optimization, small RNAs, synthetic biology, cell cybernetics

Contents

1 Introduction **4**

2 Biological Background **5**
2.1 Genes and their expression 5
2.1.1 Overview 5
2.1.2 Prokaryotic gene expression 5
2.1.3 mRNA and its structure 6
2.2 Gene regulation 8
2.2.1 Regulation of translation 8

3 Modeling and Simulation **9**
3.1 Modeling techniques 9
3.1.1 Overview 9
3.1.2 Deterministic models 10
3.2 Divider structure 10
3.3 Divider model 12
3.4 Divider Optimization 14
3.4.1 Problem Definition 14
3.4.2 Numerical solution 16
3.4.3 Analytical Analysis 17

4 Experimental Design **26**
4.1 Experimental system 26
4.2 Validation assay 28

5 Discussion **30**
5.1 Biology 30
5.2 Modeling and Simulation 30
5.3 Experiments 30

List Of Terms **31**

List Of Figures **34**

Chapter 1

Introduction

Synthetic biology is a novel field of study unifying the knowledge of classical engineering and classical biology for the shared intent to create new organisms to serve the mankind. It is known that natural organisms employ complex genetic and other regulatory networks that are able to implement many features commonly used in synthetic electromechanical systems, e.g., memory, feedback and noise suppression. Whilst many elements of living cells have been near perfectly characterized, even more remain to be explored.

Regulation of gene expression during translation is a new area of interest for both classical and synthetic biologists. Due to newly discovered experimental techniques, such as mRNA scanning, many types of previously unknown RNAs have been found inside living organisms. Initially thought to be nothing more than ”junk”, these RNAs have been later found to play an important role in gene expression regulation. Most commonly, these RNA hybridize to either of an mRNA's UTRs to block or promote translation or degradation, therefore acting as proportional regulators.

Here, we show that translational regulation by sRNA hybridization can be more than just a proportional regulator. More specifically, a class of RNA structures, named ”dividers”, is presented. These RNA structures, if located inside an UTR of mRNA, utilize sRNA regulation to generate an all-or-nothing response. Such a response is very useful in creating an integrator, the only part that guarantees perfect adaptation.

Chapter 2

Biological Background

In this chapter, we first briefly review the basics of prokaryotic gene expression, and because of its importance for small RNA regulation, the structure of an mRNA molecule. We then explore the possibilities for regulation of gene expression during translation, with emphasis on regulation by small RNAs.

2.1 Genes and their expression

2.1.1 Overview

Genes are the blueprints of life and contain every bit of information an organism needs for its survival. Gene expression, essentially the building process through which the plans - genes are transformed into the products - proteins, is used in all known life. The complexity of this process grows with the evolutionary level of the organism, but even inside the most simple organisms, the process is incredibly complex with many details yet to be explored. Effective control of the expression of each gene is mandatory and throughout the evolution, various control mechanisms have been adopted to ensure that the cell can sustain itself and is able to quickly and efficiently adapt to intracellular and environmental conditions.

2.1.2 Prokaryotic gene expression

In prokaryotes, the genetic information is stored inside the cell's cytoplasm, as opposed to the eukaryotes, where it is engulfed inside the nucleus. The prokaryotic gene expression consists of two important steps. The first is the transcription, during which the genetic information written in the DNA is rewritten, or transcribed, into a single stranded RNA molecule. The transcription is initiated by a class of molecules called sigma-factors, that move fast along the DNA and find genes suitable for transcription. The sigma factor's structure enables an enzyme called RNA polymerase to bind to the

gene's promoter and start transcription. Then the RNA polymerase moves along the DNA, unwinding the double stranded helix structure, reading the nucleotide sequence and using this information to synthesize a new RNA molecule. Once the RNA polymerase reaches a terminator region, that is located at the end of every gene, the synthesized RNA is released and transcription ends. If the synthesized RNA molecule contains information for protein synthesis, it is called a messenger-RNA (mRNA). Transcribed messenger RNAs represent an intermediate step between genes and proteins. They serve as a template in the second step of gene expression, the translation, much like the DNA is the template in the first step. If the synthesized RNA molecule does not contain information for protein synthesis, it is then called a non-coding RNA (ncRNA). Non-coding RNAs perform various other functions inside the cell, e.g., transfer RNAs that bind aminoacids and bring them to ribosomes, ribosomal RNAs that serve as building parts for ribosomes, or small RNAs, that are the subject of this thesis, that serve as activators and inhibitors of translation.

The second step of gene expression is the translation, during which the nucleotide sequence of an mRNA molecule is re-coded, or translated, into a polypeptide chain. In prokaryotes, the translation takes place in the cytoplasm as well as the transcription. The translation is carried out by ribosomes. A ribosome is a very complex structure built from both proteins and RNAs. The translation is initiated by small proteins, called translation initiation factors, or TIFs, that help the ribosome bind to the RBS of an mRNA molecule. If the ribosome succesfully binds, it proceeds to slide along the RNA molecule until it finds a start codon, usually AUG, which is recognized by an initiator tRNA, a molecule that serves as the starting point for the future polypeptide chain. After that, the elongation process begins, during which the mRNA molecule is pulled through the ribosome, while the ribosome reads the coding region of the mRNA a codon at a time, and adds a corresponding amino-acid to the growing polypeptide chain. When a stop codon, a codon that has no corresponding amino acid, is encountered, the polypeptide is finished. The ribosome then undergoes structural changes, which result in the release of the polypeptide chain, which is later folded into an active protein. [1], [2]

2.1.3 mRNA and its structure

An mRNA is a type of RNA molecule that is used as a blueprint for translation. Its primary and secondary structure determine its stability and the rate at which it is translated. It usually contains regulatory elements at the 5' and 3' ends. Every prokaryotic mRNA molecule can be divided into four important regions, which are shown on Figure 2.2 below.

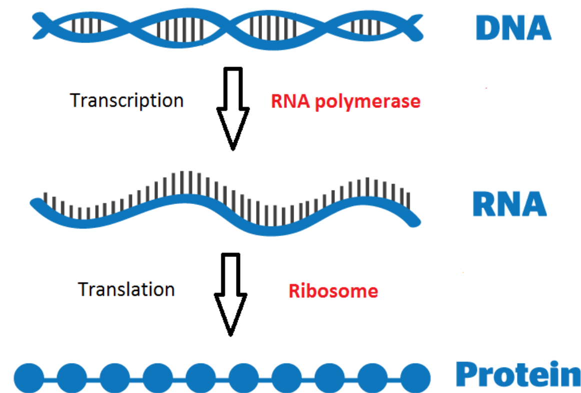


Figure 2.1: Two fundamental steps of prokaryotic gene expression. Source : `http://en.wikipedia.org/wiki/Three_prime_untranslated_region`

Firstly, there is the 5' untranslated region (5'UTR), which starts at the transcription start site and ends one nucleotide before the start codon. This region is important in regulation of degradation and translation of the mRNA molecule. The ribosome binding site (RBS) is usually located close to the 3' end of the 5'UTR. This region may contain binding sites for proteins, regulatory RNAs, and other ligands that affect the stability of the mRNA molecule and the rate at which it is translated. This region's secondary structure has been found to impact the rate of translation.

Secondly, there is the coding region, which starts after the start codon and ends before the stop codon. This region contains the codons, three nucleotide sequences, that encode the final protein's amino acid sequence.

Thirdly, there is the 3' untranslated region (3'UTR), which starts after the stop codon and ends before the poly-A-tail. It plays a crucial role in the regulation, as it contains binding sites for proteins and sRNAs that change the rate of translation and mRNA stability. Similarly to the 5'UTR, 3'UTR secondary structure has an impact on the degradation rate. Hairpin loops that prevent 3' exonucleases are often found at the very end of this region.

Last is the poly-A-tail, located at the 3' end of the mRNA molecule. This region consists of only adenines. In prokaryotes, this tail is usually 15-60 adenines long, and determines the lifespan of the mRNA molecule.

Figure 2.2: The important parts of a prokaryotic mRNA molecule.

2.2 Gene regulation

There are many different levels at which the gene expression can be regulated, differing greatly in the employed biological mechanisms. Among these is the regulation of translation. The regulatory elements are commonly divided into the *cis* and *trans* categories. The *cis* elements are located on the regulated molecule itself, while the *trans* are not.

2.2.1 Regulation of translation

Most of the regulatory elements somehow affect the translation initiation, however the are ways to regulate translation even after the elongation begins. Four structural elements in prokaryotic mRNAs are important in the initiation

1. the ribosome binding site.
2. the secondary structure of the mRNA sequence.
3. the sequences flanking the start codon.
4. the position of the start codon relative to the 5' end of the mRNA.

In this thesis, we focus on the regulation via secondary structures of the UTRs. In many prokaryotic mRNAs, the 5' ends have a considerable secondary structure that acts as a *cis* regulator. Positions and lengths of hairpin loops in the UTR have all been shown to impact the rate of translation in *in vitro* experiments. However, the secondary structure of the UTR can be greatly influenced by *trans* regulatory elements, e.g., proteins and small RNAs. Sequences called riboswitches that can completely change their structure when a ligand binds to them are a perfect example of how structural changes can turn translation on and off.

Small RNA molecules are usually 50-300 nucleotides long. They are most often encoded as separate genes and work in *trans*. These RNAs base pair, imperfectly of perfectly, with their target mRNA's UTRs, and work as either activators or inhibitors. The activators either increase the stability of the mRNA or open hairpin loops near the RBS that prevent TIFs from initiating the translation. The inhibitors decrease the stability of the mRNA, increase the secondary structure size around the RBS or simply block the RBS by base pairing to it.

Note that there also exist regulatory small RNAs that don't regulate at the level of translation. Small RNAs that alternate the protein activity or regulate the translation have also been identified. Experimental techniques development in the recent years has made it clear that RNA regulation is rich in both number and diversity. However, it is not fully understood even in the most studied bacterium E.coli. [3], [4], [5]

Chapter 3

Modeling and Simulation

In this chapter, we introduce basic techniques used for mathematical modeling of chemical reaction systems. We propose the divider structure, that is the center of this thesis. We then use these techniques to develop a mathematical model of a divider. Steady state properties of this model are then explored, first numerically, then analytically. It is shown that the divider can produce an all-or-nothing response, and guidelines for optimal parameter values are given.

3.1 Modeling techniques

3.1.1 Overview

It is desirable to develop mathematical tools that are not only able to explain experimental data and predict their future values, e.g. measured mRNA concentrations or protein fluorescence levels, but that can also be used to design new systems inside living organisms that change their behavior or employ new functions. Because of the complex nature of biological systems, a wide variety of modeling techniques exist, each operating on some level of simplification. As in many other areas of research, these techniques can be divided into three basic categories :

1. Deterministic models
2. Stochastic models
3. Hybrid models

Each of these approaches has it's own advantages and disadvantages, largely differing in complexity, accuracy and model size. A quick overview of the deterministic modeling techniques is given below. Stochastic and hybrid models can be used to study the effects of randomness, that is naturally present in biochemical systems, and its propagation, but are not utilized in this work and therefore broader discussion about them is omitted.

3.1.2 Deterministic models

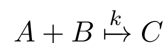
The dynamical properties of a chemical reaction network can be well approximated with a system of ordinary differential equations, if the two following assumptions hold

1. The solution is well mixed, i.e., the concentration of each species is constant w.r.t the position.

2. Number of molecules of each species is sufficiently large.

Note that systems in which the assumption 2) is not correct are better described with stochastic models.

The differential equations describing a chemical reaction system can be obtained, assuming one knows all reactions that occur in the system, using the law of mass action. This law states, that the change of species's concentrations A, B and C in a simple system



$$A + B \overset{k}{\mapsto} C$$

can be described with the following system of ordinary differential equations

$$\frac{dA}{dt} = -kC,$$
$$\frac{dB}{dt} = -kC,$$
$$\frac{dC}{dt} = +kC,$$

where k is a constant reaction rate. This law can be further generalized to be usable for all possible chemical reactions.

3.2 Divider structure

As was discussed in Section 2.2.1, sRNAs usually work as one step inhibitors/activators. These mechanisms are often based on the opening of hairpin loops by strand displacement.

The overview of strand displacement can be seen on Figure 3.1. Small RNA molecule A reacts with hairpin loop B to open it and unveil the domain Y of the molecule. A strand-displacement reaction is initiated with the hybridization of molecule A and the toehold domain of molecule B (step 1.). This allows the free domain of A to branch migrate and open the hairpin (step 2.). Branch migration is the random walk process in which one domain displaces another of identical sequence through a series of reversible single

nucleotide dissociation and hybridization [6]. When the hairpin fully opens, the branch migration is completed, and the domain Y of molecule B is free (step 3.). The rate of a strand displacement reaction is determined by diffusion and the toehold sequence.

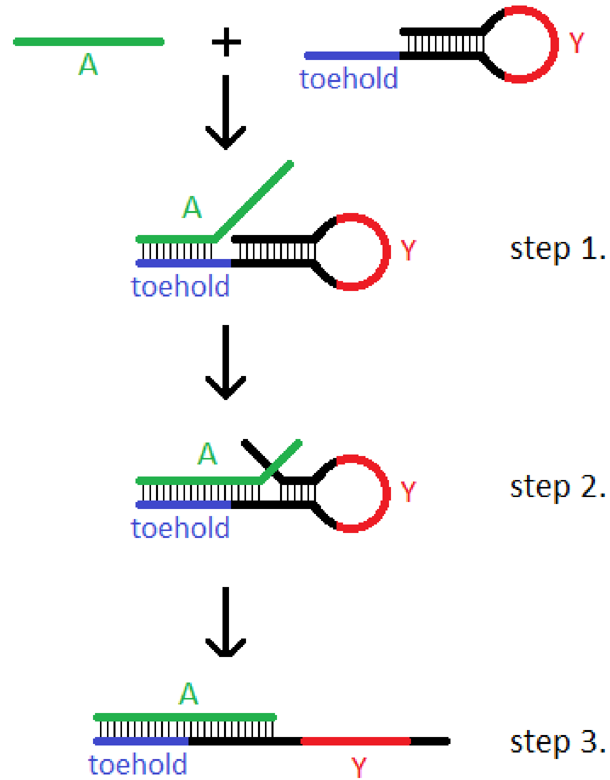


Figure 3.1: An overview of the strand displacement mechanism.

Suppose the domain Y is an RBS. The RBS is then locked inside the hairpin structure, and translation cannot be initiated until the sRNA A opens the hairpin. The sRNA then works as a one step *trans* activator, i.e., only the addition of one sRNA is required to uncover the RBS and re-enable translation.

Suppose we added another hairpin loop and a toehold inside the first hairpin loop, and we put the domain Y inside the second second hairpin loop (see Figure 3.2). Note that the second hairpin loop can only be opened after the first one. Two small RNAs would then be required to fully open the double hairpin and uncover the Y domain, creating a two step *trans* activator.

Similarly, we could insert an arbitrary number of hairpin loops with toeholds to create an activator with arbitrary number of steps. We named

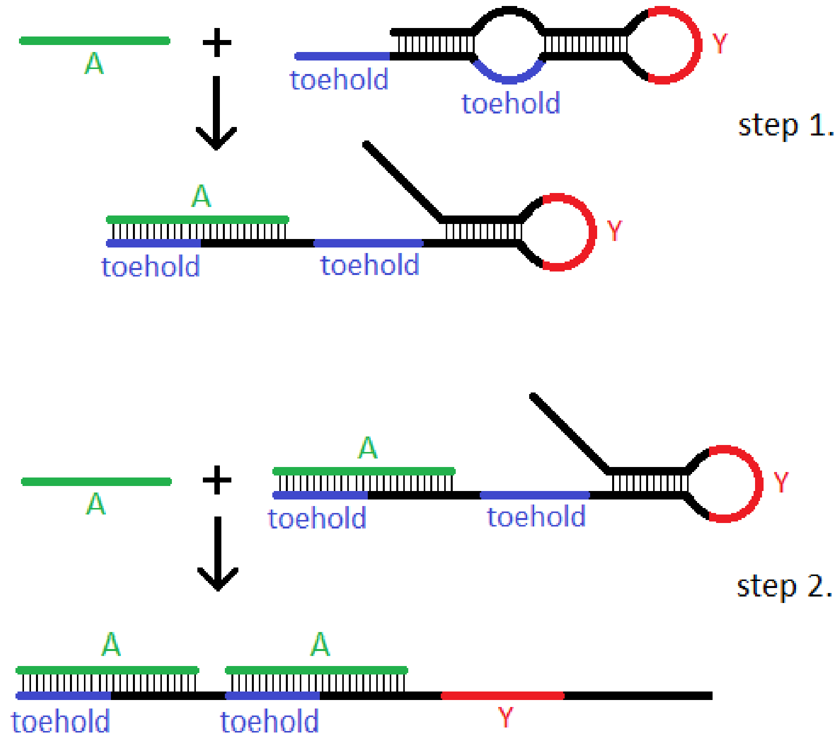


Figure 3.2: Cascaded opening of a second degree divider.

this class of structures dividers, and propose them as means to facilitate dynamical regulation with small RNAs.

Proposition 1 (Divider structure)**.** *An n-th degree divider is a series of n hairpin loops with toeholds, that open consecutively by strand displacement, from the outmost to the inmost, with the addition of n small RNAs. The small RNAs are considered as input and the fully open structure as output.*

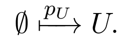
Note that the one-step and two-step activators discussed above are first and second degree divider, respectively. We hypothesized that the dividers can produce an all-or-nothing response. Below, we show that the dividers can indeed produce such a response, and discover guidelines for setting the divider parameters to get a response close to the desired.

3.3 Divider model

When constructing a model of a chemical reaction system using the mass action kinetics (section 3.1.2), one needs to first write down all reactions

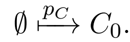
that can occur in the system. For an n-th degree divider (section 3.2), the reactions are

1. Transcription of input sRNA :



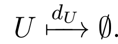
$$\emptyset \xmapsto{p_U} U.$$

2. Transcription of locked mRNA :



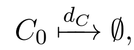
$$\emptyset \xmapsto{p_C} C_0.$$

3. Degradation of input sRNA :

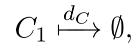


$$U \xmapsto{d_U} \emptyset.$$

4. Degradation of locked mRNA, all partially open mRNAs and fully open mRNA :

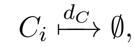


$$C_0 \xmapsto{d_C} \emptyset,$$



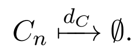
$$C_1 \xmapsto{d_C} \emptyset,$$

$$\vdots$$



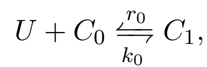
$$C_i \xmapsto{d_C} \emptyset,$$

$$\vdots$$

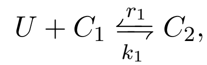


$$C_n \xmapsto{d_C} \emptyset.$$

5. Opening of hairpins by strand displacement :

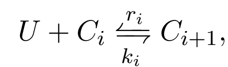


$$U + C_0 \underset{k_0}{\overset{r_0}{\rightleftharpoons}} C_1,$$



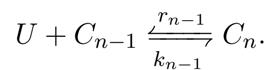
$$U + C_1 \underset{k_1}{\overset{r_1}{\rightleftharpoons}} C_2,$$

$$\vdots$$



$$U + C_i \underset{k_i}{\overset{r_i}{\rightleftharpoons}} C_{i+1},$$

$$\vdots$$



$$U + C_{n-1} \underset{k_{n-1}}{\overset{r_{n-1}}{\rightleftharpoons}} C_n.$$

Using mass actions kinetics (see 3.1.2), this set of reactions can be transformed into a system of $n+2$ nonlinear ordinary differential equations that completely describe the time evolution of species' concentrations.

$$\begin{aligned}
\frac{dU}{dt} &= p_U - d_U U - \sum_{i=0}^{n-1} k_i C_i U + \sum_{i=0}^{n-1} r_i C_{i+1}, \\
\frac{dC_0}{dt} &= p_C - d_C C_0 - k_0 C_0 U + r_0 C_1, \\
&\vdots \\
\frac{dC_i}{dt} &= -d_C C_i + k_{i-1} C_{i-1} U - r_i C_i - k_i C_i U + r_i C_{i+1}, \\
&\vdots \\
\frac{dC_n}{dt} &= -d_C C_n + k_{n-1} C_{n-1} U - r_{n-1} C_1.
\end{aligned}$$

Setting the left sides equal to zero, one can obtain a system of algebraic equations for the steady state concentrations $U_s, C_{0,ss}, C_{1,ss}, \cdots, C_{n,ss}$

$$\begin{aligned}
0 &= p_U - d_U U - \sum_{i=0}^{n-1} k_i C_i U + \sum_{i=0}^{n-1} r_i C_{i+1}, \\
0 &= p_C - d_C C_0 - k_0 C_0 U + r_0 C_1, \\
&\vdots \\
0 &= -d_C C_i + k_{i-1} C_{i-1} U - r_i C_i - k_i C_i U + r_i C_{i+1}, \\
&\vdots \\
0 &= -d_C C_n + k_{n-1} C_{n-1} U - r_{n-1} C_1.
\end{aligned}$$

Note that this is a system of $n+2$ polynomial equations in $n+2$ variables, with up to 2^{n+2} possible roots. However, not all solutions of this system are physically possible, as negative or imaginary concentrations are non-realistic. Though it was not rigorously proven, numerical simulations suggested that the system may have but one acceptable solution for any divider degree. Being interested in the steady state response of $C_{n,ss}$ w.r.t the input build rate, we consider p_U as the input and $C_{n,ss}$ as the output. The system then has $2n+3$ free parameters, which are

$$q = [d_U, p_C, d_C, k_0, k_1, \cdots, k_{n-1}, r_0, r_1, \cdots, r_{n-1}].$$

3.4 Divider Optimization

3.4.1 Problem Definition

The mass-action modeling techniques introduced above are appropriate for the study of the steady state properties of the divider structure (section 3.2).

The divider structure is hypothesized to posses an all-or-nothing steady state response w.r.t the transcription rate of the input sRNA. An all-or-nothing steady state response means that the output steady state is 0 for all input build rates smaller than a critical build rate p_{crit} and for all input build rates greater than p_{crit}, the output steady state is equal to a chosen concentration level K (see Figure 3.3).

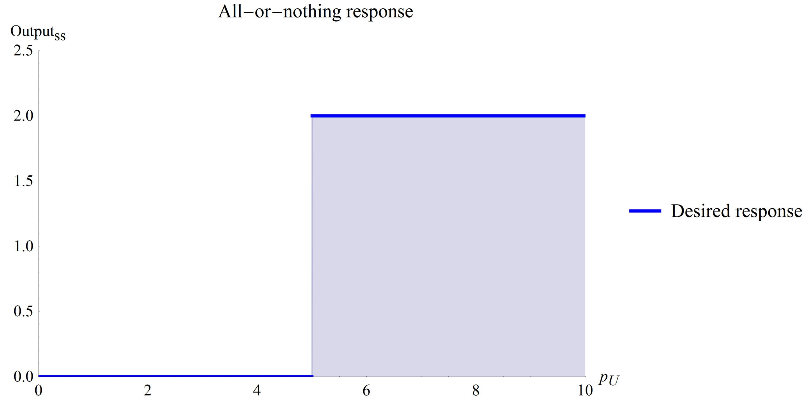


Figure 3.3: An illustration of the desired all-or-nothing steady state response of $C_{n,ss}$ w.r.t the input build rate p_U. Critical input build rate $p_{crit} = 5.0$ and end steady state $K = 2$.

In practice, the actual divider steady state response will never match the desired response completely. It is therefore desired to find the set of parameters q^*, for which the error between the divider's steady response and the desired response is minimal (e.q. w.r.t euclidean distance).

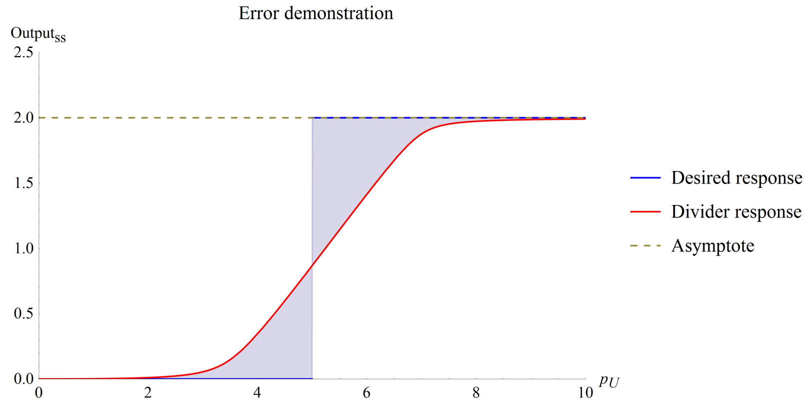


Figure 3.4: An illustration of the difference between the desired steady state response and divider's steady state response. Critical input build rate $p_{crit} = 5.0$ and end steady state $K = 2$.

Mathematically speaking, the desired steady state response $y^*_{ss}(p_U)$ can

be defined with this rule

$$y^*_{n,ss}(p_U, p_{crit}, K) = \begin{cases} 0 & \text{if } p_U \leq p_{crit} \\ K & \text{if } p_U > p_{crit}. \end{cases}$$

Denoting $y_{ss}(p_U, q)$ as the output of an n-th degree divider, the quadratic error $J(q)$ between the $y_{ss}(p_U, q)$ and $y^*_{ss}(p_U, p_{crit}, K)$ is

$$J(q) = \int_0^\infty (y^*(p_U, p_{crit}, K) - y(p_U, q))^2 \mathrm{d}p_U. \tag{3.1}$$

The task at hand is to find the optimal parameters q^* satisfying

$$q^* = \arg\min_q J(q). \tag{3.2}$$

3.4.2 Numerical solution

Initially, we used an algorithm that searches randomly through a part of the parameter space, to verify whether it is possible for a divider to have an all-or-nothing steady state response. After testing 100000 uniformly selected parameter combinations for a chosen p_{crit} and K, parameters yielding all-or-nothing behavior were identified. To explore the parameter space more thoroughly, a local optimization method, similar to gradient optimization algorithms, was performed to optimize each parameter value. This algorithm iterates on the parameter value in a direction in which the criterion $J(q)$ diminishes the most. Because of the non-linearity of the given problem, it is not possible to find this direction precisely, hence it has to be approximated. Such an algorithm does not guarantee global optimality, therefore thousands of runs with random starting points were computed for different divider degrees, critical input build rates and final steady states. All of these runs converged to one point, suggesting that the criterion may have a single local minimum that is also global for all divider degrees.

After using the above mentioned algorithm to find likely optimal parameter vectors for several divider degrees, critical build rates p_{crit} and end steady states K, we observed few basic heuristics that can be used to get the steady state responses close to the desired all-or-nothing response.

Proposition 2 (Divider design rules)**.** *To achieve all-or-nothing steady state behavior using the proposed divider structure, one needs to*

1. *set all forward reaction rates $k_0, k_1, \cdots, k_{n-2}$ to their upper bound.*
2. *keep the backward rates $r_0, r_1, \cdots, r_{n-2}$ as low as possible*
3. *The optimal values of p_C, d_C, k_{n-1} and r_{n-1} depend on the chosen n, p_{crit} and K, however, the ratio of r_{n-1} to k_{n-1} is constant. The optimal value of p_C can be well approximated from*

$$p^*_C = \frac{2p_{crit}}{2n+1}.$$

An example of best obtained fits for chosen $p_{crit} = 2.5$ and $K = 1$ are shown on Figure 3.5. Numerically found values of optimal p_C are linearly dependent on p_{crit} and inversely proportional to $2n+1$, as is shown on Figure 3.6.

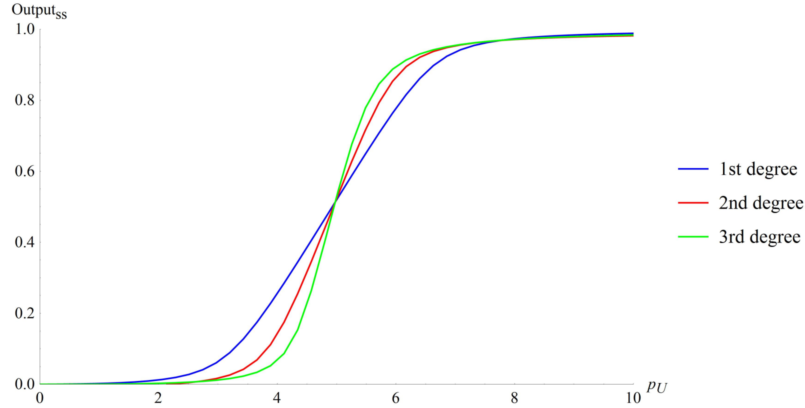


Figure 3.5: Steady state responses of dividers found with the gradient optimization algorithm ($p_{crit} = 2.5$ and $K = 1$).

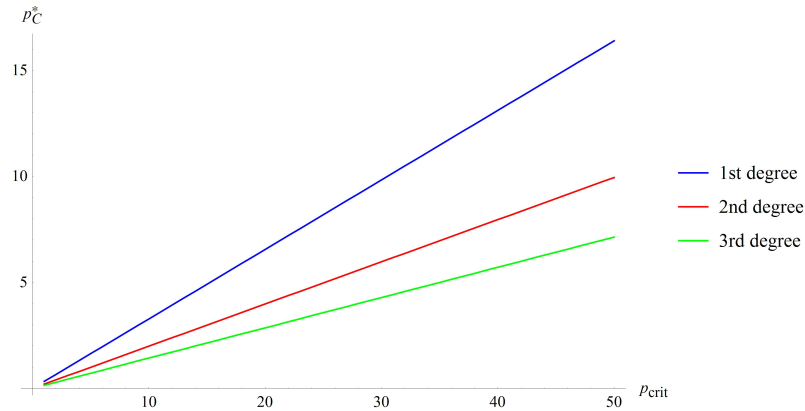


Figure 3.6: Relation between numerically found optimal value of p_C and p_{crit} ($K = 1$ and $d_C = 1$).

3.4.3 Analytical Analysis

Although analytical solutions minimizing the criterion $J(q)$ cannot be found, if certain heuristics are applied, numerically obtained results in Proposition 2 can be explained.

First degree divider

To begin, the simplest divider is considered. A first degree divider consists of a single hairpin loop, that can be fully opened by the addition of just one input sRNA. The opening of a first degree divider can be seen above in Figure 3.1. By using the divider model introduced in Section 3.3, the steady state concentrations are described by the following system of equations

$$0 = p_U - d_U U_{ss} - k_0 C_{0,ss} U_{ss} + r_0 C_{1,ss}, \tag{3.3}$$

$$0 = p_C - d_C C_{0,ss} - k_0 C_{0,ss} U_{ss} + r_0 C_{1,ss}, \tag{3.4}$$

$$0 = -d_C C_{1,ss} + k_0 C_{0,ss} U_{ss} - r_0 C_{1,ss}. \tag{3.5}$$

By substitution, a single equation in $C_{1,ss}$ is derived

$$0 = p_C p_U - (p_C + p_U + d_U \frac{d_C + r_0}{k_0}) d_C C_{1,ss} + d_C^2 C_{1,ss}^2. \tag{3.6}$$

Note that the parameters d_C, k_0, r_0 and d_U appear in the equation only in the linear coefficient of $C_{1,ss}$. Hence we can reduce the parameter space by introducing a new parameter $\beta_0 = \frac{d_C + r_0}{k_0}$. leaving us with the following equation that is subject the of this section's study :

$$0 = p_C p_U - (p_C. + p_U + d_U \beta_0) d_C C_{1,ss} + d_C^2 C_{1,ss}^2 \tag{3.7}$$

For positive values of all parameters, equation 3.7 has two positive real solutions. Of these, only the smaller is physically possible, as U_{ss} and $C_{0,ss}$ corresponding to the larger solution are always negative. Solving Equation 3.7 yields

$$C_{1,ss}(p_U) = \frac{p_C + p_U + d_U \beta_0 - \sqrt{(p_C - p_U)^2 + 2 d_U \beta_0 (p_C + p_U) + d_U^2 \beta_0^2}}{2 d_C}. \tag{3.8}$$

This equation gives us an exact relation between $C_{1,ss}$ and p_U. Note that if $p_U = 0$, that is no input sRNA is being built, then

$$\begin{aligned} C_{1,ss}(0) &= \frac{p_C + d_U \beta_0 - \sqrt{p_C^2 + 2 p_C d_U \beta_0 + d_U^2 \beta_0^2}}{2 d_C} \\ &= \frac{p_C + d_U \beta_0 - \sqrt{(p_C + d_U \beta_0)^2}}{2 d_C} \\ &= 0, \end{aligned} \tag{3.9}$$

i.e., if there is nothing to open the mRNA then no mRNA is opened. The second limiting case considers p_U going to positive infinity. This limit equals

$$\lim_{p_U \to +\infty} C_{1,ss}(p_U) = \frac{p_C}{d_C}, \tag{3.10}$$

i.e., an infinite build rate p_U results in an infinite amount of input sRNA, which results in every locked RNA being open at steady state. The final steady state of fully opened mRNA is then limited only by its build rate p_C and its degradation rate d_C. We now know, that the steady state of open mRNA $C_{1,ss}$ will change from 0 to $\frac{p_C}{d_C}$ as the input build rate p_U increases. Therefore if we wanted to increase the final steady state concentration $C_{1,ss}$, we could increase the transcription rate p_C of locked mRNA or make the mRNA more stable, reducing its degradation rate d_C. The second alternative however also extends the response time of the system, which may be a complication. Computing the slope $\frac{dC_{1,ss}}{dp_U}$ at p_U=0,

$$\left.\frac{dC_{1,ss}}{dp_U}\right|_{p_U=0} = \frac{p_C}{d_C(p_C + d_U\beta_0)} \tag{3.11}$$

reveals that small values of β_0 result in faster growth of $C1, ss$ as p_U increases, and vice versa. An example of the $C_{1,ss}(p_U)$ is shown on Figure 3.7. It can be seen that the first degree divider can not create an all-or-nothing response, only a graded response, which is in agreement with the current thinking that small RNAs serve as proportional regulators.

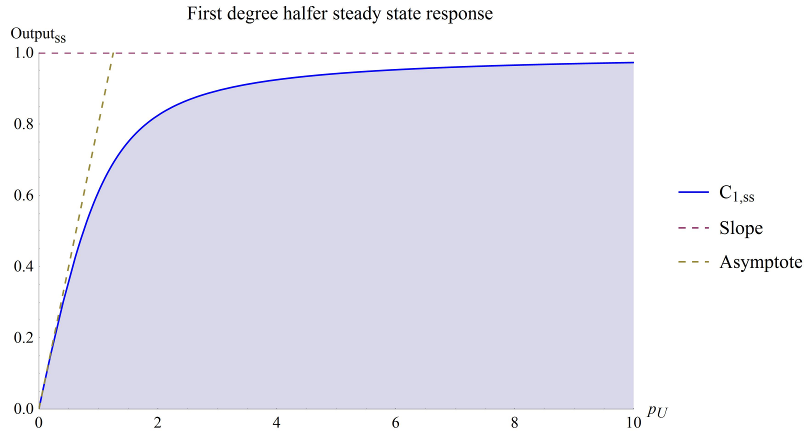


Figure 3.7: Relationship between $C_{1,ss}$ and p_U for 1st degree divider with asymptotes.

Second degree divider

Next, a second degree divider is considered and it is shown that it can produce an all-or-nothing steady state response. A second degree divider consists of two inter-spaced hairpin loops, which open one by one and only the fully opened mRNA can then be translated. The second hairpin loop can only open if the first one is open (see Section 3.2). The opening of a second degree divider can be seen above in Figure 3.2. By using the divider model

introduced in Section 3.3, the steady state concentrations can be described by the following system of equations

$$0 = p_U - d_U U_{ss} - k_0 C_{0,ss} U_{ss} + r_0 C_{1,ss} - k_1 C_{1,ss} U_{ss} + r_1 C_{2,ss} \tag{3.12}$$
$$0 = p_C - d_C C_{0,ss} - k_0 C_{0,ss} U_{ss} + r_0 C_{1,ss} \tag{3.13}$$
$$0 = -d_C C_{1,ss} + k_0 C_{0,ss} U_{ss} - r_0 C_{1,ss} - k_1 C_{1,ss} U_{ss} + r_1 C_{2,ss} \tag{3.14}$$
$$0 = -d_C C_{2,ss} + k_1 C_{1,ss} U_{ss} - r_1 C_{2,ss} \tag{3.15}$$

By substitution and introduction of new parameters $\beta_0 = \frac{d_C + r_0}{k_0}$ and $\beta_1 = \frac{d_C + r_1}{k_1}$, a single equation describing the output steady state concentration $C_{2,ss}$ can be derived

$$\begin{aligned} 0 = &- p_C p_U^2 \beta_0 + \\ &+ (p_C^2 \beta_1 + \beta_0 (p_U + d_U \beta_0)(p_U + d_U \beta_1) + p_C (p_U (3\beta_0 - \beta_1) + 2 d_U \beta_0 \beta_1)) d_C C_{2,ss} + \\ &+ (-2 p_C \beta_0 + p_U (-3\beta_0 + \beta_1) + d_U (-\beta_0^2 - 2\beta_0 \beta_1 + \beta_1^2)) d_C^2 C_{2,ss}^2 + \\ &+ (2\beta_0 - \beta_1) d_C^3 C_{2,ss}^3. \end{aligned} \tag{3.16}$$

If we apply the facts we have learned during the numerical analysis (see Proposition 2), this equation can be further simplified. To achieve the desired steady state behavior, all forward reaction rates $k_0, k_1, \cdots, k_{n-1}$ of all but the last reaction have to be set as high as possible, while keeping the backward reaction rates $r_0, r_1, \cdots, r_{n-1}$ as low as possible. For a second degree divider, this means to set k_0 as high as possible and set r_0 as low as possible. Suppose now for a moment, that we could set the forward reaction rate k_0 infinitely high, i.e., $\beta_0 = 0$. Equation 3.16 then reduces to

$$C_{2,ss}(d_C^2 C_{2,ss}^2 - (p_U + d_U \beta_1) d_C C_{2,ss} + p_C p_U - p_C^2) = 0. \tag{3.17}$$

For $p_U \leq p_C$, only the zero root of Equation 3.17 results in a physically possible steady state. For $p_U > p_C$, only the smaller of the roots of the quadratic part of Equation 3.17 results in a a physically possible steady state. This can be summarized as

$$C_{2,ss}(p_U) = \begin{cases} 0 & \text{if } p_U \leq p_C \\ \frac{d_U \beta_1 + p_U - \sqrt{(d_U \beta_1 + p_U)^2 - 4 p_C p_U + 4 p_C^2}}{2 d_C} & \text{if } p_U > p_C. \end{cases} \tag{3.18}$$

If we introduce a shifted build rate

$$p'_U = p_U - p_C, \tag{3.19}$$

expression 3.18 changes to

$$C_{2,ss}(p_U) = \begin{cases} 0 & \text{if } p'_U \leq 0 \\ \frac{p_C + p'_U + d_U \beta_1 - \sqrt{(p_C - p'_U)^2 + 2 d_U \beta_1 (p_C + p'_U) + d_U^2 \beta_1^2}}{2 d_C} & \text{if } p'_U > 0. \end{cases} \tag{3.20}$$

Note that Equation 3.8 and Equation 3.20 are the same for $\beta_0 = \beta_1$ and $p_U = p'_U$. We can write

$$C_{2,ss}(p_U) = \begin{cases} 0 & \text{if } p_U \leq p_C \\ C_{1,ss}(p'_U)\Big|_{\beta_0=\beta_1} = C_{1,ss}(p_U - p_C)\Big|_{\beta_0=\beta_1} & \text{if } p_U > p_C. \end{cases} \tag{3.21}$$

This means that the steady state response of a second degree divider w.r.t the input build rate is the same as the response of a first degree divider, only shifted by p_C to the right (see Figure 3.8), i.e, the addition of the second lock creates an initial phase, where the output is zero, but otherwise does not change the response.

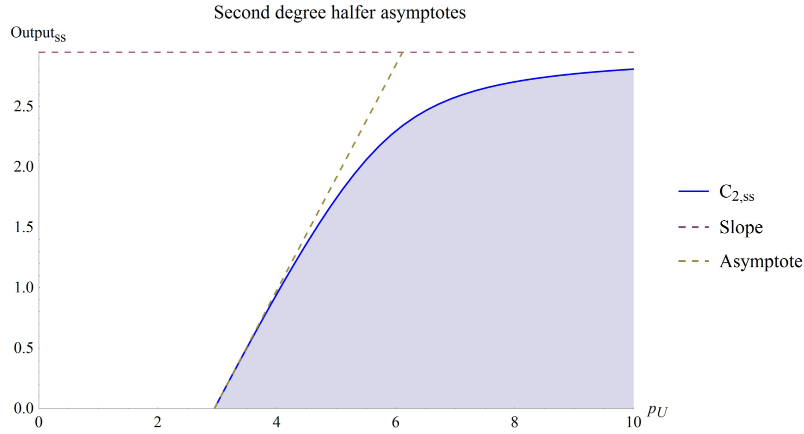


Figure 3.8: A limit case of $C_{2,ss}$ as a function of input build rate for $p_C = 2.95$.

This is in turn essential for creating an all-or-nothing response, because by changing p_C, we can create an arbitrarily long interval of p_U with zero output. A good approximate on the optimal value of p_C for a given p_{crit} can be computed from the point where $C_{2,ss}$ reaches half of its end steady state K.

$$p^*_C = \frac{2p_{crit}}{3}. \tag{3.22}$$

Value of d_C then has to be $\frac{p_C}{K}$ to achieve the desired end steady state. Note that this is in agreement with the numerically obtained result in Proposition 2. In natural conditions the forward reaction rate k_0 cannot be set infinitely high. For sufficiently high k_0, the divider steady state response stays close to the asymptotic response introduced above for infinite k_0. The parameter β_1 has the biggest impact on how the actual response follows the asymptotic. The optimal value of β_1 cannot be expressed analytically, but a good solution can be quickly found numerically. A response of a an optimally set divider

can be seen on Figure 3.9. Below, on Figure 3.10, is a demonstration of the impact of β_1 on the steady state response.

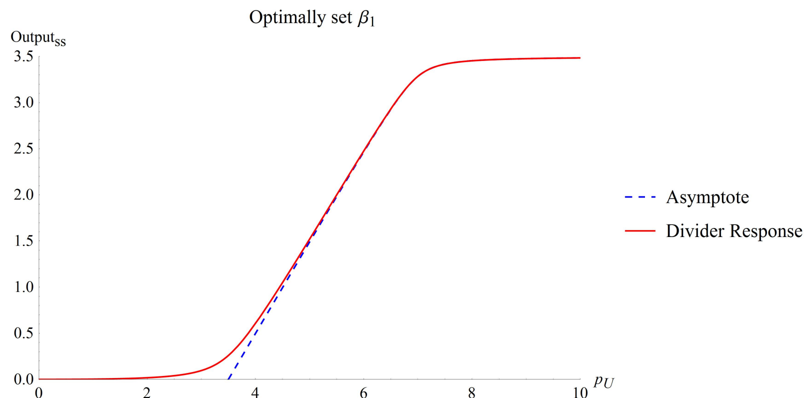


Figure 3.9: A second degree divider's output steady state as a function of input build rate, numerically optimized for $p_{crit} = 5.0$ and $K = 3.5$.

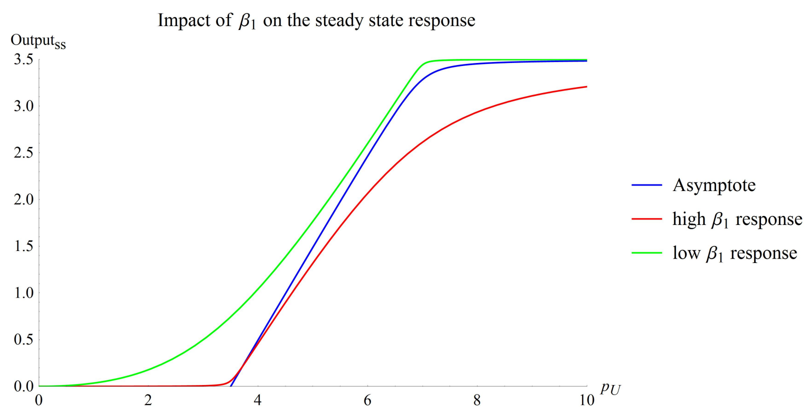


Figure 3.10: Impact of β_1 on $C_{n,ss}(p_U)$, set for $p_{crit} = 5.0$ and $K = 3.5$.

Higher degree dividers

Finally, a divider of degree n is considered. The steady states of an n-th degree divider can be computed from a system of $n+2$ polynomial equations (see Section 3.3). By substitution and introduction of new variables, as in previous sections, this system of polynomial equations reduces to a single polynomial equation in $C_{n,ss}$ of degree $n+1$. Analysis of the roots of this equation in its full form is impossible. However, similarly to Equation 3.16, by setting the forward reaction rates to positive infinity for all but the n-th

reaction, the polynomial equation simplifies to

$$C_{n,ss}^{n-1}(d_C^2 C_{n,ss}^2 - ((2-n)p_C + p_U + d_U\beta_{n-1})d_C C_{n,ss} + p_C p_U - (n-1)p_C^2) = 0. \tag{3.23}$$

Where $\beta_{n-1} = \frac{d_C + r_{n-1}}{k_{n-1}}$. If we introduce a shifted build rate

$$p'_U = p_U - (n-1)p_C, \tag{3.24}$$

Equation 3.23 changes to

$$C_{n,ss}^{n-1}(d_C^2 C_{n,ss}^2 - (p_C + p'_U + d_U\beta_{n-1})d_C C_{n,ss} + p_C p'_U) = 0. \tag{3.25}$$

Similarly to Equation 3.16, for $p_U \leq (n-1)p_C$, i.e, $p'_U < 0$, it can be shown that only the zero root of Equation 3.25 results in a physically possible steady state. For $p_U > (n-1)p_C$, i.e, $p'_U > 0$, only the smaller of the roots of the quadratic part of Equation 3.25 results in a a physically possible steady state. This can be summarized as

$$C_{n,ss}(p'_U) = \begin{cases} 0 & \text{if } p'_U \leq 0 \\ \frac{p_C + p'_U + d_U\beta_{n-1} - \sqrt{(p_C - p'_U)^2 + 2d_U\beta_{n-1}(p_C + p'_U) + d_U^2\beta_{n-1}^2}}{2d_C} & \text{if } p'_U > 0. \end{cases} \tag{3.26}$$

Note that Equation 3.8 and Equation 3.26 are the same for $\beta_0 = \beta_n$ and $p_U = p'_U$. Equation 3.26 can then be rewritten as

$$C_{n,ss}(p_U) = \begin{cases} 0 & \text{if } p_U \leq (n-1)p_C \\ C_{1,ss}(p_U - (n-1)p_C)\Big|_{\beta_0 = \beta_{n-1}} & \text{if } p_U > (n-1)p_C. \end{cases} \tag{3.27}$$

This means that the steady state response of an n-th degree divider w.r.t the input build rate is the same as the response of a first degree divider, only shifted by $(n-1)p_C$ to the right (see Figure 3.11), i.e, the addition of $n-1$ other locks creates a delay phase in the input build rate, where the output is zero, but otherwise does not change the response.

Similarly to the second degree divider, a good approximate on the optimal value of p_C for a given p_{crit} can be computed from the point where $C_{n,ss}$ reaches half of its end steady state K.

$$p_C^* = \frac{2p_{crit}}{2n+1},$$

which is again in agreement with the numerically obtained result in Proposition 2. The optimal value of d_C can then be computed from the condition on the end steady state. The optimal value of β_{n-1} has to be computed numerically and depends on how high we can set the forward reaction rates $k_0, k_1, \cdots, k_{n-2}$.

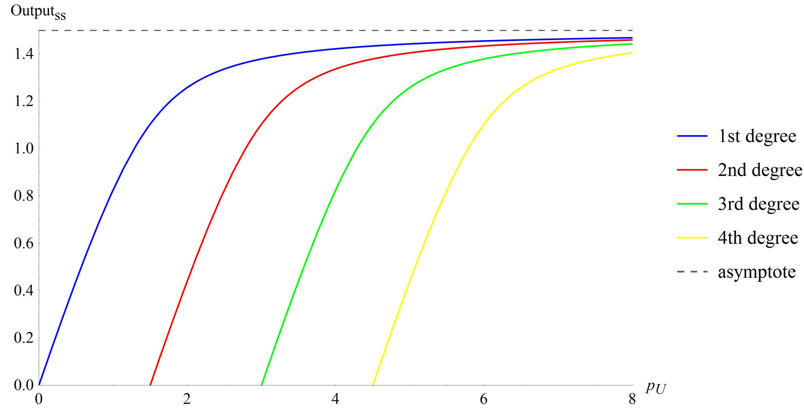


Figure 3.11: A comparative of asymptotic behavior of divider's of degree one to four with the same parameters ($p_C = 1.5$).

It can be shown that as the divider degree n increases, if optimally set, the difference between the divider response and an all-or-nothing response decreases (see Figure 3.12). More specifically, if we approximate the error between the desired and the actual response by a set of triangles, the error is inversely proportional to the divider degree n (see Figure 3.4.3).

$$\min_q J(q) \approx \frac{const.}{n}$$

This implies that in the limit of n going to positive infinity, the divider steady state response will match an all-or-nothing response completely. However, for higher degree dividers, the optimal values of p_C, d_C and β_{n-1} may be out of bounds of physical possibilities.

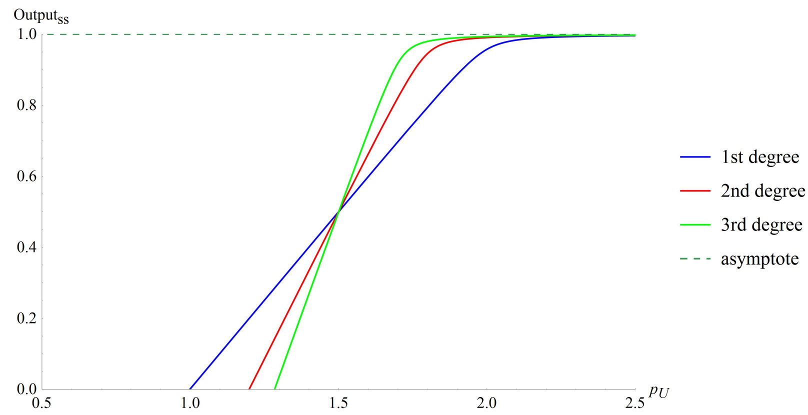


Figure 3.12: A comparative of asymptotic divider output steady states as a function of the input build rate, with each divider's parameters optimally set for $p_{crit} = 1.5$ and $K = 1$.

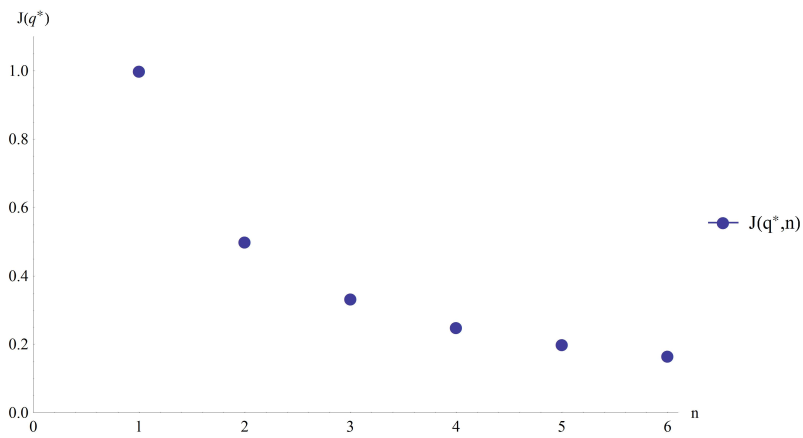


Figure 3.13: Relationship between $J(q^*)$ and the divider degree n.

Chapter 4

Experimental Design

The model presented above is based on many assumptions and simplifications and experimental verification is necessary. We propose a simple, yet sufficient experimental system that can be used to verify the behavior of a first degree divider *in vivo* inside the bacterium E. coli. This system is modular, and can be easily extended to higher degree dividers.

4.1 Experimental system

The experimental system consists of two important parts cloned into a common vector having the ampicillin resistance gene for selection and a bacterial center of origin for replication.

The first part is the input sRNA generator. The purpose of this part is to transcribe at a tunable rate the sRNA that that is the input to the divider described in Chapter [cptrnum]. The part is made up of seven sub-parts.

1. 5' gibson assembly overhang.
2. Promoter. We chose an inducible promoter pBAD (`http://parts.igem.org/Part:BBa_K206000`). This L-arabinose inducible promoter was very well characterized by the British Columbia's iGEM team in 2009. It enables transcriptional regulation of the sRNA level and therefore characterization of the system response to different input build rates.
3. PstI restriction site.
4. sRNA coding sequence. The sequence was designed to open the divider in output's 5'UTR. A stabilizing hairpin that prevents degradation by the 3' exonuclease was added to the 3' end of the sRNA.
5. XbaI restriction site.

6. Double terminator. Binding sites for sequencing primers m13F and m13R were added between the two terminators. Sequencing allows verification of correct nucleotide sequence.

7. overhang for the Gibson assembly.

The whole sequence was manufactured by IDT using their gblock technology.

The second part is a fluorescence protein generator.. Fluorescence measurements are commonly used as means to measure the intracellular conditions. We chose the red fluorescent protein (RFP) as our output. A well characterized RFP generator `http://parts.igem.org/Part:BBa_J04450` was selected. This part has lactose inducible promoter, which adds another level of tunability to the system. Using the site-directed mutagenesis kit, a first degree divider was inserted to the 5'UTR region of this gene. The divider blocks the RBS, preventing the translation. The sRNA base pairs with the toehold region of the divider, and unlocks the RBS by strand displacement, re-enabling the translation of the RFP mRNA. The strand displacement reaction diagram can be seen on Figure 4.2. A restriction site for EcorI was added to the hairpin loop of the divider. This provides us with means to quickly add other hairpins and increase the divider degree.

Note that there are, in total, four enzyme restriction sites recognized by the standard restriction enzymes in the whole system. These sites were added for the purpose of future expansion of the system. Overview of these sites and their respective purposes is given below.

1. SnabI - located at the 5' end of the sRNA gene and at the 3' end of the RFP gene, this site allows quick separation of the vector from the insert. This is necessary for gel verification of correct joining.

2. PstI - located at the beginning of the sRNA sequence, this site allows the addition of additional nucleotides to the 5' end of the sRNA sequence.

3. XbaI - located at the end of the sRNA sequence, this site allows the addition of additional nucleotides to the 3' end of the sRNA sequence.

4. EcorI - located inside the hairpin loop in the 5' UTR of the RFP gene, this site allows the addition of another hairpin loop, increasing the divider degree.

All parts were joined together using the Gibson Assembly method.

Figure 4.1: The subparts of the sRNA gene.

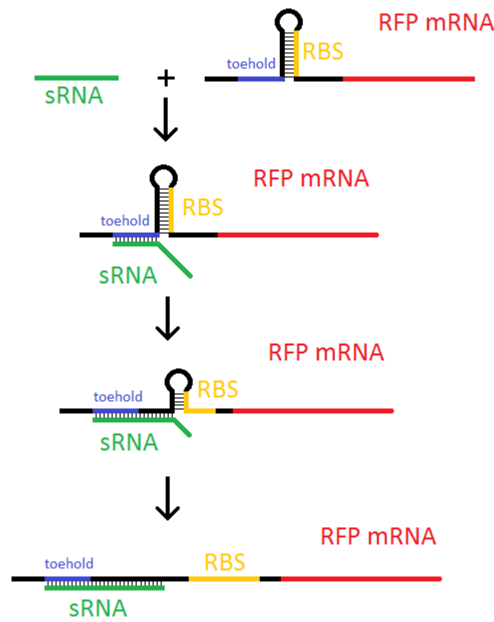


Figure 4.2: Opening of the divider blocking the RBS of RFP by sRNA.

4.2 Validation assay

Following assembly of all parts into a single plasmid, a fluorescence assay of responses for different sRNA transcription rates, determined by arabinose concentrations, will be performed. A second system is constructed a second system for negative control. This second system is an exact copy of the first system, but doesn't have a divider inside the 5'UTR of the RFP gene. Translation of RFP is therefore uninhibited. The control system is used to normalize growth in the fluorescence assays. Normalized fluorescence levels should show a relation similar to a step function w.r.t. the arabinose concentration, i.e., fluorescence levels should be minimal up to a critical arabinose concentration, after which the fluorescence levels should rise quickly to their maximum value. Expected time responses of the experimental system for three different arabinose concentrations can be seen in Figure 4.3. Expected relation between the time-averaged normalized fluorescence levels and arabinose concentration can be seen in Figure 4.4.

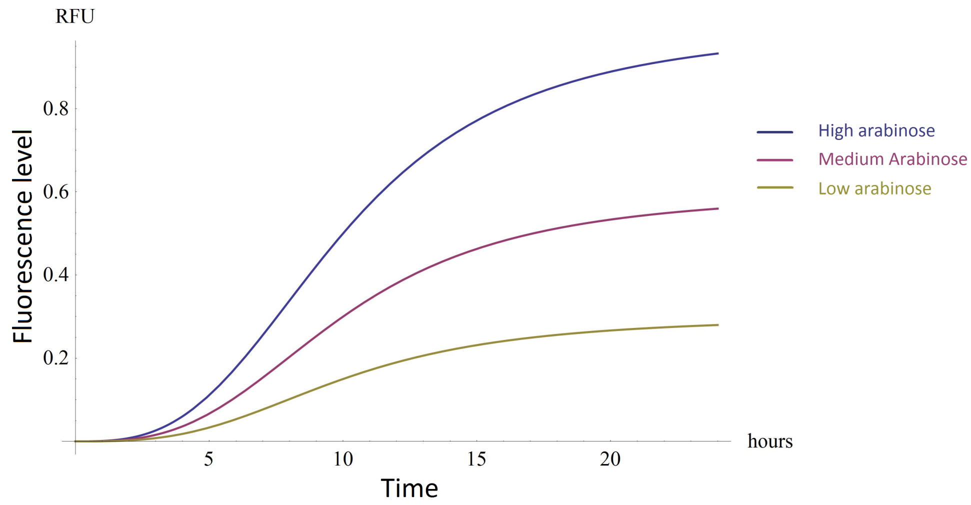


Figure 4.3: Expected time response of the experimental system to three different arabinose concentrations.

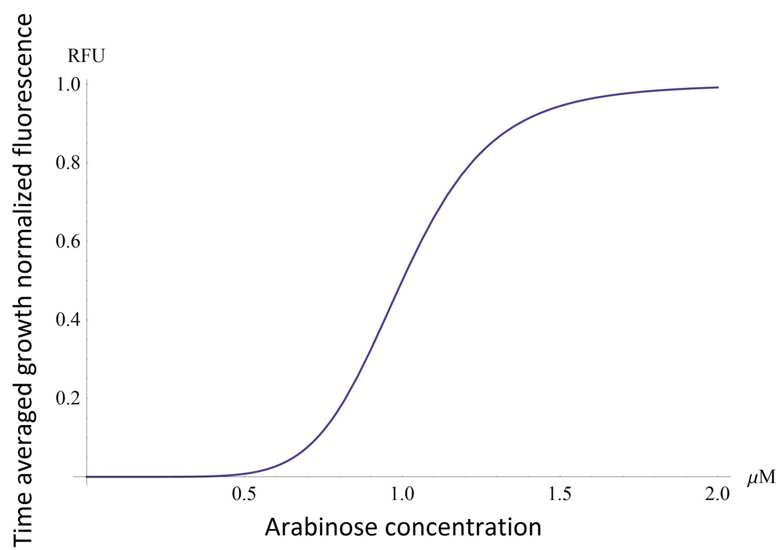


Figure 4.4: Expected relation between the time-averaged growth normalized fluorescence levels and arabinose concentration.

Chapter 5

Discussion

5.1 Biology

While transcriptional regulation has been studied for many years, and a lot of parts are now well characterized, translational regulation by small RNAs is a relatively novel topic. Though much progress has been made, and simple one step regulation is now often used, e.g, for gene silencing, much remains unknown about sRNA regulation in living organisms. The divider structure presented here shows that RNA may implement dynamic regulation, e.g, the all-or-nothing response. Hence the divider structure may be a programmable alternative to natural protein regulation mechanisms that are based on principles of cooperativity.

5.2 Modeling and Simulation

The introduced deterministic model is a first step in designing working dividers. Both numerical and analytical analysis of the model proved that the system is capable of generating the all-or-nothing response. Though this model predicts previously expected results, it is based on many assumptions that must be experimentally validated. To create a more precise model, strand displacement dynamics may also be taken into account. In extension of this work, we would like to create a stochastic model of the divider structure to study its effects on noise propagation. It is hypothesized that it can suppress noise and work as a proofreading mechanism.

5.3 Experiments

Experiments designed to validate the divider concept have been designed and are currently under way. Affirmative experimental results would for the first time illustrate engineered RNA based dynamic regulation.

List Of Terms

Codon A set of three adjacent nucleotides in an mRNA that encodes one amino acid of the translated protein. There are 64 codons, but only 21 amino acids, which makes genetic code redundant and which also enables a way to regulate translation.

DNA A double stranded nucleic acid that contains the genetic information of an organism. DNA is stored differently in eukaryotes and prokaryotes.

Eukaryote An organism whose cells contain a nucleus and other organelles enclosed within membranes.

Gene A segment of DNA that codes for a polypeptide chain or for an RNA that has a function in the organism.

Genome The entire organism's hereditary information, most often encoded in DNA.

Hairpin loop a DNA/RNA structural pattern that occurs when two reversely complement regions, interspaced by a short sequence, hybridize to form a structure similar to a hairpin. **mRNA** A type of RNA molecule that was transcribed from a gene and will later be translated into a polypeptide chain.

Nucleotide A basic building block of nucleic acids,made of a nitrogenous base, a five carbon sugar and atleast one phosphatase group.

Prokaryote An organism whose cells lack a membrane-bound nucleus.

Protein A biological macromolecule consisting of one or more chains of amino acids. Proteins perform a vast array of functions within living organisms including catalytizing reactions,cell signalling and many more.

RBS or ribosomal binding site is a region on a mRNA molecule that allows

ribosome binding and translation of the mRNA molecule. Many different RBS sequences have been identified, with. Note that RBSs exist only in prokaryotes, in eukaryotes 5'-cap and IRESs functionally replace RBSs.

RNA A nucleic acid that performs various functions within the cell. Many different types of RNA molecules have been found in living organisms, e.g. siRNAs, miRNAs,tRNAs and many more. Use of different nucleotides and being single stranded are its key differences from DNA.

RNA primary structure The RNA's linear sequence of nucleotides, most common and

RNA secondary structure The structure of base pairing of RNA's nucleotides, that can be characterized by base pairing probability matrix or, more commonly, in bracket notation. RNA's function may be fully predicted from it's secondary structure and various algorithms exist that can predict RNA's secondary structure *in vitro* from it's primary structure.

RNA tertiary structure The locations of the RNA's atoms in three-dimensional space, not often used for it's heavy computational load and lack of effective tertiary structure prediction algorithms.

UTR An untranslated region of an mRNA, that most commonly embraces the mRNA's coding region from both sides and stores information about the mRNA molecule, such as degradation rate, translation rate and ligand binding rates.

List of Figures

2.1 Two fundamental steps of prokaryotic gene expression. Source : `http://en.wikipedia.org/wiki/Three_prime_untranslated_region` 7
2.2 The important parts of a prokaryotic mRNA molecule. 7

3.1 An overview of the strand displacement mechanism. 11
3.2 Cascaded opening of a second degree divider. 12
3.3 An illustration of the desired all-or-nothing steady state response of $C_{n,ss}$ w.r.t the input build rate p_U. Critical input build rate $p_{crit} = 5.0$ and end steady state $K = 2$. 15
3.4 An illustration of the difference between the desired steady state response and divider's steady state response. Critical input build rate $p_{crit} = 5.0$ and end steady state $K = 2$. . . . 15
3.5 Steady state responses of dividers found with the gradient optimization algorithm ($p_{crit} = 2.5$ and $K = 1$). 17
3.6 Relation between numerically found optimal value of p_C and p_{crit} ($K = 1$ and $d_C = 1$). 17
3.7 Relationship between $C_{1,ss}$ and p_U for 1st degree divider with asymptotes. 19
3.8 A limit case of $C_{2,ss}$ as a function of input build rate for $p_C = 2.95$. 21
3.9 A second degree divider's output steady state as a function of input build rate, numerically optimized for $p_{crit} = 5.0$ and $K = 3.5$. 22
3.10 Impact of β_1 on $C_{n,ss}(p_U)$, set for $p_{crit} = 5.0$ and $K = 3.5$. . . 22
3.11 A comparative of asymptotic behavior of divider's of degree one to four with the same parameters ($p_C = 1.5$). 24
3.12 A comparative of asymptotic divider output steady states as a function of the input build rate, with each divider's parameters optimally set for $p_{crit} = 1.5$ and $K = 1$. 25
3.13 Relationship between $J(q^*)$ and the divider degree n. 25

4.1 The subparts of the sRNA gene. 28

4.2 Opening of the divider blocking the RBS of RFP by sRNA. . 28
4.3 Expected time response of the experimental system to three different arabinose concentrations. 29
4.4 Expected relation between the time-averaged growth normalized fluorescence levels and arabinose concentration. 29

Bibliography

[1] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. Watson, *Molecular Biology of the Cell.* Garland, 4th ed., 2002.

[2] D. Colgan and J. L. Manley, "Mechanism and regulation of mrna polyadenylation," *Genes and Development*, 2007.

[3] M. Kozak, "Regulation of translation via mrna structure in prokaryotes and eukaryotes," *Gene*, vol. 361, no. 0, pp. 13 – 37, 2005.

[4] S. Shabalina and E. Koonin, "Origins and evolution of eukaryotic rna interference," *Trends Ecol Evol.*, 2008.

[5] G. Wilkie, K. Dickson, and N. Gray, "Regulation of mrna translation by 5'- and 3'-utr-binding factors," *TRENDS in Biochemical Sciences*, vol. 28, no. 4, pp. 182 – 188, 2003.

[6] D. Y. Zhang and G. Seelig, "Dynamic dna nanotechnology using strand-displacement reactions," *Nature chemistry*, vol. 3, no. 2, pp. 103–113, 2011.