Analysis, design, and *in vitro* implementation of robust biochemical networks

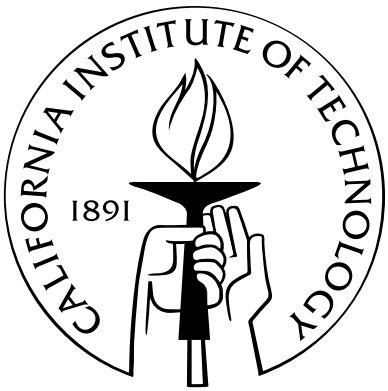
Thesis by

Elisa Franco

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2012

(Defended June 27, 2011)

Alla mia cara zia Giovanna

Acknowledgements

My years at Caltech have been a true learning journey: I am grateful to many, many people for accompanying me through it.

First of all, I want to thank my advisor Richard Murray, who made the journey possible and provided me with invaluable guidance. His enthusiasm, support, and vision turned into exciting challenges many of what I thought were research dead ends or unsurmountable issues. I am very grateful to Erik Winfree, who generously allowed me to become an extended member of his laboratory. His sharp ideas, constructive criticism, and amazing attention to detail have been essential stepping stones for the development of my graduate work. I would also like to thank Paul Rothemund and John Doyle for being part of my thesis committee.

My deepest gratitude goes to Jongmin Kim, for introducing me to *in vitro* synthetic circuits and being a wonderful collaborator and friend for many years. I am also thankful to all the members of the Winfree lab, and in particular to Nadine Dabby, Rebecca Schulman, and Rizal Hariadi for their support, feedback, and friendship.

I am in great debt to my friend Franco Blanchini: he helped turn a casual conference hallway conversation into a good research project. I also want to thank Domitilla Del Vecchio for helpful discussions and advice. I am grateful for having collaborated with John Dabiri at the beginning of my graduate studies.

I was lucky to work with bright and curious undergraduates: Per-Ola Forsberg, with whom I started the flux regulation project, Chris Sturk, and Fei Chen.

All the members of the Murray group have been extraordinary journey companions – literally so, during the fun and adventurous Murray group trips. In particular, I want to thank Mary Dunlop, Julia Braman, Vanessa Jönsson, Ophelia Venturelli, Tim Chung, Pete Trautman, Marcos Nahmad, Dionysios Barmpoutis, and Shuo Han. I am also in debt to Dennice Gayme and Genti Buzi for our "job search" sessions, and to Katie Galloway and Josh Michener for the stimulating bio-control journal club meetings.

For helping with laboratory supplies, traveling, and many other the practical details of my doctorate years, I am grateful to Gloria Bain, Karolyn Knoll, and Anissa Scott.

Without my Italian friends in Los Angeles, adjusting to life in the US would have been a much

harder process: in particular, I would like to thank Daniela Maniezzo, Maria Rita D'Orsogna, Guido Maretto, Roberto Congiu, Martina Carbone, Diego Fazi, Pierpaolo Bergamo, Luca Giacchino, Riccardo Schmid, Andrea Censi, and Luigi Perotti.

Finally, I have to thank my family for their endless support. In particular, my mother Marina, my aunt Lucia, and my husband Mohsen, for their love and patience; and my old friend Tariq for teaching me to be persistent.

Abstract

The functionalities of every living organism are wired in the biochemical interactions among proteins, nucleic acids, and all the other molecules that constitute life's building blocks. Understanding the general design principles of this "hardware of life" is an exciting and challenging task for modern bioengineers. In this thesis, I focus on the topic of molecular network robustness: I investigate several design rules guaranteeing desired functionalities in specific systems, despite their components variability. Experimental verifications of such design schemes are carried out using *in vitro* transcriptional circuits, a minimal analogue of cellular genetic networks.

The first problem I consider is flux control, which is a fundamental feature for the correct performance of biochemical systems. I describe a simple model problem where two reagents bind stoichiometrically to form an output product. In the absence of any regulation, imbalances in the reagent production rates can cause accumulation of unused molecules, and limit the output flow. To match the reagents' flux robustly with respect to the open loop rates, I propose the use of negative or positive feedback schemes that rely on competitive binding. Such schemes are modeled through ordinary differential equations and implemented using transcriptional circuits; data are presented showing the performance of the two approaches.

The second topic I examine is the functional robustness of interconnected networks. Molecular devices characterized in isolation may lose their properties once interconnected. This challenge is illustrated with a case study: a synthetic transcriptional clock is used to time conformational changes in a molecular nanomachine called DNA tweezers. Mass conservation introduces parasitic interactions that perturb the oscillator trajectories proportionally to the total amount of tweezers "load". To overcome this problem, we can use a transcriptional switch that acts as a buffer amplifier, achieving signal propagation and at the same time reducing the perturbations on the source of signal.

Finally, I describe a general class of control-theoretic methods to analyze structural robustness in natural biological systems. Using Lyapunov theory and set invariance, the stability properties of several well-known case studies are analytically demonstrated. The key feature of this analysis is its reliance on parameter-independent models, which only capture essential dynamic interactions between molecular species.

Contents

Chapter 1

Introduction

1.1 Design principles for robust molecular networks

All living organisms, from bacteria to humans, share a remarkable feature: to survive, they must be able to sense external stimuli and implement adequate responses. The ability to effectively control their own behavior based on the "measured" environment is what makes individuals fit and successful. But how do living things make decisions that are crucial to their survival? This question branches out in many directions: from neuro-economics to ethology to molecular biology, several research fields have focused on different aspects of how "control" happens at every layer of what we call life.

At the simplest level, we find that single cells are individually capable of interacting with their surroundings: in this context, decision making and control are embedded in biochemical events. One of the most classical examples is given by the famous experiments of Jacob and Monod [53] in the 1960s, which showed that *E. coli* adapts its gene expression profile to the type of nutrient available. When lactose is abundant, but glucose is not, a set of genes called the *lac* operon is activated through a lactose-dependent cascade of reactions. The proteins expressed from the *lac* operon allow the cells to metabolize lactose and grow. In the absence of lactose, or when glucose is present at high concentrations, the *lac* operon genes are repressed: thus, cells do not waste energy to produce unnecessary lactose-digesting enzymes. This is a clear example of how the control center of a cell is in large part constituted by chemical reactions. It is appropriate to classify a set of molecules that interact and thereby induce specific cellular behaviors as a molecular or biochemical network.

Continuing with our example, the metabolites, genes, and proteins involved in the *lac* operon genetic switch should respond consistently to variations in the available nutrient. However, cells generally differ from each other in size, and therefore in the number and distribution of metabolites and proteins present. Moreover, the intracellular environment is crowded and its content is affected by several parameters, such as temperature and external inputs. Potentially, undesired interactions between the *lac* operon network components and other molecules may slow down or disrupt the

network behavior.

So are there features that confer robustness to a biochemical network? Although evolution operates more as a tinkerer than as an engineer [6], several examples of engineering design principles have been identified in biological systems. For example, negative feedback is the control theorists' favorite tool to confer robustness to a system [12], because it structurally reduces the impact of parametric uncertainty and disturbances. In the biological world, there is evidence that negative auto-regulation in gene expression reduces the variability of protein concentration in cellular populations [16]. Negative feedback has also been related to the response robustness (and speed) of the heat shock response in *E. coli* [31].

A classical example of robust molecular circuitry is probably given by bacterial chemotaxis [14, 9, 117]. The action of the flagellar motor of *E. coli* is driven by a cascade of signaling proteins, whose active or inactive state is determined by the presence of nutrient in the environment. Both analysis on a simplified ordinary differential equation (ODE) model [14] and experiments [9] showed how the *E. coli* flagellar motion presents a robustly stable steady-state: steps in the nutrient concentration only temporarily alter the motor equilibrium. Cells are therefore sensitive to nutrient gradients, but always return to their steady state motion (such property is also referred to as adaptability). Such stable steady state can be described as a function of the concentrations of the signaling cascade protein components and a few binding rates, and is therefore independent of external inputs. Further analysis also demonstrated how integral feedback is present in the chemotaxis network, and guarantees robustness (perfect adaptation) of the equilibrium [134].

Experimental and theoretical studies aimed at unraveling the design principles of existing biological networks generally fall under the category of systems biology. A different approach is represented by synthetic biology [87], which instead focuses on the design of new biological circuits. However, creating new functionalities can be also useful for probing existing systems. On the one hand, for instance, bacteria and yeast have been engineered to become micro-scale factories to produce fuel, anti-malarial drug precursors, insulin, and even silk [69, 100, 131]. On the other hand, we can cite the example of the MAPK pathway synthetic re-wiring, which has been extremely helpful in clarifying the role of several proteins involved in the cascade [17, 15]. Another class of examples is given by the many artificial oscillators synthesized in the past decade [13, 25, 33, 44, 128], which provide insights into the design principles underlying natural cellular clocks and circadian rhythms. Robustness of negative-feedback-loop-based oscillators, for instance, has been experimentally linked to the presence of delays [122], in agreement with classical control theory results [99, 11]. The synthetic approach has also given interesting insights regarding organism-level network robustness: in [52], for instance, it was demonstrated that survival of *E. coli* was not significantly altered by promoter recombinations adding new links across different networks. Some of the re-wired networks actually conferred a fitness advantage under various selection pressures.

It is imperative to characterize and study molecular networks in their own operational context, the cell. However, the complexity of the cellular environment may be an insurmountable obstacle to a detailed understanding of molecular interactions. In fact, quantitative predictions on the dynamic behaviors of *in vivo* molecular networks are limited to small systems, mostly due to the lack of knowledge of the system parameters and to the presence of unmodeled reactions. Synthetic, cell-free biochemical approaches offer a bottom-up, simplified alternative to the study of molecular circuitry.

1.2 Cell-free methods

Operating in an *in vitro* environment with a limited number of biological parts offers several advantages. First, many layers of complexity present *in vivo* may be eliminated, allowing scientists to focus on specific phenomena more quantitatively. Second, fully artificial biological design principles and chemistries can be explored, opening new doors for technology and for understanding the evolution of life.

Cell-free transcription and translation regulatory circuits have been successfully reproduced in [90], with the purpose of achieving a high level of detail (relative to *in vivo* studies) in the investigation of genetic network behaviors. A good example of how *in vitro* assays can reveal new information about natural networks is given by [88], where the reconstruction of circadian oscillations of cyanobacterial KaiC phosphorylation showed that this process is independent of transcription and translation. Recently, a similar *in vitro* set of experiments showed that the dynamics of this oscillator are determined by intermolecular associations: for instance, mutations altering the binding rates of KaiB to KaiC will modulate the oscillator period [96]. A faithful reproduction of *in vitro* cellular environment is still challenging, requiring many components [112] or not-well-characterized extracts [90]. However, transcription-translation kits for cell-free protein production are now commercially available; such kits are particularly useful for the synthesis of unnaturally modified aminoacids [114, 113].

The quest for the minimal biochemistry that supports life [125, 73] is another area where *in vitro* experiments are essential. A related topic of great interest is the role of nucleic acids in general, and of RNA in particular, in the development of life and regulation of gene expression [38, 22].

In vitro synthetic biology and nanotechnology are rapidly evolving [119, 30] in many directions: one relevant trend is the use of nucleic acids for the implementation of natural algorithms and chemical reaction networks. The most attractive property of nucleic acids is programmability [118]: established methodologies are available to reliably predict structure and hybridization pathways of an oligonucleotide molecule, starting from its plain sequence information [141, 80, 2, 28]. If we can predict the structure of a given nucleic acid strand, the ability to design systems of strands that interact according to desired reaction pathways is only a few steps further. Large sets of

nucleic acids have been designed to self-assemble into arbitrary shapes [102, 59]; to create devices moving on programmed paths [74, 137] and performing tasks [47]; and to construct biochemical logic circuitry [110, 140] and molecular machines [139, 24].

The programmability of nucleic acids makes them an ideal candidate for theoretical and experimental studies regarding general chemical reaction networks. In [120], the authors propose motifs for the implementation of arbitrary chemical dynamics with nucleic acids: such dynamics are generated through toehold-mediated branch migration [138, 111], and their speed can be tuned by suitably choosing the length of the toehold domains. (I will return to the topic of branch migration in Section 1.2.1.) Numerical tools for the automated generation of DNA strands implementing a desired reaction network are also available [93].

Although nucleic acid catalytic devices are available, it is interesting to explore the computational and dynamical capabilities of systems integrating proteins and nucleic acids. This is an attractive setup for two main reasons: first, we have a chance to work with molecular network scenarios that may be closer to those of natural networks; second, we can develop useful ground knowledge for the simultaneous programmability of both nucleic acids and amino-acid sequences. Predicting and programming enzyme folding and function is a very active research area [77]: however, custom protein synthesis (with a specified structure and function) is still not possible.

One of the first attempts to construct *in vitro* molecular circuitry using DNA and proteins is the predator-prey system in [4], which consisted of DNA templates and only three proteins: T7 RNA polymerase, M-MLV reverse transcriptase, and RNase H cloned from *E. coli*. The accumulation of sequence mutations is one of the likely reasons for the limited success of those experiments. More recently, logic gates using several enzymes [126] and full metabolic platforms [55] have been characterized. Transcriptional circuits, developed by J. Kim in the Winfree lab at Caltech, are a versatile tool for building molecular networks, and will be described in detail in the next section.

1.2.1 *In vitro* transcriptional circuits

Synthetic *in vitro* genetic transcriptional circuits [61, 63] consist of nucleic acids and two protein species, T7 RNA polymerase (RNAP) and *E. coli* RNase H. Here I will describe their general features, providing the relevant background information for Chapters 2 and 3 of this thesis. Starting with Figure 1.1 A, from now on nucleic acids will be graphically represented as linear strings of letters corresponding to their bases (the helical geometry of double-stranded DNA and RNA will not be shown); the backbone 5'-3' direction will be indicated with an arrow at the 3' end. When appropriate, specific functional areas, or domains, of a nucleic acid strand will be associated with different colors (e.g., domains d1 and d2 in Figure 1.1 A); complementary strands will have the same color (e.g., domains d1 and d1' in Figure 1.1 A).

The fundamental components of a transcriptional network are biochemical switches accepting

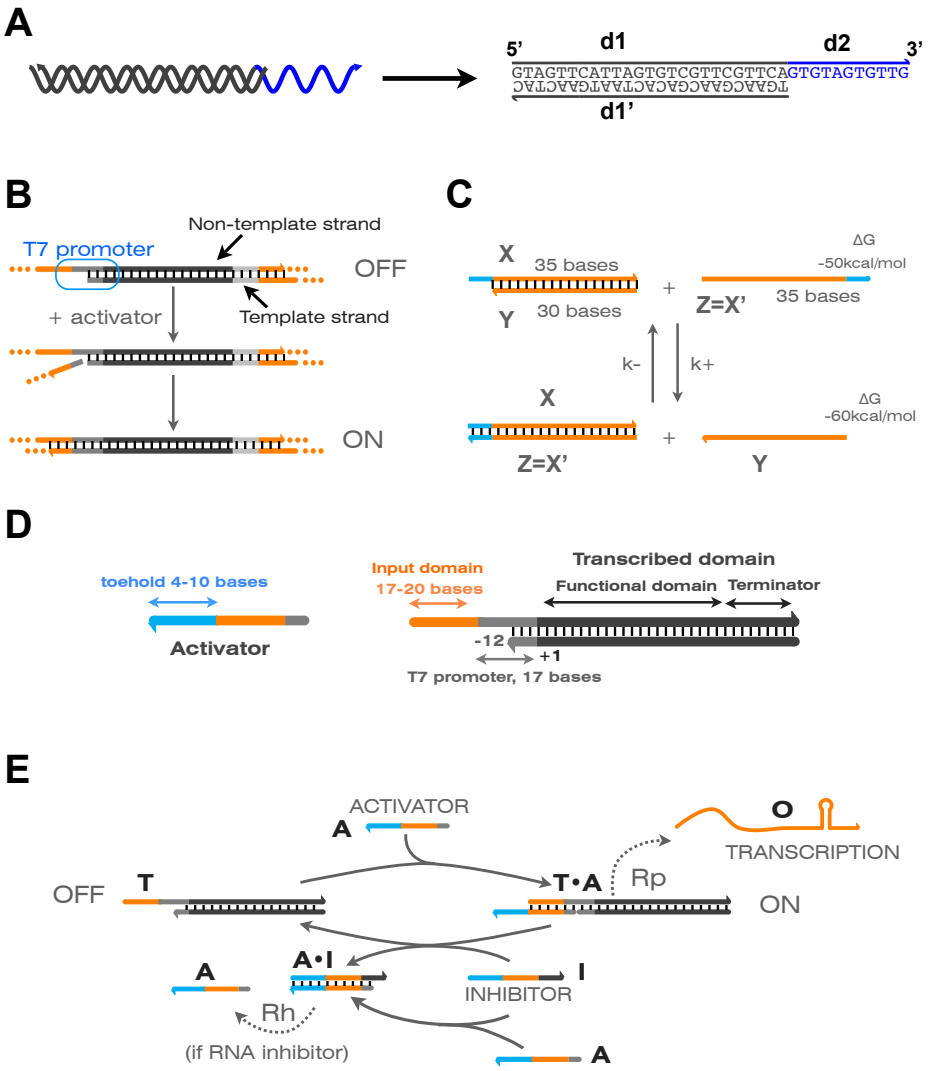


Figure 1.1: A. Double-stranded nucleic acids present the well-known double helical structure. In this thesis, nucleic acids will be represented by strings of letters and graphically sketched as shown on the right. The sugar backbone will be marked with an arrowhead indicating the 5'-3' direction. Contiguous sequences having a specific function will be associated with an identifying color (e.g., d1 vs d2); complementary domains are associated with the same color (e.g., d1 and d1'). B. Transcription of a DNA sequence can be inhibited if the promoter region is partially single-stranded and missing part of the template strand. When the template strand is completed by binding of an activator strand, the binding affinity of this site to RNAP is restored. C. Schematic representation of branch migration reaction. The exposed cyan region in the complex X·Y is called a toehold: the complementary cyan region of strand Z will bind to the toehold and initiate the branch migration process, which is accompanied by an overall decrease in the system's free energy. (The reverse binding rate k^- is negligible.) D. General domains and their lengths for *in vitro* genelets. E. Scheme for resulting reactions in a genelet when an activator and an inhibitor are present. Note that inhibitors can be RNA species. Reactions 1.1 correspond to this figure panel.

one (or more) inputs and generating one (or more) outputs, which can be used to interconnect different switches [61]. Such switches can be implemented as short, linear, synthetic genes whose activity can be turned on and off by altering their promoter region. From now on I will refer to these short artificial genes in transcriptional networks as templates or "genelets", a term originally suggested by Prof. E. Klavins. I will now introduce two notions that are helpful for understanding how the state of such genelets can be systematically switched.

• **Switching promoter activity:** Promoters are double-stranded genetic domains having a high binding affinity for RNA polymerase. The binding affinity can be lost when the structure [57] or sequence [49] of the promoter region is altered, resulting in weaker transcription of the downstream region. A promoter that is partially single-stranded, where the template strand is missing, does not represent a good binding site for RNAP [57]. Referring to Figure 1.1 B, top, if the non-coding strand of the promoter is single-stranded, the genelet can be effectively considered off. The transcription rate of this incomplete promoter is, in general, below 10% of the transcription rate of a fully double-stranded promoter. This residual transcription activity is here called transcription "leak", and we find that it is dependent on the promoter flanking sequences 3.7.13. When a DNA strand complementary to the promoter single-stranded domain is added in solution, the transcription efficiency is recovered and the gene can be considered on. (Data comparing the on and off transcription efficiency of some of the genelets used in this thesis are shown in Section 3.35.) The single-stranded DNA species switching on the genelet will be called an activator. Details regarding the optimal design of the nicked promoter can be found in [61], Section 3.4. So far, only the bacteriophage T7 promoter has been used in transcriptional circuits, due to its high binding affinity and transcription efficiency for the T7 RNA polymerase enzyme, which is commercially available from most biotechnology vendors.

• **Branch migration:** Consider the two nucleic acid complexes shown in Figure 1.1 C, top. One is formed by strands X and Y, the second is a single-stranded species Z, which is fully complementary to X. The complex formed by strands X and Y is partially single-stranded: the blue overhang is an exposed domain, to which the corresponding blue domain of strand Z will initiate binding, subsequently peeling off X from Y. In fact, the system switches quickly to a final, thermodynamically more favorable configuration, where X is bound to its complement Z=X' and Y is released in solution. The blue overhang, where the migration of strands is initiated, is called a toehold. The speed of the reaction is determined by the length of the toehold, as shown in [138] through fluorescence experiments.

The two above notions can be combined: a genelet may be designed to be switched on by an activator strand added in solution, and switched off by branch migration. Branch migration is operated by a single-stranded inhibitor binding to the exposed toehold of an activator/genelet

complex, stripping off the activator strand. General genelet design specifications for the required domains and their lengths, are shown in Figure 1.1 D. The overall mechanism for switching on and off a genelet is depicted in Figure 1.1 E.

Genelets can be interconnected through their RNA outputs by means of an inhibition or activation pathway. The RNA output of a genelet can serve as an inhibitor for a downstream genelet; alternatively, the RNA output can be used to release an activator otherwise sequestered in an activator/inhibitor complex. RNA has the potential to activate a DNA template by binding to the single-stranded activation domain, thereby completing the promoter; however, due to the constraints of our system, this is pathway is not used [82]. Degradation is introduced in the system using the endonuclease RNase H, which targets DNA-RNA hybrids, hydrolyzing the RNA strand and releasing the DNA strand.

The general theoretical foundations for transcriptional circuits were laid out in [62], where the computational capability of these molecular networks is demonstrated to be equivalent to that of neural networks. In general, it is possible to systematically model these circuits using ODEs. (Typically, transcriptional circuits experiments are run at high molecular counts: stochasticity can be safely neglected.) For instance, referring to Figure 1.1 E, consider a genelet T having a DNA activator A, an RNA inhibitor I, and an RNA output O. The chemical reactions expected to occur by design are:

Activation	$\mathrm{T} + \mathrm{A} \overset{k_{TA}}{\rightarrow} \mathrm{T \cdot A}$
Inhibition	$\mathrm{T \cdot A} + \mathrm{I} \overset{k_{TAI}}{\rightarrow} \mathrm{T} + \mathrm{I \cdot A}$
Annihilation	$\mathrm{A} + \mathrm{I} \overset{k_{AI}}{\rightarrow} \mathrm{A \cdot I}$
Transcription: on state	$\mathrm{RNAP} + \mathrm{T \cdot A} \underset{k_{ON}^{-}}{\overset{k_{ON}^{+}}{\rightleftarrows}} \mathrm{RNAP \cdot T \cdot A} \overset{k_{catON}}{\rightarrow} \mathrm{RNAP} + \mathrm{T \cdot A} + \mathrm{O}$ (1.1)
Transcription: off state	$\mathrm{RNAP} + \mathrm{T} \underset{k_{OFF}^{-}}{\overset{k_{OFF}^{+}}{\rightleftarrows}} \mathrm{RNAP \cdot T} \overset{k_{catOFF}}{\rightarrow} \mathrm{RNAP} + \mathrm{T} + \mathrm{O}$
Degradation	$\mathrm{RNaseH} + \mathrm{A \cdot I} \underset{k_{H}^{-}}{\overset{k_{H}^{+}}{\rightleftarrows}} \mathrm{RNaseH \cdot A \cdot I} \overset{k_{catH}}{\rightarrow} \mathrm{RNaseH} + \mathrm{A}.$

(All hybridization reactions are reversible, but the reverse reaction is extremely slow and can be neglected in practice.) The corresponding ODEs can be derived immediately, following the general rules for mass action kinetics. In general, nucleic acid hybridization rates can be measured or estimated from the literature, while enzymatic parameters are more difficult to establish and have a higher variability [61, 63, 64]. (Enzymatic parameter uncertainty will be discussed in particular in Chapter 3.)

The concentrations of activators and inhibitors represent tunable thresholds. Branch migration reactions yielding inhibition, annihilation, or activation are stoichiometric, competitive binding processes. Competitive binding easily generates ultrasensitive responses of the switches [21, 81]: this is an important design feature of transcriptional circuits, and is particularly crucial to achieve oscillatory dynamics.

Several networks have been experimentally characterized using transcriptional circuits: self-inhibiting and self-activating genelets [61], a bistable toggle switch [63], and negative-feedback-based oscillators differing for their topology [64]. In this thesis, I will use this tool kit to construct systems achieving robust properties to be defined later.

1.3 Thesis overview and contribution

Let us go back to our initial question: what are the features that confer robustness to a biochemical network? In this thesis, I will focus on three different topics related to this question. Two chapters include work that follows a "synthetic", bottom-up approach: I will consider specific robust design objectives for biochemical networks, followed by synthesis using transcriptional circuits. The last chapter will instead follow a "systems" approach, reporting more general theoretical robustness results for existing molecular pathways.

• **Chapter 2: Flux control for molecular networks.** Flux control is a fundamental feature for the correct performance of large scale networks, of which familiar examples are the Internet, power grids, or even pipe networks. In the biological world, cells rely as heavily for their survival on a regulated flow of nucleic acids, transcription factors, and other metabolites. It is therefore interesting to explore and understand molecular flow rate control at the molecular level, especially to develop systematic design principles for large biochemical circuits.

In this chapter I will propose two network architectures based on negative and positive feedback, to regulate and match the output flow rate of two interconnected systems. Feedback is implemented through mass action chemical reactions, which down- or up-regulate the activity of the molecules generating the network output. To my knowledge, this design has not been considered elsewhere in the literature. First, negative auto-regulation and positive cross-regulation will be introduced through a very simple, intuitive ODE model. Then, I will describe the implementation of these networks using transcriptional circuits, showing preliminary experimental results. Numerical simulations and data suggest that feedback confers robustness to the system with respect to certain parametric variations and to initial conditions.

The general idea of flux control through positive and negative feedback has been previously presented in two conference papers [40, 42]. I developed the initial design idea and implementation details for

the negative auto-regulation circuit; the first experiments and numerical simulations were carried out by an undergraduate student, Per-Ola Forsberg (SURF program at Caltech). Richard Murray suggested studying the cross-activation scheme. All the analysis, data, and numerical simulations reported in this thesis were performed by me.

• **Chapter 3: Modularity of interconnected systems.** An important research direction in synthetic biology is the systematic design and construction of large molecular networks. Ideally, biological devices should behave modularly, i.e., they should maintain their functionalities (characterized in isolation) when interconnected to other devices. This can be rephrased as a question of robustness: by design, the properties of a system should not be disrupted by the interconnection with other systems. Achieving modularity is a challenge in most engineering fields: classical examples include voltage drops at the output of non-ideal voltage generators, pressure losses in pipe networks and level changes in systems of tanks.

This chapter is dedicated to the experimental study of a molecular oscillator to be used as a clock for a downstream molecular device. Mathematical modeling and experiments show that interconnecting the oscillator to its load in a direct manner, i.e., by stoichiometric binding and release, results in undesired back-action effects and loss of the original signal. Loosely speaking, the back-action is primarily caused by mass conservation constraints. This issue is mitigated by the introduction of a molecular insulator, a node draining a small amount of molecules from the oscillator and using them to amplify its signal [27]. Experiments are carried out using the tool kit of transcriptional circuits.

The project presented in this chapter was developed in close collaboration with the group of Prof. Friedrich Simmel at the Technical University in Munich. F. Simmel and E. Friedrichs had the original idea of using the transcriptional oscillator proposed in [64] to time conformational switching in the well-known molecular tweezers system [139]. Jongmin Kim initially suggested connecting another genelet to the oscillator, using its RNA output to induce switching in the tweezers; this eventually became our insulator design. My contribution was the idea of using this system as a benchmark to study the general challenges of molecular modularity and insulation; such idea was largely inspired by [27] and by several discussions with Prof. Domitilla Del Vecchio. While several experiments I performed were originally designed by the group at TUM, I developed many control experiments to better understand the retroactivity effects and the tweezers behavior. Specific challenges I tackled were data reproducibility, oscillation frequency and amplitude tuning, and the development of a new transcription protocol to avoid the use of commercial kits. In this thesis I will only report experiments performed by me at Caltech, unless explicitly noted in the text or figures. I also contributed the analysis on the simplified model system illustrating the challenge of retroactivity in Section 3.2. Detailed first principles models and parameter fitting were performed by J. Kim and R. Jungmann. In this chapter I will use materials from [41]. The general idea of insulation in the

context of transcriptional circuits was also presented in [39].

• **Chapter 4: Robust properties of natural networks.** As already noted, the molecular circuitry of living organisms performs remarkably robust regulatory tasks, despite the intrinsic variability of its components. A large body of research has in fact highlighted that robustness is often a structural property of biological systems. However, there are few systematic methods to mathematically model and describe structural robustness. With a few exceptions, numerical studies have been the *de facto* standard for this type of investigation.

In this chapter I will highlight how robust stability of equilibria in biological networks can be analyzed using Lyapunov and invariant sets theory. In particular, the analysis is focused on the structure of ODE models rather than on their specific functional expressions. Without resorting to extensive numerical simulations, the stability properties of well-known biological networks will be rigorously proved to be robust. Several case studies will be considered, including the *lac* operon and the mitogen-activated protein kinase (MAPK) pathway.

This project was developed with Prof. Franco Blanchini at the University of Udine. F. Blanchini and I conceived together the general idea of structural analysis of biological models using Lyapunov functions. F. Blanchini mainly focused on the technical results; I contributed the models and assessed the key assumptions and interpretations of the results in a biological context. This chapter reports results from [19].

Chapter 2

Flux control for biochemical networks

2.1 Introduction

Cellular pathways rely heavily on a regulated flux of nucleic acids, transcription factors, and other metabolites. In the era of synthetic biology, it is important to understand and optimize the mechanisms that control and optimize molecular flows. This will contribute to the formulation of systematic design rules for constructing large biochemical networks [92]. (In the following I will use the words flux and flow interchangeably.)

Here, I will consider a simple model problem: given two reagents that bind to form a product, how can we equate their flow through the design of suitable feedback loops? If the two flows are not matched, we could fall in a scenario where (1) the reagent with the higher flux will accumulate, creating a potentially harmful excess of such species and (2) the flow of product will be limited by the lower reagent flux. Two different network design solutions to these problems will be proposed, both based on the use of feedback. A desirable feature of such designs would be their robustness (low sensitivity) with respect to the open loop production rate of the reagents.

The first scheme relies on the use of negative auto-regulation: either species in excess is designed to down-regulate its own production rate. Situation (1) is therefore avoided. The second scheme is based on positive cross-regulation: if one of the reagents is in excess, it will increase the production rate of the second reagent. This second architecture aims at avoiding point (2). The main feature of both these schemes is that feedback is implemented using stoichiometric reactions and without making time-scale separation arguments, which typically yield Michaelis-Menten or Hill functions.

The flux-matching problem and the outlined solutions will first be described with a simple system of ODEs. Then, I will outline how the properties of these feedback schemes can be assessed experimentally using transcriptional circuits. Experiments on the implementation of the negative auto-regulation scheme satisfactorily agree with the numerical predictions, and suggest that flow-

matching is achieved robustly with respect to the open loop rates. On the contrary, the positive cross-regulatory scheme presents several design challenges and the data currently available do not verify the flux matching property conclusively.

2.2 Problem formulation

Consider a simple chemical reaction network

$$\begin{array}{l} \mathrm{T}_1 \overset{\beta_1}{\rightharpoonup} \mathrm{R}_1 + \mathrm{T}_1, \\ \mathrm{T}_2 \overset{\beta_2}{\rightharpoonup} \mathrm{R}_2 + \mathrm{T}_2, \end{array} \qquad \mathrm{R}_1 + \mathrm{R}_2 \overset{\mathrm{k}}{\rightharpoonup} \mathrm{P}. \tag{2.1}$$

Two chemical species T_1 and T_2 produce, respectively, reactants R_1 and R_2, at rates β_1, β_2. The reactants then bind to form an output product P. T_1 and T_2 could be, for instance, two genes whose mRNA or protein outputs R_1 and R_2 must interact stoichiometrically to form a complex useful for a downstream process. A pictorial representation of the network is given in Figure 2.1 A. The differential equation corresponding to the dynamics of R_i is:

$$\frac{\mathrm{d}[\mathrm{R}_\mathrm{i}]}{\mathrm{dt}} = \beta_\mathrm{i} \cdot [\mathrm{T}_\mathrm{i}] - \mathrm{k}\,[\mathrm{R}_\mathrm{i}][\mathrm{R}_\mathrm{j}], \quad \mathrm{i},\mathrm{j} \in \{1,2\}, \mathrm{i} \neq \mathrm{j}. \tag{2.2}$$

The build-up of the product P is clearly conditioned by the rates β_1, β_2 and the concentrations $[\mathrm{T}_1]$ and $[\mathrm{T}_2]$. If the production rates for R_1 and R_2 are significantly different, one can make two observations. First, the reactant produced at the higher rate will accumulate in the system. Second, the lower production rate becomes a bottleneck for the formation of P. For instance, if $[\mathrm{T}_1] \ll [\mathrm{T}_2]$, the concentration of R_2 builds up; at the same time the yield of P is limited by the production rate of R_1. If reactions (2.1) represent a genetic circuit in a cellular host, an excess of R_2 could harm the organism, besides causing a waste of resources. Ideally, biochemical or metabolic networks should include feedback loops able to eliminate excess production of molecules that are not utilized by the system, and increase insufficient production of molecules in high demand. The solution trajectories for equation (2.2) are shown in Figure 2.2. Parameters were chosen as $\beta_\mathrm{i} = 0.01/\mathrm{M}$, $\mathrm{k} = 2 \cdot 10^3/\mathrm{M/s}$, $\mathrm{T}_1 = 100$ nM, $\mathrm{T}_2 = 200$ nM.

In this work, I will consider the model system (2.1) when the production rates for R_i are not balanced. The question that will be asked is: If we could design R_1 and R_2 to interact with the generating species T_1 and T_2, could we achieve self-regulation and matching of the flux rates for the two reactants, robustly with respect to the open loop rates? I will investigate this question by looking at the effects of the feedback loops that can be generated by R_1 and R_2. In particular, I will study the cases where the two reactants self-repress (scheme shown in Figure 2.1 B) or cross-activate

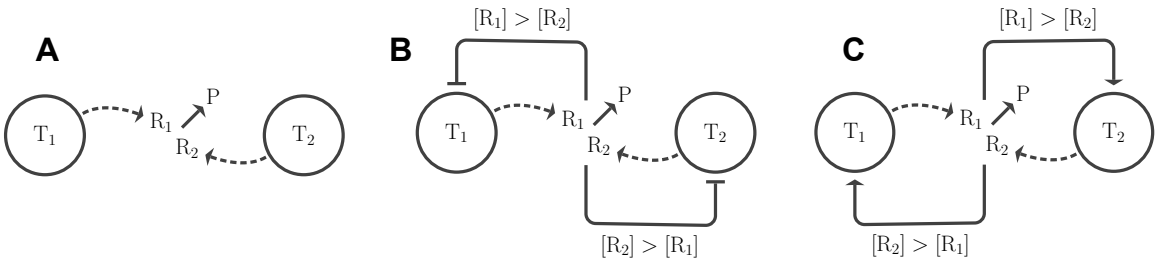


Figure 2.1: A. Schematic representation for our model problem (2.1). B. Negative feedback scheme to control the flow of R_1 and R_2, corresponding to equations (2.3). The comparison between the concentrations of R_1 and R_2 is implicit, due to the formation of the product P. C. Positive feedback interconnection to control the flow of R_1 and R_2, corresponding to equations (2.7).

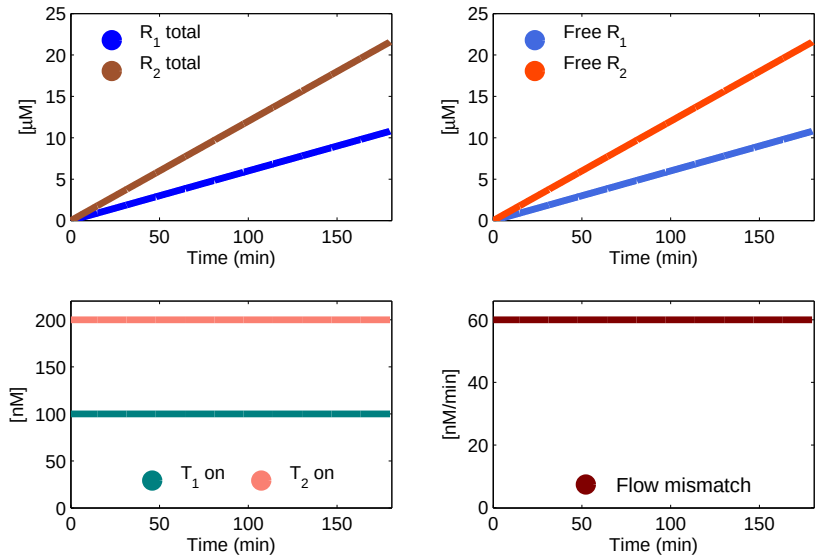


Figure 2.2: Numerical solution to the differential equations (2.2). Bottom right: absolute value of the flux mismatch between the total amount of species R_1^{tot} and R_2^{tot}.

(Figure 2.1 C). I will assume that the feedback occurs by mass action chemical reactions.

2.2.1 Self-repression

Free molecules of R_i, $i = 1, 2$, bind to active T_i thereby inactivating it:

$$R_i + T_i \overset{\delta_i}{\rightharpoonup} T_i^*,$$

$$T_i^* \overset{\alpha_i}{\rightharpoonup} T_i,$$

where T_i^* is an inactive complex. We assume that $T_i^{tot} = T_i + T_i^*$, and that T_i^* naturally reverts to its active state with a first-order rate α_i. The total amount of R_i is $[R_i^{tot}] = [R_i] + [T_i^*] + [P]$. A pictorial representation of this feedback interconnection is shown in Figure 2.1 B. The corresponding

differential equations are:

$$\frac{\mathrm{d}[\mathrm{T_i}]}{\mathrm{dt}} = \alpha_\mathrm{i}\left([\mathrm{T_i^{tot}}] - [\mathrm{T_i}]\right) - \delta_\mathrm{i}\,[\mathrm{R_i}][\mathrm{T_i}],$$
$$\frac{\mathrm{d}[\mathrm{R_i}]}{\mathrm{dt}} = \beta_\mathrm{i}\,[\mathrm{T_i}] - \mathrm{k}\,[\mathrm{R_i}][\mathrm{R_j}] - \delta_\mathrm{i}\,[\mathrm{R_i}][\mathrm{T_i}]. \tag{2.3}$$

For illustrative purposes, the above differential equations are solved numerically. The parameters chosen are: $\alpha_1 = \alpha_2 = 3 \cdot 10^{-4}$ /s, $\beta_1 = \beta_2 = 0.01$ /s, $\delta_1 = \delta_2 = 5 \cdot 10^2$ /M/s, and $\mathrm{k} = 2 \cdot 10^3$/M/s. An imbalance in the production rates of $\mathrm{R_1}$ and $\mathrm{R_2}$ is created by setting $[\mathrm{T_1}](0) = [\mathrm{T_1^{tot}}] = 100$ nM and $[\mathrm{T_2}](0) = [\mathrm{T_2^{tot}}] = 200$ nM, while $[\mathrm{R_1}](0) = [\mathrm{R_2}](0) = 0$. The overall result of this feedback interconnection is that the mismatch in the flow rate of $\mathrm{R_1}$ and $\mathrm{R_2}$ is reduced, as shown in Figure 2.3. The flow rate is defined as the derivative of the total amount of $[\mathrm{R_i^{tot}}]$. The flow rate mismatch is defined as the absolute value of the difference between the two fluxes. The effect of changing the feedback strength, for simplicity chosen as $\delta_1 = \delta_2$, is shown in Figure 2.4: the figure shows the mean active fraction of $[\mathrm{T_i}]$ and the mean flow mismatch over a trajectory simulated for 10 hours. The mean is shown, rather than steady-state values, to capture the behavior of the system over the whole trajectory.

It is possible to examine the nullclines relating $\mathrm{T_1}$ and $\bar{\mathrm{T}}_2$, and find the equilibria $\bar{\mathrm{T}}_1$ and $\bar{\mathrm{T}}_2$ as intersection of these nullclines:

$$\dot{\mathrm{T}}_\mathrm{i} = 0 \quad \Longrightarrow \quad \mathrm{R_i} = \frac{\alpha_\mathrm{i}(\mathrm{T_i^{tot}} - \mathrm{T_i})}{\delta_\mathrm{i}\mathrm{T_i}},$$
$$\dot{\mathrm{R}}_\mathrm{i} = 0 \quad \Longrightarrow \quad \mathrm{R_i} = \frac{\beta_\mathrm{i}\mathrm{T_i}}{\mathrm{kR_j} + \delta_\mathrm{i}\mathrm{T_i}}.$$

To simplify the derivation, we set $\delta_1 = \delta_2 = \delta$, $\beta_1 = \beta_2 = \beta$, $\alpha_1 = \alpha_2 = \alpha$. Equating the two expressions for R_i, we get the following equations (for $i = 1, 2$ and $j = 1, 2$):

$$\left(\frac{\alpha}{\delta}\right)^2 \mathrm{k}\left(\frac{\mathrm{T_i^{tot}} - \mathrm{T_i}}{\mathrm{T_i}}\right)\left(\frac{\mathrm{T_j^{tot}} - \mathrm{T_j}}{\mathrm{T_j}}\right) + \alpha(\mathrm{T_i^{tot}} - \mathrm{T_i}) - \beta\mathrm{T_i} = 0.$$

We can find an expression of the nullclines by introducing a change of variables $\mathrm{u} = \left(\frac{\mathrm{T_1^{tot}} - \mathrm{T_1}}{\mathrm{T_1}}\right)$ and $\mathrm{v} = \left(\frac{\mathrm{T_2^{tot}} - \mathrm{T_2}}{\mathrm{T_2}}\right)$, and defining $\phi_1 = \psi_1 = \left(\frac{\alpha}{\delta}\right)^2 \mathrm{k}$, $\phi_2 = \alpha\mathrm{T_1^{tot}}$, $\psi_2 = \alpha\mathrm{T_2^{tot}}$, $\phi_3 = \beta\mathrm{T_2^{tot}}$, and finally $\psi_3 = \beta\mathrm{T_1^{tot}}$:

$$\mathrm{u}^2(\phi_1\mathrm{v}) + \mathrm{u}(\phi_1\mathrm{v} + \phi_2 - \phi_3\frac{1}{1+\mathrm{v}}) - \phi_3\frac{1}{1+\mathrm{v}} = 0, \tag{2.4}$$
$$\mathrm{v}^2(\psi_1\mathrm{u}) + \mathrm{v}(\phi_1\mathrm{u} + \psi_2 - \psi_3\frac{1}{1+\mathrm{u}}) - \psi_3\frac{1}{1+\mathrm{u}} = 0. \tag{2.5}$$

The roots of the equations above represent the nullclines of the system. Because all the parameters in these equations are positive, there is always a single root. The nullclines are numerically solved, for varying δ, in Figure 2.5.

A condition for flow matching at steady-state can be derived as follows:

$$\dot{\mathrm{R}}_1 - \dot{\mathrm{R}}_2 = 0,$$
$$\beta_1 \mathrm{T}_1 - \delta_1 \mathrm{T}_1 \mathrm{R}_1 = \beta_2 \mathrm{T}_2 - \delta_2 \mathrm{T}_2 \mathrm{R}_2.$$

Substituting the expressions for R_1 and R_2 that can be derived by setting $\dot{\mathrm{T}}_1 = 0 = \dot{\mathrm{T}}_2$, we get:

$$\beta_1 \bar{\mathrm{T}}_1 - \alpha_1(\mathrm{T}_1^{\mathrm{tot}} - \bar{\mathrm{T}}_1) = \beta_2 \bar{\mathrm{T}}_2 - \alpha_2(\mathrm{T}_2^{\mathrm{tot}} - \bar{\mathrm{T}}_2).$$

Taking $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$ we get:

$$\bar{\mathrm{T}}_2 = \bar{\mathrm{T}}_1 + \frac{\alpha}{\alpha + \beta}(\mathrm{T}_2^{\mathrm{tot}} - \mathrm{T}_1^{\mathrm{tot}}). \tag{2.6}$$

The flow matching condition above is shown in Figure 2.5, red dashed line. If $\beta \gg \alpha$, i.e., the production of $\mathrm{R_i}$ is much faster than the generating species $\mathrm{T_i}$ inactivation rate, then the condition can be rewritten as:

$$\bar{T}_1 \approx \bar{T}_2.$$

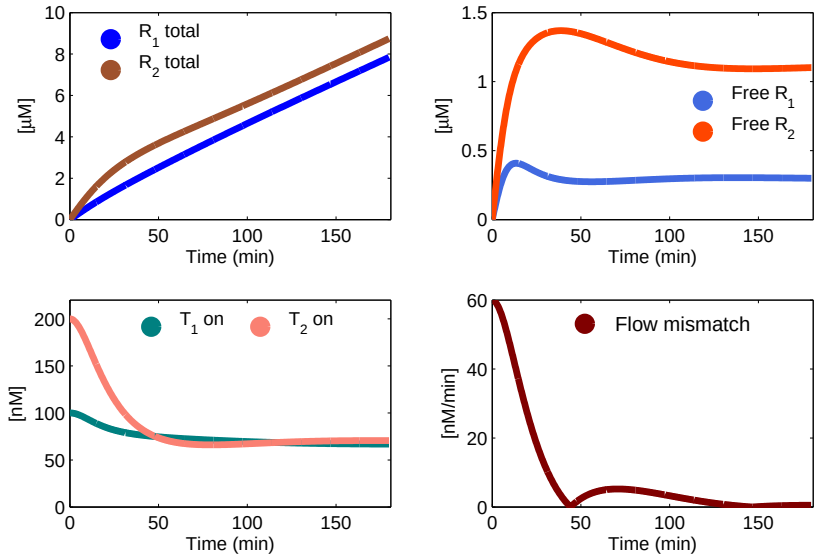


Figure 2.3: Numerical simulation showing the solution to the negative feedback architecture (Figure 2.1 B) modeled with equations (2.3). The flow mismatch between R_1 and R_2 is shown in the bottom-right panel.

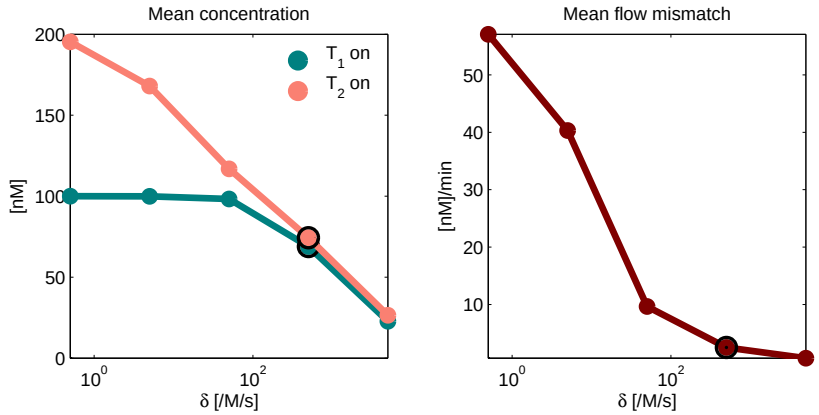


Figure 2.4: Numerical simulation for the negative feedback scheme (2.3), showing the mean concentration of active generating species T_1 and T_2 and the mean flow mismatch as a function of the feedback parameter rate δ. The points corresponding to the set of nominal parameters (trajectories in Figure 2.6) are circled in black. The mean is taken over a trajectory of 10 hours. The stronger the negative feedback, the less R_1 and R_2 are produced by the two subsystems.

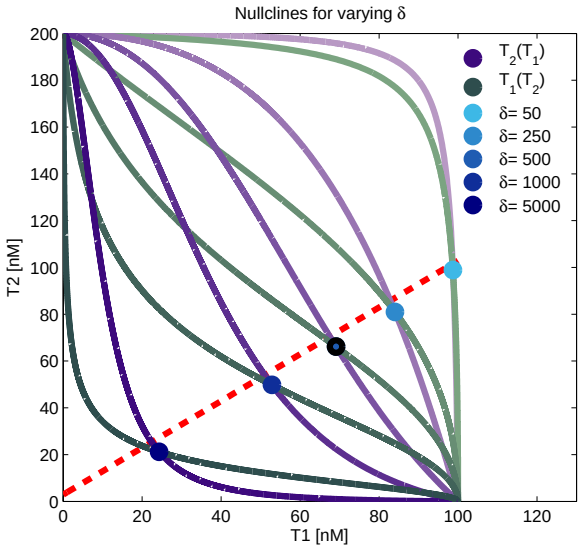


Figure 2.5: Numerical simulation: nullclines of the negative feedback scheme (2.3) in the T_1-T_2 plane, calculated for different values of δ, finding the roots of equations (2.4) and (2.5). The equilibrium corresponding to the set of nominal parameters (trajectories in Figure 2.3) is circled in black. The flow matching condition (2.6) is shown in the red dashed line. The fluxes are not exactly matched; however, we can get arbitrarily close to satisfying the matching condition if the ratio $\alpha/(\alpha+\beta)$ is close to one.

2.2.2 Cross-activation

Free molecules of R_i bind to inactive T_j and activate it:

$$R_i + T_j^* \xrightarrow{\delta_{ij}} T_j$$
$$T_i \xrightarrow{\alpha_i} T_i^*,$$

where again T_i^* is an inactive complex and $T_i^{tot} = T_i + T_i^*$. The total amount of R_i is $[R_i^{tot}] = [R_i]+[T_j]+[P]$. We now assume that T_i naturally reverts to its inactive state with rate α_i. Figure 2.1 B shows the scheme associated with this feedback interconnection. The corresponding differential equations are

$$\begin{aligned}\frac{d[T_i]}{dt} &= -\alpha_i\,[T_i] + \delta_{ji}\,[R_j]([T_i^{tot}] - [T_i]),\\ \frac{d[R_i]}{dt} &= \beta_i\,[T_i] - k\,[R_i][R_j] - \delta_{ij}\,[R_i]([T_j^{tot}] - [T_j]).\end{aligned} \tag{2.7}$$

The above differential equations were solved numerically. The parameters were chosen for illustrative purposes as $\alpha_1 = \alpha_2 = 3 \cdot 10^{-4}$ /s, $\beta_1 = \beta_2 = 0.01$ /s, $\delta_1 = \delta_2 = 5 \cdot 10^2$ /M/s, and $k = 2 \cdot 10^3$/M/s. The total amount of templates was chosen as $[T_1^{tot}] = 100$ nM, $[T_2^{tot}] = 200$ nM. The initial conditions of active $[T_i]$ are set as $[T_1](0) = 10$ nM and $[T_2](0) = 160$ nM, while $[R_1](0) = [R_2](0) = 0$. The overall result of this positive feedback interconnection is shown in Figure 2.6. The flow rate is defined again as the derivative of the total amount of $[R_i^{tot}]$. The flux mismatch is defined as the absolute value of the difference between the two flow rates. The effect of changing the feedback strength, where for simplicity $\delta_1 = \delta_2$, is shown in Figure 2.7, which plots the mean active fraction of $[T_i]$ and the mean flow mismatch over a trajectory simulated for 10 hours. The mean is shown, rather than steady-state values, to capture the behavior of the system over the whole trajectory. The right panel in Figure 2.7 seems to indicate that the flux mismatch of the two circuits is minimized for a certain range of δ around the nominal value of $\delta = 5 \cdot 10^2$. However, for values of δ that are much smaller or much larger than the nominal value of $5 \cdot 10^2$, the system dynamics do not reach steady-state within the simulated 10 hours. We will further explore the behavior of the system's equilibria and flow matching conditions, as done for the negative feedback scheme.

The nullclines of the system in the T_1-T_2 space can be calculated as done for the negative feedback design. Taking equations (2.7), we find:

$$\dot{T}_j = 0 \quad \Longrightarrow \quad R_i = \frac{\alpha_j T_j}{\delta_{ij}(T_j^{tot} - T_j)},$$
$$\dot{R}_i = 0 \quad \Longrightarrow \quad R_i = \frac{\beta_i T_i}{kR_j + \delta_{ij}(T_j^{tot} - T_j)}.$$

To simplify the derivation, we set $\delta_{12} = \delta_{21} = \delta$, $\beta_1 = \beta_2 = \beta$, $\alpha_1 = \alpha_2 = \alpha$. Equating the two expressions for R_i, we get the following equations (for $i = 1, 2$ and $j = 1, 2$):

$$\left(\frac{\alpha}{\delta}\right)^2 k \left(\frac{T_i}{T_i^{tot} - T_i}\right)\left(\frac{T_j}{T_j^{tot} - T_j}\right) + \alpha T_i - \beta T_j = 0. \tag{2.8}$$

We can find an expression of the nullclines by introducing a change of variables $z = \left(\frac{T_1}{T_1^{tot} - T_1}\right)$ and $w = \left(\frac{T_2}{T_2^{tot} - T_2}\right)$, and defining $\phi_1 = \psi_1 = \left(\frac{\alpha}{\delta}\right)^2 k$, $\phi_2 = \alpha T_1^{tot}$, $\psi_2 = \alpha T_2^{tot}$, $\phi_3 = \beta T_2^{tot}$, and finally $\psi_3 = \beta T_1^{tot}$:

$$z^2(\phi_1 v) + z(\phi_1 w + \phi_2 - \phi_3 \frac{w}{1+w}) - \phi_3 \frac{w}{1+w} = 0, \tag{2.9}$$
$$w^2(\psi_1 z) + w(\phi_1 z + \psi_2 - \psi_3 \frac{z}{1+z}) - \psi_3 \frac{z}{1+z} = 0. \tag{2.10}$$

The roots of the equations above represent the nullclines of the system. Because all the parameters in these equations are positive, there is always a single root. The nullclines are numerically solved, for varying δ, in Figure 2.8.

A condition for flow matching at steady-state can be derived as follows:

$$\dot{R}_1 - \dot{R}_2 = 0,$$
$$\beta_1 T_1 - \delta_{21} R_1 (T_2^{tot} - T_2) = \beta_2 T_2 - \delta_{12} R_2 (T_1^{tot} - T_1).$$

Substituting the expressions for R_1 and R_2 that can be derived by setting $\dot{T}_1 = 0 = \dot{T}_2$, we get:

$$\beta_1 \bar{T}_1 - \frac{\delta_{21}}{\delta_{12}} \alpha_2 \bar{T}_2 = \beta_2 \bar{T}_2 - \frac{\delta_{12}}{\delta_{21}} \alpha_1 \bar{T}_1.$$

Taking $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$, and $\delta_{12} = \delta_{21} = \delta$ we get:

$$\bar{T}_2 = \bar{T}_1. \tag{2.11}$$

This flow matching condition is shown in Figure 2.8 in the red dashed line. Decreasing α (inactivation

rate for the generating species) or increasing δ (speed of the positive feedback), with respect to the nominal values chosen here, causes the equilibrium of the system to be pushed toward the upper right corner of Figure 2.8. Moreover, when decreasing α or increasing δ the system reaches equilibrium on a timescale in the order of several dozens of hours. Explicit tradeoffs on the effects of α and δ may be found by further analysis on the nullclines and on the locus of equilibria in equation (2.8).

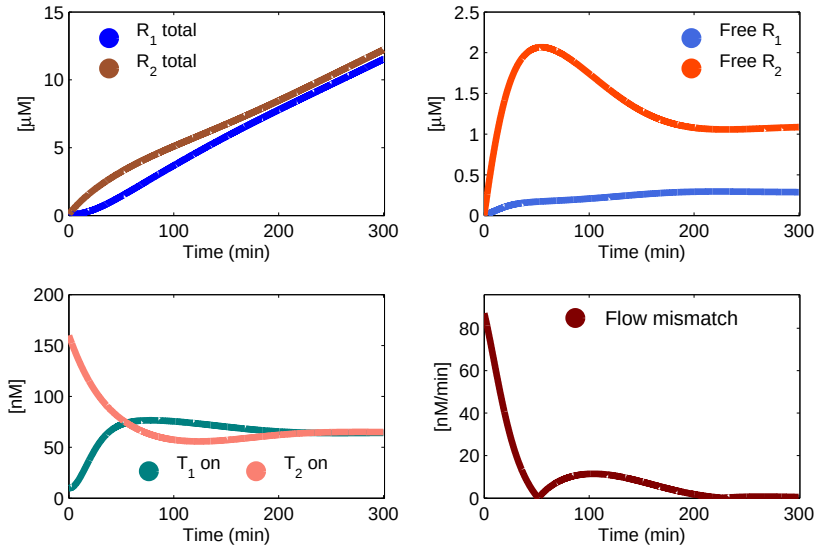


Figure 2.6: Numerical simulation showing the solution to the positive feedback architecture (Figure 2.1 C) modeled with equations (2.7). The flow mismatch between R_1 and R_2 is shown in the bottom-right panel.

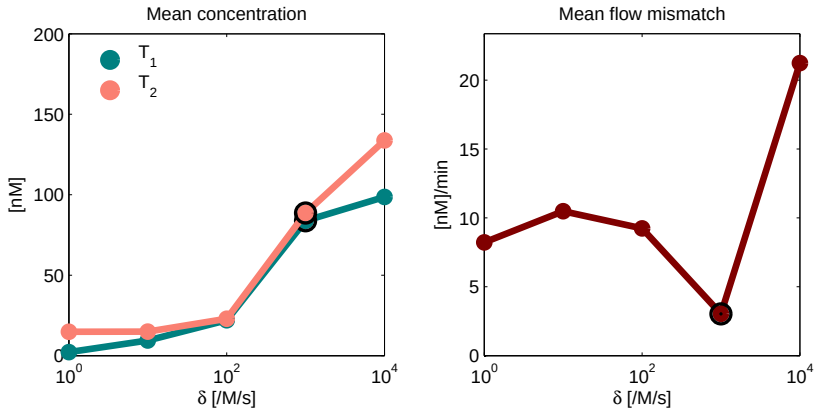


Figure 2.7: Numerical simulation for the positive feedback scheme model (2.7), showing the mean concentration of active generating species T_1 and T_2 and the mean flux mismatch as a function of the cross-activation rate parameter rate δ. The points corresponding to the set of nominal parameters (trajectories in Figure 2.6) are circled in black. The mean is taken over a trajectory of 10 hours. The flux mismatch of the two circuits seems to be minimized for a certain range of δ around the nominal value of $\delta = 5 \cdot 10^2$. However, for values of δ that are much smaller or much larger than the nominal value of $5 \cdot 10^2$, the system dynamics do not reach steady-state within the simulated 10 hours. Figure 2.8 shows the numerically computed nullclines of the system and the corresponding equilibria for varying δ.

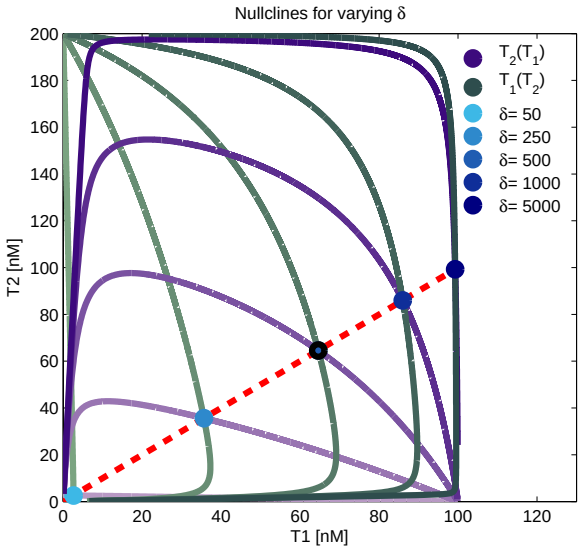


Figure 2.8: Numerical simulation: nullclines of the positive feedback scheme (2.7) in the T_1-T_2 plane, calculated for different values of δ finding the roots of equations (2.9) and (2.10). The equilibrium corresponding to the set of nominal parameters (trajectories in Figure 2.6) is circled in black. The flow matching condition (2.11) is shown in the red dashed line. The flow matching condition is satisfied by the equilibria $\bar{T}_1$ and $\bar{T}_2$ for $\delta = 5 \cdot 10^3$.

2.3 Implementation with transcriptional circuits

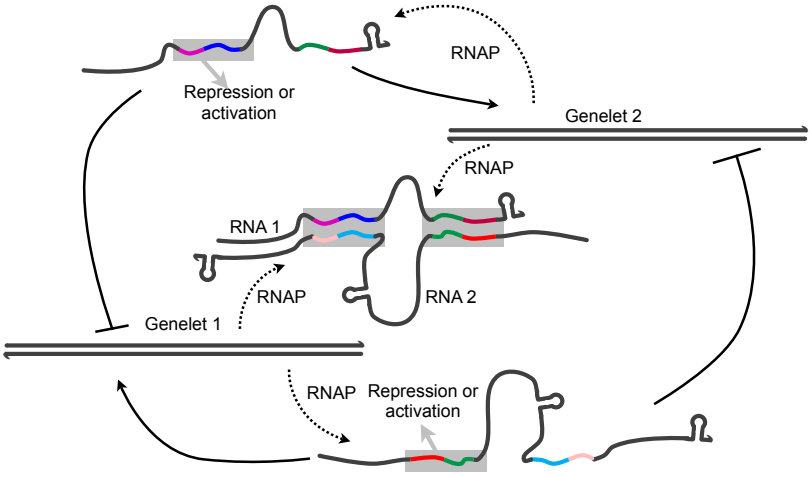


Figure 2.9: Scheme highlighting the general idea behind the transcriptional circuits implementation of the two feedback interconnections shown in Figures 2.1 B and C. Two RNA species bind to form a product, and their regulatory domains are sequestered. The feedback loops are active when either species is in excess, and therefore its regulatory domains are not covered.

The model problem described above can be experimentally tested using transcriptional circuits. The two species T_1 and T_2 correspond to two switches, whose RNA transcripts are the output reagents R_1 and R_2. Such transcripts are designed to bind and form an RNA complex P. Since the focus of this work is the investigation of the effects of feedback, the structure of P and its functionality as a standalone complex will be neglected. Depending on the feedback scheme to be implemented, the RNA species R_1 and R_2 will be designed to have different domains. However, once R_1 and R_2 are bound and form P, it will be required that the complex is inert and all the regulatory domains for negative auto-regulation or cross-activation are covered. This idea is depicted in Figure 2.9.

2.3.1 Self-repression

A graphical sketch of the domain-level design for the self-repression interconnection is shown in Figure 2.10 A. The RNA outputs of each genelet are designed to be complementary to the corresponding activator strand. However, the two RNA species are also complementary. This specification on the design of the transcripts introduces a binding domain between T_i and R_j, which is considered another off state, as shown in Figure 2.10 B. Such complex is a substrate for RNase H and the RNA strand is degraded by the enzyme, releasing the genelet activation domain. We assume that the transcription efficiency of an RNA-DNA promoter complex is very low. This hypothesis was not experimentally challenged for this specific system; however, in Section 3.7.14 we show that this assumption is valid for other genelets with the same promoter domain.

The self-inhibitory genelet design was first characterized in [61]. The circuit design proposed

here, with two-domain RNA transcripts, was originally presented in [40].

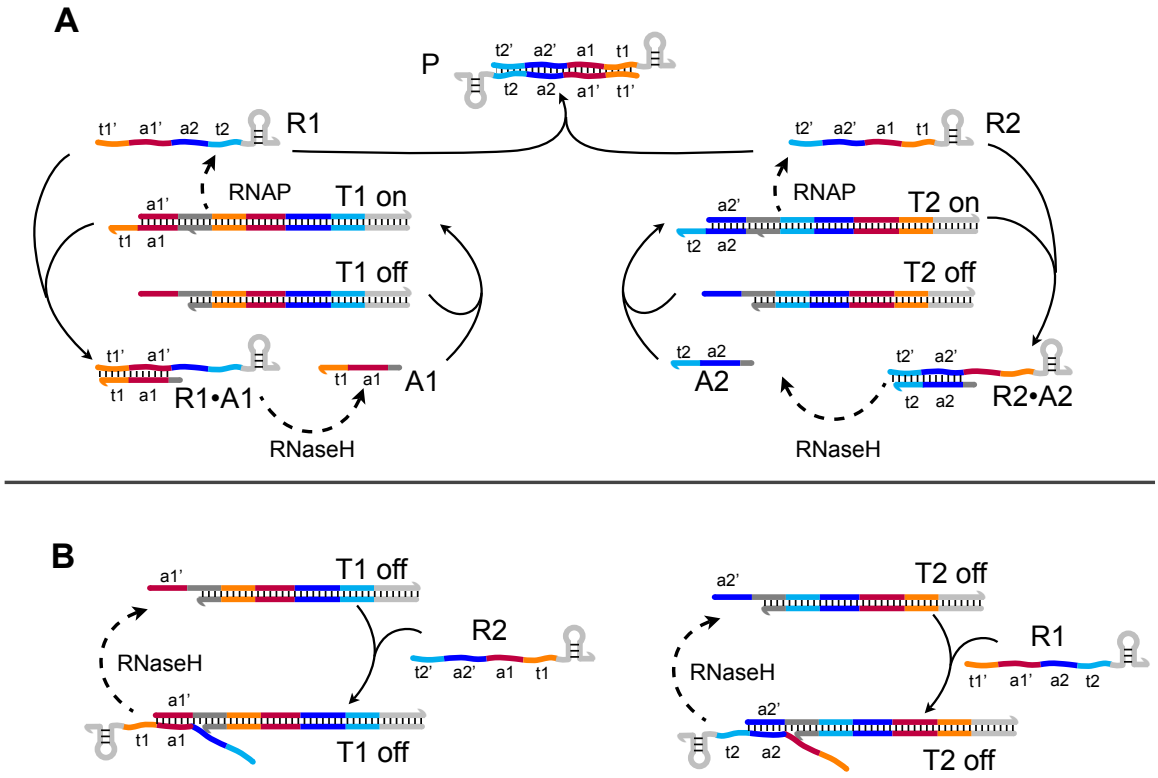


Figure 2.10: General reaction scheme representing a transcriptional circuit implementation of the negative feedback scheme in Figure 2.1 B. Complementary domains have the same color. Promoters are in dark gray, terminator hairpin sequences in light gray. The RNA output of each genelet is designed to be complementary to its corresponding activator strand. The two RNA species are also complementary. A. Desired self-inhibition loops. B. Undesired cross-hybridization and RNase H mediated degradation of the RNA-template complexes.

2.3.1.1 Modeling

Based on the outlined design specifications and the resulting molecular interactions, we can build a model for the system. The switches T_i and T_j can have three possible states: the on state where activator and template are bound and form the complex T_iA_i; the off state given by free T_i; the off state represented by R_j bound to T_i forming T_iR_j. An off state still allows for RNAP weak binding and transcription. Throughout this derivation, the dissociation constants are omitted when assumed to be negligible. It is hypothesized that the concentration of enzymes is considerably lower than that of the DNA molecules, allowing the classical steady-state assumption for Michaelis-Menten kinetics.

The overall reactions are, for $\mathrm{i} \in \{1,2\}$, $j \in \{2,1\}$:

Activation	$\mathrm{T_i} + \mathrm{A_i} \overset{\mathrm{k_{T_iA_i}}}{\rightharpoonup} \mathrm{T_i \cdot A_i}$
Inhibition	$\mathrm{R_i} + \mathrm{T_i \cdot A_i} \overset{\mathrm{k_{R_iT_iA_i}}}{\rightharpoonup} \mathrm{R_i \cdot A_i} + \mathrm{T_i}$
Annihilation	$\mathrm{R_i} + \mathrm{A_i} \overset{\mathrm{k_{R_iA_i}}}{\rightharpoonup} \mathrm{R_i \cdot A_i}$
Output formation	$\mathrm{R_i} + \mathrm{R_j} \overset{\mathrm{k_{R_iR_j}}}{\rightharpoonup} \mathrm{R_i \cdot R_j}$
Undesired hybridization	$\mathrm{R_j} + \mathrm{T_i} \overset{\mathrm{k_{R_jT_i}}}{\rightharpoonup} \mathrm{R_j \cdot T_i}.$

The enzymatic reactions are, for $\mathrm{i} \in \{1,2\}$, $j \in \{2,1\}$:

Transcription: on state	$\mathrm{RNAP} + \mathrm{T_i \cdot A_i} \underset{\mathrm{k^-_{ONii}}}{\overset{\mathrm{k^+_{ONii}}}{\rightleftharpoons}} \mathrm{RNAP \cdot T_i \cdot A_i} \overset{\mathrm{k_{catONii}}}{\rightharpoonup} \mathrm{RNAP} + \mathrm{T_iA_i} + \mathrm{R_i}$
Transcription: off state	$\mathrm{RNAP} + \mathrm{T_i} \underset{\mathrm{k^-_{OFFii}}}{\overset{\mathrm{k^+_{OFFii}}}{\rightleftharpoons}} \mathrm{RNAP \cdot T_i} \overset{\mathrm{k_{catOFFii}}}{\rightharpoonup} \mathrm{RNAP} + \mathrm{T_i} + \mathrm{R_i}$
Transcription: off state, undesired	$\mathrm{RNAP} + \mathrm{R_j \cdot T_i} \underset{\mathrm{k^-_{OFFji}}}{\overset{\mathrm{k^+_{OFFji}}}{\rightleftharpoons}} \mathrm{RNAP \cdot R_j \cdot T_i} \overset{\mathrm{k_{catOFFji}}}{\rightharpoonup} \mathrm{RNAP} + \mathrm{R_j \cdot T_i} + \mathrm{R_i}$
Degradation	$\mathrm{RNaseH} + \mathrm{R_i \cdot A_i} \underset{\mathrm{k^-_{Hii}}}{\overset{\mathrm{k^+_{Hii}}}{\rightleftharpoons}} \mathrm{RNaseH \cdot R_i \cdot A_i} \overset{\mathrm{k_{catHii}}}{\rightharpoonup} \mathrm{RNaseH} + \mathrm{A_i}$
	$\mathrm{RNaseH} + \mathrm{R_j \cdot T_i} \underset{\mathrm{k^-_{Hji}}}{\overset{\mathrm{k^+_{Hji}}}{\rightleftharpoons}} \mathrm{RNaseH \cdot R_j \cdot T_i} \overset{\mathrm{k_{catHji}}}{\rightharpoonup} \mathrm{RNaseH} + \mathrm{T_i}.$

Given the above reactions, it is straightforward to derive a set of ODEs as follows:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{T_i}] =& -\mathrm{k_{T_iA_i}}\,[\mathrm{T_i}]\,[\mathrm{A_i}] + \mathrm{k_{R_iT_iA_i}}\,[\mathrm{R_i}]\,[\mathrm{T_i \cdot A_i}] - \mathrm{k_{R_jT_i}}\,[\mathrm{R_j}]\,[\mathrm{T_i}] + \mathrm{k_{catHji}}\,[\mathrm{RNaseH \cdot R_j \cdot T_i}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{A_i}] =& -\mathrm{k_{T_iA_i}}\,[\mathrm{T_i}]\,[\mathrm{A_i}] - \mathrm{k_{R_iA_i}}\,[\mathrm{R_i}]\,[\mathrm{A_i}] + \mathrm{k_{catHii}}\,[\mathrm{RNaseH \cdot R_i \cdot A_i}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{R_i}] =& -\mathrm{k_{R_iR_j}}\,[\mathrm{R_i}]\,[\mathrm{R_j}] - \mathrm{k_{R_iT_iA_i}}\,[\mathrm{R_i}]\,[\mathrm{T_i \cdot A_i}] - \mathrm{k_{R_iT_j}}\,[\mathrm{R_i}]\,[\mathrm{T_j}] - \mathrm{k_{R_iA_i}}\,[\mathrm{R_i}]\,[\mathrm{A_i}]\\
&+ \mathrm{k_{catONii}}\,[\mathrm{RNAP \cdot T_i \cdot A_i}] + \mathrm{k_{catOFFii}}\,[\mathrm{RNAP \cdot T_i}] + \mathrm{k_{catOFFji}}\,[\mathrm{RNAP \cdot R_j \cdot T_i}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{R_i \cdot R_j}] =& +\mathrm{k_{R_iR_j}}\,[\mathrm{R_i}]\,[\mathrm{R_j}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{R_j \cdot T_i}] =& +\mathrm{k_{R_jT_i}}\,[\mathrm{R_j}]\,[\mathrm{T_i}] - \mathrm{k_{catHji}}[\mathrm{RNaseH \cdot R_j \cdot T_i}].
\end{aligned}
\tag{2.12}
$$

The molecular complexes that appear in the right-hand side of the above equations can be expressed using mass conservation: $[\mathrm{T_i \cdot A_i}] = [\mathrm{T_i^{tot}}] - [\mathrm{T_i}] - [\mathrm{R_j \cdot T_i}]$ and $[\mathrm{R_i \cdot A_i}] = [\mathrm{A_i^{tot}}] - [\mathrm{A_i}] - [\mathrm{T_i \cdot A_i}]$.

We assume that binding of enzymes to their substrate is faster than the subsequent catalytic step, and that the substrate concentration is much larger than the amount of enzyme. This allows us to use the standard Michaelis-Menten quasi-steady-state expressions. We need to define the Michaelis-Menten coefficients: for instance, for the ON state of the template, define: $\mathrm{k_{MONii}} = \frac{\mathrm{k^-_{ONii}}+\mathrm{k_{catONii}}}{\mathrm{k^+_{ONii}}}$. Then the following expressions hold:

$$[\mathrm{RNAP^{tot}}] = [\mathrm{RNAP}]\left(1+\frac{[\mathrm{T_1\cdot A_1}]}{\mathrm{k_{MON11}}}+\frac{[\mathrm{T_1}]}{\mathrm{k_{MOFF11}}}+\frac{[\mathrm{T_2\cdot A_2}]}{\mathrm{k_{MON22}}}+\frac{[\mathrm{T_2}]}{\mathrm{K_{MOFF22}}}+\frac{[\mathrm{R_2\cdot T_1}]}{\mathrm{k_{MOFF21}}}+\frac{[\mathrm{R_1\cdot T_2}]}{\mathrm{k_{MOFF12}}}\right),$$
$$[\mathrm{RNaseH^{tot}}] = [\mathrm{RNaseH}]\left(1+\frac{[\mathrm{R_1\cdot A_1}]}{\mathrm{k_{MH11}}}+\frac{[\mathrm{R_2\cdot A_2}]}{\mathrm{k_{MH22}}}+\frac{[\mathrm{R_2\cdot T_1}]}{\mathrm{k_{MH21}}}+\frac{[\mathrm{R_1\cdot T_2}]}{\mathrm{k_{MH12}}}\right).$$

We can easily rewrite the above equations as $[\mathrm{RNAP}] = \frac{[\mathrm{RNAP^{tot}}]}{\mathrm{P}}$ and $[\mathrm{RNaseH}] = \frac{[\mathrm{RNaseH^{tot}}]}{\mathrm{H}}$, with a straightforward definition of the coefficients P and H. Finally:

$$[\mathrm{RNAP\cdot T_i\cdot A_i}] = \frac{[\mathrm{RNAP^{tot}}]\,[\mathrm{T_i\cdot A_i}]}{\mathrm{P\cdot k_{MONii}}},$$
$$[\mathrm{RNAP\cdot R_j\cdot T_i}] = \frac{[\mathrm{RNAP^{tot}}]\,[\mathrm{R_j\cdot T_i}]}{\mathrm{P\cdot k_{MOFFji}}},$$
$$[\mathrm{RNAP\cdot T_i}] = \frac{[\mathrm{RNAP^{tot}}]\,[\mathrm{T_i}]}{\mathrm{P\cdot k_{MOFFii}}},$$
$$[\mathrm{RNaseH\cdot R_i\cdot A_i}] = \frac{[\mathrm{RNaseH^{tot}}]\,[\mathrm{R_i\cdot A_i}]}{\mathrm{H\cdot k_{MHii}}},$$
$$[\mathrm{RNaseH\cdot R_j\cdot T_i}] = \frac{[\mathrm{RNaseH^{tot}}]\,[\mathrm{R_j\cdot T_i}]}{\mathrm{H\cdot k_{MHji}}},$$

which can be substituted in equations (2.12).

The nonlinear set of equations (2.12) is analyzed numerically. The parameter values used in these simulations are reported in Table 2.1. Such parameters are consistent with those in [63]; this is a fair assumption since the design of this system is essentially identical to that of a repressible switch. For simplicity we assume that the circuits are symmetric, and their parameters are therefore identical. We can assess the performance of the circuit by just creating an imbalance in the concentration of the templates. Figure 2.11 shows the system trajectories that correspond to zero initial conditions for $[\mathrm{A_i}]$ and $[\mathrm{R_i}]$, while the complexes $[\mathrm{T_1A_1}] = [\mathrm{T_1^{tot}}] = 100$ nM, $[\mathrm{T_2A_2}] = [\mathrm{T_2^{tot}}] = 50$ nM, $[\mathrm{A_1^{tot}}] = 100$ nM and $[\mathrm{A_2^{tot}}] = 50$ nM. (The simulation first allows for equilibration of all the DNA strands in the absence of enzymes. Only the portion of trajectories after addition of enzymes is shown.) The total concentration of enzymes is assumed to be $[\mathrm{RNAP^{tot}}] = 80$ nM and $[\mathrm{RNaseH^{tot}}] = 8.8$ nM. The RNAP and RNase H concentrations were chosen based on typical experimental conditions. (For a brief discussion on estimating enzyme concentrations, see Table 3.3, Section 3.7.4.) Note that the concentration of RNAP is not negligible relative to the total amount of genelets present: this means that the Michaelis-Menten approximation may not be accurate in this case. The simulation results are shown in Figure 2.11 and are consistent with the traces obtained for the simple model system

shown at Figure 2.3.

Table 2.1: Simulation Parameters for Equations (2.12)

Units: [1/M/s]	Units: [1/s]	Units: [M]
$k_{T_iA_i} = 4 \cdot 10^4$	$k_{catON_{ii}} = 0.06$	$k_{MON_{ii}} = 250 \cdot 10^{-9}$
$k_{T_iA_iR_i} = 5 \cdot 10^4$	$k_{catOFF_{ii}} = 1 \cdot 10^{-3}$	$k_{MOFF_i} = 1 \cdot 10^{-6}$
$k_{A_iR_i} = 5 \cdot 10^4$	$k_{catOFF_{ij}} = 1 \cdot 10^{-3}$	$k_{MOFF_{ij}} = 1 \cdot 10^{-6}$
$k_{R_iT_j} = 1 \cdot 10^3$	$k_{catH_{ii}} = .1$	$k_{MH_{ii}} = 50 \cdot 10^{-9}$
$k_{R_iR_j} = 1 \cdot 10^6$	$k_{catH_{ji}} = .1$	$k_{MH_{ji}} = 50 \cdot 10^{-9}$

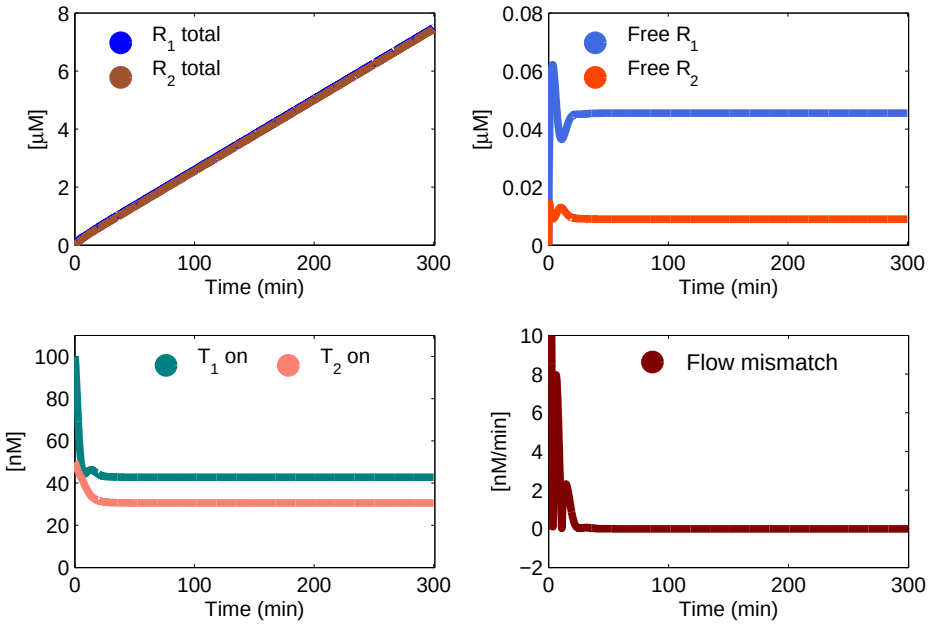


Figure 2.11: Numerical simulation for equations (2.12). Parameters are chosen as in Table 2.1. $[T_1A_1] = [T_1^{tot}] = 100$ nM, $[T_2A_2] = [T_2^{tot}] = 50$ nM, $[A_1^{tot}] = 100$ nM, and $[A_2^{tot}] = 50$ nM, $[RNAP^{tot}] = 80$ nM, and $[RNaseH^{tot}] = 8.8$ nM. These results are consistent with those of the simple model proposed in equations (2.3), and analyzed numerically in Figure 2.3.

2.3.1.2 Experimental results

We expect the feedback scheme to downregulate the production of either RNA species when in excess with respect to the other. For instance, if the concentration of $[T_1 \cdot A_1]$ is twice the concentration of $[T_2 \cdot A_2]$, the concentration of R_1 produced will clearly exceed that of R_2. If the feedback scheme is working correctly, we expect to notice a decrease in the percentage of template $[T_1 \cdot A_1]$. We can easily verify this hypothesis by labeling the 5' end of the non-template strand of the genelets with different fluorescent dyes, and by labeling the corresponding activator strand with a quencher on the 3' end. Inactive templates will emit a high fluorescence signal, while the signal of active

templates will be quenched. For instance, when A_1 is stripped off T_1, the T_1 fluorescence signal will increase. For convenience, the fluorescence traces will be processed to map the measured signal to the corresponding active genelet concentrations. In all the fluorescence traces shown here, the total amount of activators is stoichiometric to the total amount of templates: $[A_i^{tot}] = [T_i^{tot}]$.

Figure 2.12 A shows the behavior of the two genelets in isolation: we can verify that each genelet self-inhibits after the enzymes are added. (For details on the data normalization procedure, refer to Section 2.3.1.3.) The concentration of RNA present in solution can be measured through gel electrophoresis, as shown in Figure 2.12 B: lanes 1 and 2 show that transcription is effectively absent. When the two genelets are present in solution in stoichiometric amount, their RNA outputs bind quickly to form a double-stranded complex, and therefore the feedback loops become a secondary reaction (by design thermodynamically less favorable than the $R_1 \cdot R_2$ complex formation). As shown in Figure 2.12 C, the two genelets only moderately self-repress. The total RNA concentration in solution is high, as shown in the denaturing gel in Figure 2.12 B, lanes 3 and 4. A discussion on the accuracy of the gel data is in Section 2.3.1.3.

When the templates $[T_1^{tot}]$ and $[T_2^{tot}]$ are in different ratios, the system behavior is shown in Figure 2.13 A. We can plot the resulting initial active template ratio (which corresponds to the total template ratio) versus the steady-state one: we find that the system behaves symmetrically and the steady-state ratio is close to one across all the initial ratios. Therefore, given open loop transcription rates that differ across a factor of 1–3, these results suggest that the system robustly matches the flux of R_1 and R_2. If the concentration of $[T_i^{tot}]$ and $[A_i^{tot}]$ is changed over time, the steady-state concentration of active genelets adjusts as shown in Figure 2.14 A and B. Samples from this set of experiments were analyzed using a denaturing gel: the results are shown in Figure 2.14 C and D (corresponding to the traces in Figure 2.14 A and B, respectively) and show the total RNA amount in solution and that $[R_1^{tot}] \approx [R_2^{tot}]$, as desired (Figure 2.14 E and F).

The data in Figure 2.13 A were fitted using MATLAB, restricting the search algorithm to optimize a subset of parameters that are shown in Table 2.2. This subset of parameters was chosen to assess whether varying the branch migration rates and the enzyme speeds could satisfactorily explain the data collected. Such parameters were used to numerically compute equations (2.12), generating the simulated time traces shown in dashed lines in Figures 2.12 and 2.13. The fitted parameters differ from the initially postulated parameters: in particular, the binding rates for activation, inhibition, and output formation are much faster than what initially was assumed (Table 2.1); in particular, the fitted output formation rate is too high and not physically acceptable. Clearly, the current fits may be improved by extending the parameter space; this will be part of the future work on this system.

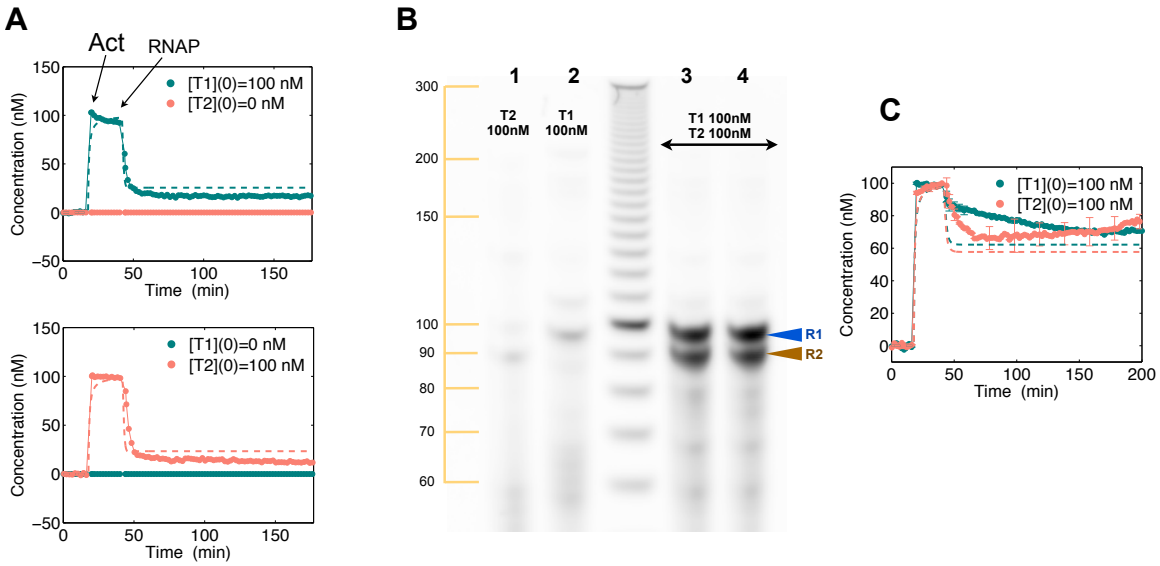


Figure 2.12: A. Experimental data showing the isolated active genelet concentrations as a function of time: the self-inhibition reaction turns the switches off, and the RNA concentration in solution is negligible, as verified in the gel electrophoresis data in panel B, lanes 1 and 2 (samples taken at steady-state after 2 h). Dashed lines represent numerical trajectories of equations (2.12), using the fitted parameters in Table 2.2. B. Denaturing gel image: lanes 1 and 2 show that the switches in isolation self-inhibit and no significant transcription is measured. Lanes 3 and 4 show the total RNA amount in samples from the experiment shown at panel C, taken at steady-state after 2 h. When the genelets are in stoichiometric amount, their flow rates are already balanced and there is only a moderate self-inhibition.

Table 2.2: Fitted Parameters for (2.12)

.

Units: [1/M/s]	Units: [1/s]
$k_{T_iA_i} = 2.9 \cdot 10^5$	$k_{catON_{ii}} = 0.06$
$k_{T_iA_iR_i} = 5 \cdot 10^5$	$k_{catH_{ii}} = .09$
$k_{A_iR_i} = 5 \cdot 10^4$	$k_{catH_{ji}} = .09$
$k_{R_iT_j} = 1 \cdot 10^3$	
$k_{R_iR_j} = 2 \cdot 10^7$	

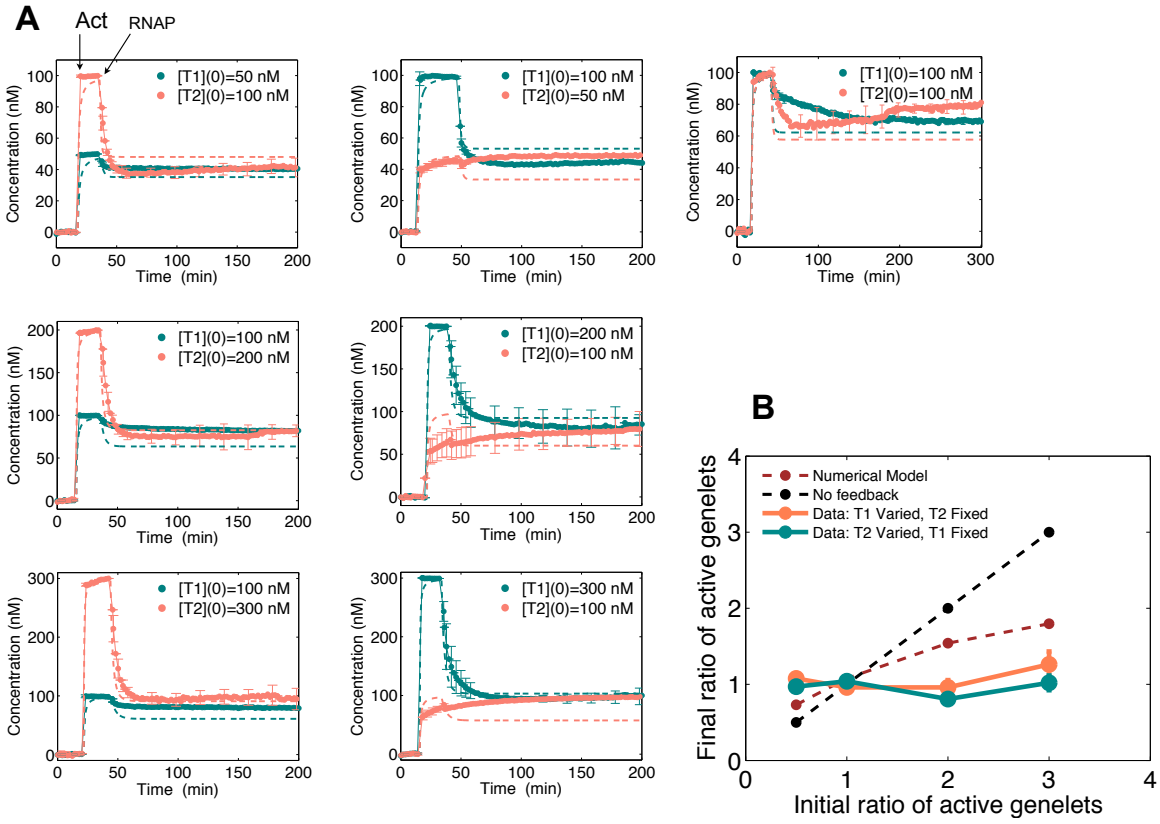


Figure 2.13: A. Concentration of active genelets over time at different total templates concentration. The concentration of activators is always stoichiometric to the amount of corresponding template. B. Overall plot for the total/initial ratio of templates versus the final ratio of active templates. Dashed lines in all the figures correspond to numerical simulations for model (2.12), using the parameters in Table 2.2.

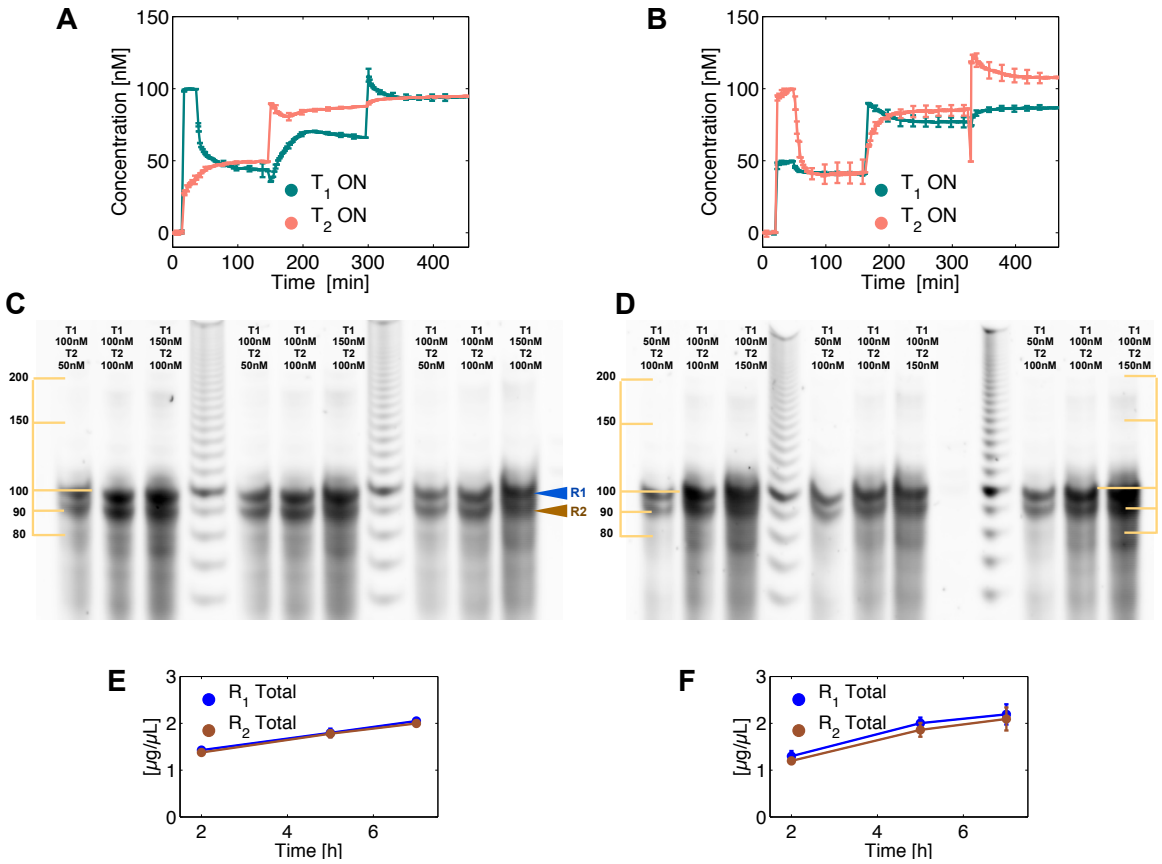


Figure 2.14: A and B. Fluorescent traces showing the adaptation of the active fraction of genelets, when the total amount of templates is varied over time. C and D. Samples from the experiments shown in panels A and B, respectively, were analyzed with gel electrophoresis. E and F show the concentrations of RNA species, estimated from the gel samples. The RNA concentration was estimated using the DNA ladder as a control. For a brief discussion on the accuracy of the RNA concentration estimates, refer to Section 2.3.1.3.

2.3.1.3 Materials and methods

DNA oligonucleotides and enzymes

The DNA strands were designed by thermodynamic analysis using the Winfree lab DNA design toolbox for MATLAB, Nupack [2] and Mfold [141]. The strands were optimized to yield free energy gains favoring the desired reactions, and to avoid unwanted secondary structures and crosstalk. Further constraints on the length and structure of the strands, which can affect the transcription efficiency and fidelity, were taken into account referring to [61], Chapter 3.4. All the strands were purchased from Integrated DNA Technologies, Coralville, IA [1]. $T_1 - nt$ is labeled with TAMRA at the $5'$ end, $T_2 - nt$ is labeled with Texas Red at the $5'$ end, both activators A_1 and A_2 are labeled with the IOWA black RQ quencher at the $3'$ end. The transcription buffer mix was prepared prior to each experiment run (two to four samples) using the T7 Megashortscript kit (#1354), Ambion, Austin, TX which includes the T7 RNA polymerase enzyme mix, the transcription buffer, and rNTPs utilized in the experiments. *E. coli* RNase H was purchased from Ambion (#2292).

Transcription protocol

The templates were annealed with 10% (v/v) 10× transcription buffer from 90°C to 37°C for 1 h 30 min at a concentration 5–10× the target concentration. The DNA activators were added to the annealed templates from a higher concentration stock, in a solution with 10% (v/v), 10× transcription buffer, 7.5 mM each NTP, 4% (v/v) T7 RNA polymerase, and .44% (v/v) *E. coli* RNase H. Each transcription experiment for fluorescence spectroscopy was prepared for a total target volume of 70 μl. Samples for gel studies were stopped by denaturing dye (80% formamide, 10 mM EDTA, 0.01g XCFF).

Data acquisition

The fluorescence was measured at 37°C every two minutes with a Horiba/Jobin Yvon Fluorolog 3 system. Excitation and emission maxima for TAMRA were set to 559 nm and 583 nm, respectively, according to the IDT reccommendation; for Texas Red the maxima for the spectrum were set to 598–617 nm. Slit widths were set to 2 nM for excitation and 4 nM for emission. The raw fluorescence data $\Phi(t)$ were converted to estimated switch activity by normalizing with respect to maximum fluorescence Φ_{max}(measured before adding activators and enzymes) and to minimum fluorescence Φ_{min} (measured after adding activators and before adding enzymes):

$$[T_iA_i](t) = [T_i^{tot}] \cdot \left(1 - \frac{\Phi(t) - \Phi_{min}}{\Phi_{max} - \Phi_{min}}\right).$$

No correction for the dilution caused by the addition of enzymes (roughly 7%) was implemented. Denaturing polyacrylamide gels (8% 19:1 acrylamide:bis and 7 M urea in TBE buffer, 100 mM Tris, 90 mM boric acid, 1 mM EDTA) were run at 67°C for 45 min with 10 V/cm in TBE buffer. Samples were loaded using Xylene Cyanol FF dye. For quantitation, denaturing gels were stained

with SYBR Gold (Molecular Probes, Eugene, OR; #S-11494). In the control lane a 10-base DNA ladder (Invitrogen, Carlsbad, CA; #1082-015) was utilized. The DNA ladder 100 bp band was used as a control to estimate the concentrations of the RNA species in solution in Figure 2.14 E and F. Gels were scanned using the Molecular Imager FX (Biorad, Hercules, CA) and analyzed using the Quantity One software (Biorad, Hercules, CA). Note that comparing the intensity of DNA and RNA bands may result in inaccurate concentration estimates. Also, the high amount of RNA produced may have caused saturation in the gel data processing, which has been neglected so far. Additional experiments are needed to clarify these issues.

Numerical simulations

The system was numerically analyzed using MATLAB (The MathWorks). Differential equations were solved using the `ode23` routine. The preliminary numerical studies to obtain some insight on the circuit behavior were performed taking the parameters from [63] and the references cited therein. After collecting the fluorescence traces shown in Figures 2.12 and 2.13, a subset of the parameters was fit using the MATLAB `fmincon` routine. The parameters fitted are the rates $k_{T_iA_i}$, $k_{T_iA_iR_i}$, $k_{A_iR_i}$, $k_{R_1R_2}$, $k_{R_iT_j}$, and the parameters $k_{catON_{ii}}$ and $k_{catH_{ij}}$. This specific subset of parameters was chosen to gain intuition on the effects of the branch migration rates (which are tunable by design of the toehold lengths), and of the enzyme speed. Clearly, the data fits shown could be improved, and to this purpose future fits will be performed on a larger set of parameters, including additional fluorescence and gel electrophoresis data. The amount of RNAP in the fit is fixed at 80 nM, wile that of RNase H is 8.8 nM, based on estimates for stock enzyme concentrations of around 1.25 μM (for more discussions on estimating enzyme concentrations, see Table 3.3, Section 3.7.4). The parameters used in this paper are reported in Tables 2.1 and 2.2.

Oligonucleotide sequences

Due to technical constraints of the supplier IDT DNA, T_1-nt and T_2-nt were shortened with respect to the nominal design to have a length of 125 bases. The strands used in the experiments are those denoted below as "Short". These modifications did not alter the regulatory domains of the transcripts R_1 and R_2. Also the full length of the main transcription products was not affected, as verified by gel electrophoresis in Figure 2.12 B.

T_1-nt Full (134-mer) 5'-CTA ATG AAC TAC TAC TAC ACA CTA ATA CGA CTC ACT ATA GGG AGA AAC AAG AAC GAC ACT AAT GAA CTA CTA CTA CAC ACC AAC CAC AAC TTT ACC TTA ACC TTA CTT ACC ACG GCA GCT GAC AAA GTC AGA AA-3' (not synthesized)

T_1-nt Short (125-mer) 5'-Tamra-CT AAT GAA CTA CTA CTA CAC ACT AAT ACG ACT CAC TAT AGG GAG AAA CAA GAA CGA CAC TAA TGA ACT ACT ACT ACA CAC CAA CCA CAA CTT TAC CTT AAC CTT ACT TAC CAC GGC AGC TGA CAA-3'

T_1-t (107-mer) 5'-TTT CTG ACT TTG TCA GCT GCC GTG GTA AGT AAG GTT AAG GTA AAG TTG TGG TTG GTG TGT AGT AGT AGT TCA TTA GTG TCG TTC TTG TTT CTC

CCT ATA GTG AGT CG-3'

A_1 (35-mer) 5'-TAT TAG TGT GTA GTA GTA GTT CAT TAG TGT CGT TC-3'

T_2-nt Full (126-mer) 5'-GGT TAA GGT AAA GTT GTG GTT GTA ATA CGA CTC ACT ATA GGG AGA AAC AAG TAA GTA AGG TTA AGG TAA AGT TGT GGT TGG TGT GTA GTA GTA GTT CAT TAG TGT CGT TCC TGA CAA AGT CAG AAA-3' (not synthesized)

T_2-nt Short (126-mer) 5'-TexasRed-GG TTA AGG TAA AGT TGT GGT TGT AAT ACG ACT CAC TAT AGG GAG AAA CAA GTA AGT AAG GTT AAG GTA AAG TTG TGG TTG GTG TGT AGT AGT AGT TCA TTA GTG TCG TTC CTG ACA AAG TCA GAA-3'

T_2-t (99-mer) 5'-TTT CTG ACT TTG TCA GGA ACG ACA CTA ATG AAC TAC TAC TAC ACA CCA ACC ACA ACT TTA CCT TAA CCT TAC TTA CTT GTT TCT CCC TAT AGT GAG TCG-3'

A_2 (35-mer) 5'-TAT TAC AAC CAC AAC TTT ACC TTA ACC TTA CTT AC-3'

R_1 (95-mer) 5' - GGG AGA AAC AAG AAC GAC ACU AAU GAA CUA CUA CUA CAC ACC AAC CAC AAC UUU ACC UUA ACC UUA CUU ACC ACG GCA GCU GAC AAA GUC AGA AA -3'

R_2 (87-mer) 5'-GGG AGA AAC AAG UAA GUA AGG UUA AGG UAA AGU UGU GGU UGG UGU GUA GUA GUA GUU CAU UAG UGU CGU UCC UGA CAA AGU CAG AAA -3'

2.3.2 Cross-activation

The implementation of the model system (2.7), graphically represented in Figure 2.1 C, presents several challenges. The general design idea that has been pursued for this systems is sketched in Figure 2.15 A. Both genelets are constitutively inhibited by a DNA inhibitor I_i. Each RNA output R_i is designed to bind to the inhibitor I_j (domains indicated as q_j-a_j-t_j), thereby releasing the activator A_j for binding to T_j. Because R_i should also cover the active domain of R_j in the formation of P, then R_i must also be complementary to A_i (domains t_i'-a_i'-q_i'): therefore, this design is structurally affected by binding of RNA to templates (as for the self-repressing circuit), and by RNA-mediated self-inhibition loops, as shown in the reaction scheme in Figure 2.15 C. Both design challenges depend on the length and sequences of the complementarity domains shared by R_i and R_j. For instance, we could avoid inserting in the RNA species the toehold sequences t_1, t_1', t_2, and t_2' to minimize the self inhibition; however, this would facilitate the formation of complexes $A_i \cdot I_i \cdot R_j$ that would slow down the release of A_i. In this thesis, I will report only the data taken on the most successful experimental design so far.

2.3.2.1 Modeling

To construct a dynamic model for the cross-activating circuit represented in Figure 2.15 A, we start from a list of all the chemical reactions that can occur.

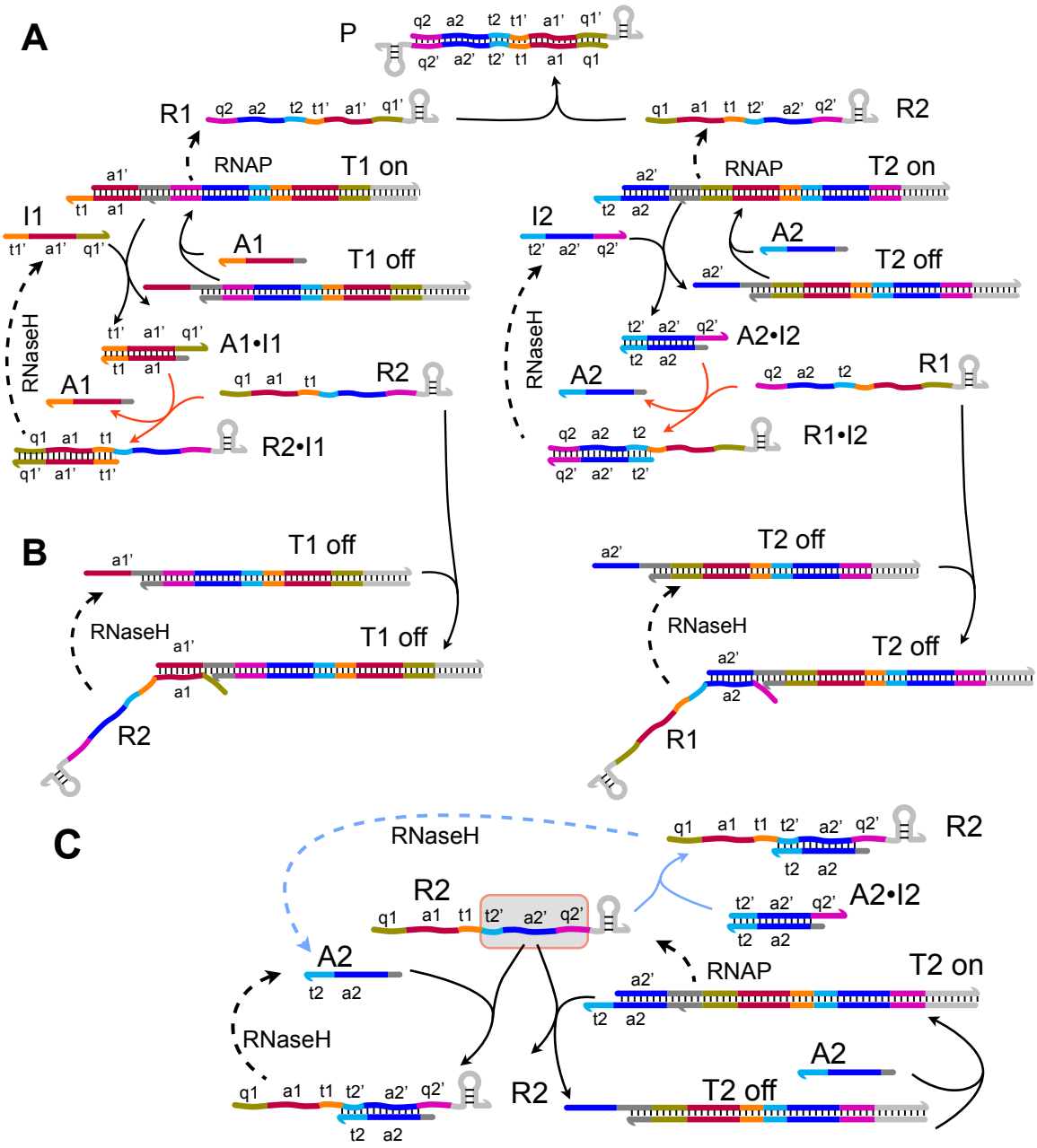


Figure 2.15: General reaction scheme of the adopted transcriptional circuits implementation for the positive feedback scheme in Figure 2.1 C. Complementary domains are represented with the same color. Promoters are colored in dark gray domains, while hairpin terminator sequences are in light gray. A. Desired cross-activation loops. The activation reaction arrows are colored in red. B. Undesired cross-activation and RNase H-mediated degradation of the RNA-template complexes. C. Undesired self-inhibition. The inhibition pathway in cyan arrows nominally should not occur, since there is no exposed toehold to favor it. However, this reaction has been observed in preliminary experiments not shown in this thesis and is therefore also included in the models.

Activation	$\mathrm{T_i + A_i \overset{k_{T_iA_i}}{\rightarrow} T_i \cdot A_i}$
Inhibition	$\mathrm{T_i \cdot A_i + I_i \overset{k_{T_iA_iI_i}}{\rightarrow} T_i + I_i \cdot A_i}$
Annihilation	$\mathrm{A_i + I_i \overset{k_{A_iI_i}}{\rightarrow} A_i \cdot I_i}$
Release	$\mathrm{R_i + A_j \cdot I_j \overset{k_{R_iA_jI_j}}{\rightarrow} R_i \cdot I_j + A_i}$
Annihilation	$\mathrm{R_i + I_j \overset{k_{R_iI_j}}{\rightarrow} R_i \cdot I_j}$
Output formation	$\mathrm{R_i + R_j \overset{k_{R_iR_j}}{\rightarrow} R_i \cdot R_j}$
Undesired interactions	$\mathrm{R_i + A_i \overset{k_{R_iA_i}}{\rightarrow} R_i \cdot A_i}$
	$\mathrm{R_i + T_j \overset{k_{R_iT_j}}{\rightarrow} R_i \cdot T_j}$

Transcription: on state	$\mathrm{RNAP + T_i \cdot A_i \underset{k^-_{ON_{ii}}}{\overset{k^+_{ON_{ii}}}{\rightleftarrows}} RNAP \cdot T_i \cdot A_i \overset{k_{catON_{ii}}}{\rightarrow} RNAP + T_i \cdot A_i + R_i}$
Transcription: off state	$\mathrm{RNAP + T_i \underset{k^-_{OFF_i}}{\overset{k^+_{OFF_i}}{\rightleftarrows}} RNAP \cdot Ti \overset{k_{catOFF_i}}{\rightarrow} RNAP + T_i + R_i}$
	$\mathrm{RNAP + R_i \cdot T_j \underset{k^-_{OFF_{ij}}}{\overset{k^+_{OFF_{ij}}}{\rightleftarrows}} RNAP \cdot R_i \cdot T_j \overset{k_{catOFF_{ij}}}{\rightarrow} RNAP + R_i \cdot T_j + R_j}$
Degradation	$\mathrm{RNaseH + R_i \cdot I_j \underset{k^-_{H_{I_j}}}{\overset{k^+_{H_{I_j}}}{\rightleftarrows}} RNaseH \cdot R_i \cdot I_j \overset{k_{catH_{I_j}}}{\rightarrow} RNaseH + I_j}$
	$\mathrm{RNaseH + R_i \cdot A_i \underset{k^-_{H_{A_i}}}{\overset{k^+_{H_{A_i}}}{\rightleftarrows}} RNaseH \cdot R_i \cdot A_i \overset{k_{catH_{A_i}}}{\rightarrow} RNaseH + A_i}$
	$\mathrm{RNaseH + R_i \cdot T_j \underset{k^-_{H_{T_j}}}{\overset{k^+_{H_{T_j}}}{\rightleftarrows}} RNaseH \cdot R_i \cdot T_j \overset{k_{catH_{T_j}}}{\rightarrow} RNaseH + T_j.}$

The resulting set of ordinary differential equations is:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{T_i}] &= -\,\mathrm{k_{T_iA_i}}\,[\mathrm{T_i}]\,[\mathrm{A_i}] - \mathrm{k_{R_jT_i}}\,[\mathrm{R_j}]\,[\mathrm{T_i}] + \mathrm{k_{T_iA_iI_i}}\,[\mathrm{T_i\cdot A_i}]\,[\mathrm{I_i}] + \mathrm{k_{catH_{T_i}}}\,[\mathrm{RNaseH\cdot R_j\cdot T_i}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{A_i}] &= -\,\mathrm{k_{T_iA_i}}\,[\mathrm{T_i}]\,[\mathrm{A_i}] - \mathrm{k_{A_iI_i}}\,[\mathrm{A_i}]\,[\mathrm{I_i}] - \mathrm{k_{R_iA_i}}\,[\mathrm{R_i}]\,[\mathrm{A_i}] + \mathrm{k_{catH_{A_i}}}\,[\mathrm{RNaseH\cdot R_i\cdot A_i}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{I_i}] &= -\,\mathrm{k_{A_iI_i}}\,[\mathrm{A_i}]\,[\mathrm{I_i}] - \mathrm{k_{T_iA_iI_i}}\,[\mathrm{T_i\cdot A_i}]\,[\mathrm{I_i}] - \mathrm{k_{R_jI_i}}\,[\mathrm{R_j}]\,[\mathrm{I_i}] + \mathrm{k_{catH_{I_i}}}\,[\mathrm{RNaseH\cdot R_j\cdot I_i}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{R_i}] &= -\,\mathrm{k_{R_iA_jI_j}}\,[\mathrm{R_i}]\,[\mathrm{A_j\cdot I_j}] - \mathrm{k_{R_iR_j}}\,[\mathrm{R_i}]\,[\mathrm{R_j}] - \mathrm{k_{R_iT_j}}\,[\mathrm{R_i}]\,[\mathrm{T_j}] - \mathrm{k_{R_iI_j}}\,[\mathrm{R_i}]\,[\mathrm{I_j}] - \mathrm{k_{R_iA_i}}\,[\mathrm{R_i}]\,[\mathrm{A_i}]\\
&\quad + \mathrm{k_{catON_{ii}}}\,[\mathrm{RNAP\cdot T_i\cdot A_i}] + \mathrm{k_{catOFF_i}}\,[\mathrm{RNAP\cdot T_i}] + \mathrm{k_{catOFF_{ji}}}\,[\mathrm{RNAP\cdot R_j\cdot T_i}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{R_i\cdot T_j}] &= +\,\mathrm{k_{R_iT_j}}\,[\mathrm{R_i}]\,[\mathrm{T_j}] - \mathrm{k_{catH_{T_j}}}\,[\mathrm{RNaseH\cdot R_i\cdot T_j}],\\
\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{R_i\cdot R_j}] &= +\,\mathrm{k_{R_iR_j}}\,[\mathrm{R_i}]\,[\mathrm{R_j}].
\end{aligned}
\tag{2.13}
$$

As previously done for the self-inhibiting circuit model, we can express the enzyme-substrate complexes using the Michaelis-Menten approximation. For the RNAP substrate, for instance, we find:

$$
[\mathrm{RNAP\cdot T_i\cdot A_i}] = \frac{[\mathrm{RNAP^{tot}}]}{\left(1 + \sum_{\mathrm{i,j}} \frac{[\mathrm{T_i\cdot A_i}]}{\mathrm{k_{MON_{ii}}}} + \frac{[\mathrm{T_i}]}{\mathrm{k_{MOFF_i}}} + \frac{[\mathrm{R_i\cdot T_j}]}{\mathrm{k_{MOFF_{ij}}}}\right)}. \tag{2.14}
$$

Analogous expressions can be derived for all other complexes.

Equations (2.13) are numerically solved using the MATLAB ode23s solver. Table 2.3 shows the parameters used for the simulations. Such generic parameters are consistent with the those in [63]. For simplicity we assume that the two sub-circuits are symmetric and have the same binding rates. We check the behavior of the system by creating an imbalance in the total concentration of inhibitors: $[\mathrm{T_1^{tot}}] = [\mathrm{A_1^{tot}}] = 50$ nM, $[\mathrm{T_2^{tot}}] = [\mathrm{A_2^{tot}}] = 100$ nM, while $[\mathrm{I_1^{tot}}] = 20$ nM and $[\mathrm{I_2^{tot}}] = 120$ nM. The simulation first allows for equilibration of all the DNA strands in the absence of enzymes. The plot shows the trajectories after addition of the enzymes, whose total concentration is assumed to be $[\mathrm{RNAP^{tot}}] = 80$ nM and $[\mathrm{RNaseH^{tot}}] = 8.8$ nM, based on typical experimental conditions (for a brief discussion on estimating enzyme concentrations, see Table 3.3, Section 3.7.4). As noted before for the self-inhibitory scheme, the concentration of RNAP is not negligible relative to the total amount of genelets present and therefore the Michaelis-Menten approximation may not be accurate in this case. The simulation results are shown in Figure 2.16 and are consistent with the traces obtained for the simple model system shown at Figure 2.6: the templates cross activate and reach an equilibrium where the flow of total RNA is matched. A comparison between the performance of the transcriptional negative and positive feedback circuits models was also done in [42].

Table 2.3: Parameters for the Initial Numerical Analysis of the Cross Activating Circuit

Units: [1/M/s]	Units: [1/s]	Units: [M]
$k_{T_iA_i} = 4 \cdot 10^4$	$k_{catON_{ii}} = 0.06$	$k_{MON_{ii}} = 250 \cdot 10^{-9}$
$k_{T_iA_iI_i} = 5 \cdot 10^4$	$k_{catOFF_i} = 1 \cdot 10^{-3}$	$k_{MOFF_i} = 1 \cdot 10^{-6}$
$k_{A_iI_i} = 5 \cdot 10^4$	$k_{catOFF_{ij}} = 1 \cdot 10^{-3}$	$k_{MOFF_{ij}} = 1 \cdot 10^{-6}$
$k_{R_jA_iI_i} = 5 \cdot 10^5$	$k_{catH_{I_i}} = .1$	$k_{MH_{I_i}} = 50 \cdot 10^{-9}$
$k_{R_iI_i} = 5 \cdot 10^5$	$k_{catH_{T_i}} = .1$	$k_{MH_{T_i}} = 50 \cdot 10^{-9}$
$k_{R_iT_j} = 1 \cdot 10^3$	$k_{catH_{A_i}} = .1$	$k_{MH_{A_i}} = 50 \cdot 10^{-9}$
$k_{R_iA_i} = 1 \cdot 10^3$		
$k_{R_iR_j} = 2 \cdot 10^5$		

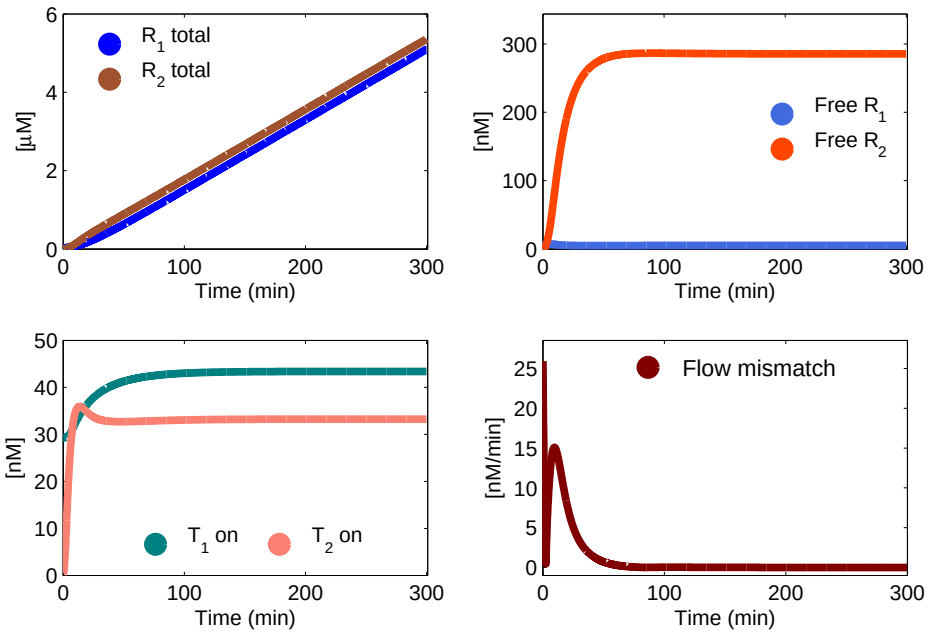


Figure 2.16: Numerical simulation for equations (2.13). Parameters are chosen as in Table 2.3. $[T_1^{tot}] = [A_1^{tot}] = 50$ nM, $[T_2^{tot}] = [A_2^{tot}] = 100$ nM, while $[I_1^{tot}] = 20$ nM, and $[I_2^{tot}] = 120$ nM. $[RNAP^{tot}] = 80$ nM and $[RNaseH^{tot}] = 8.8$ nM. These numerical results are in general consistent with those obtained for the simple model (2.7), shown in Figure 2.6.

2.3.2.2 Preliminary experimental results

I will report here preliminary results obtained on the most satisfactory transcriptional circuit design that has been considered so far. As previously noted, the challenge is to generate the cross-activation loops without introducing undesired self-inhibition or other side complexes.

The time traces shown in this section are obtained using fluorescence spectroscopy: the two genelet non-template strands were labeled on the 5' end of their activation domain with fluorescent dyes, and their corresponding activators were labeled with quenchers on the 3' end. The data were processed to represent the fraction of the genelets in their on state.

The current circuit design is not immune from the self-inhibition problem: Figure 2.17 A shows the concentration of the two activated switches in separate samples, with the addition of RNA polymerase and RNase H in sequence. Both panels show a decrease in the concentration of active genelets after addition of RNAP; such effect is more dramatic for the active T_2 switch. Addition of RNase H reduces the observed self-inhibition (although not completely for the T_2 switch, a detail which has still not been explained), confirming that it is due to RNA-DNA interactions.

The circuit was still tested (as done in the self-repression design) by observing the behavior of the two genelets when present in solution at different concentrations. To overcome the self-inhibition problem, I decided to implement three experimental strategies: use a high concentration of DNA inhibitors, a high concentration of RNase H, and lengthen the toeholds for the branch migration reaction of oligos A_2 and I_2 (toehold t_2, in reference to Figure 2.15.)

All traces shown in Figure 2.17 B were measured with a concentration of $[A_1^{tot}] = [A_2^{tot}] = 300$ nM, and $[I_1^{tot}] = [I_2^{tot}] = 1$ μM: this means that initially both genelets, regardless of the template concentrations, should be in a fully inhibited state. However, the leak in the off-state transcription appears to be sufficient to trigger cross-activation of the two circuits. This is particularly evident for the genelet at lower concentration, which is activated in most traces. The resulting ratio of the steady-state on fraction versus the total ratio of the genelets in solution is plotted in Figure 2.17 C. The reasons behind the circuit asymmetry are currently not clear. However, the data presented are encouraging and suggest that this feedback scheme, once properly debugged, may provide a robust method to maximize and match the flux of R_1 and R_2 across a range of open loop transcription rates.

The preliminary data shown in Figure 2.17 B were compared to the predictions of the dynamic model (2.13), using a set of modified parameters and an additional reaction, which has been observed to occur in preliminary data not reported here:

$$R_i + T_i \cdot A_i \overset{k_{R_i T_i A_i}}{\rightarrow} T_i + R_i \cdot A_i.$$

The activation, inhibition, product formation, and transcription rates were adjusted to the parame-

ters shown in Table 2.4. All other parameters are unchanged with respect to Table 2.3. The adjusted set of parameters was chosen to explore whether variations in the branch migration speeds and the transcription rates from an off state could easily explain the observed data. These data fits, however, do not fully explain the circuit behavior. Systematic fitting on additional experimental data, and on a larger set of parameters, is necessary to improve our understanding of the system's dynamics.

Additional experiments on the current design are needed to address primarily two issues: 1) The leaky transcription from the inhibited switches must be characterized with separate gel experiments. The off-state transcription, in the absence of activators and inhibitors, has been tested and is negligible. Therefore, the inhibition efficiency of the current designs for I_1 and I_2 might not be sufficient to keep the circuits fully inactive. 2) The cross-activation binding rates must be assessed in isolation using the RNA transcripts, to branch migrate fluorescently labeled DNA activators and inhibitors. Future and ongoing work on this project are further described in Section 5.1.

Table 2.4: Parameters for the Numerical Analysis of the Cross Activating Circuit after Fitting

Units: [1/M/s]	Units: [1/s]
$k_{T_iA_i} = 9 \cdot 10^4$	$k_{catON_{ii}} = 0.2$
$k_{T_iA_iI_i} = 1.3 \cdot 10^4$	$k_{catOFF_i} = 1 \cdot 10^{-2}$
$k_{A_iI_i} = 1.3 \cdot 10^4$	$k_{catOFF_{ij}} = 1 \cdot 10^{-2}$
$k_{R_jA_iI_i} = 1 \cdot 10^5$	
$k_{R_jI_i} = 1 \cdot 10^5$	
$k_{R_iT_j} = 1 \cdot 10^3$	
$k_{R_1A_1} = 1 \cdot 10^3$	
$k_{R_2A_2} = 1 \cdot 10^4$	
$k_{R_1T_1A_1} = 1 \cdot 10^3$	
$k_{R_2T_2A_2} = 1 \cdot 10^4$	
$k_{R_1R_2} = 1 \cdot 10^6$	

2.3.2.3 Materials and methods

DNA oligonucleotides and enzymes

The DNA strands were designed using the Winfree lab DNA design toolbox for MATLAB, Nupack [2] and Mfold [141], following the constraint guidelines in [61], Chapter 3.4. All the strands were purchased from Integrated DNA Technologies, Coralville, IA [1]. $T_1 -$ nt is labeled with TYE 563 at the $5'$ end, $T_2 -$ nt is labeled with TYE665 at the $5'$ end, both activators A_1 and A_2 are labeled with the IOWA black RQ quencher at the $3'$ end. T7 RNAP was purchased from Epicentre Biotechnologies, Cat. n. TM910K (200 U/μl). *E. coli* cloned RNase H was purchased from Ambion, Cat. n. AM2292 (10 U/μl). Inorganic lyophilized pyrophosphatase added to the transcription

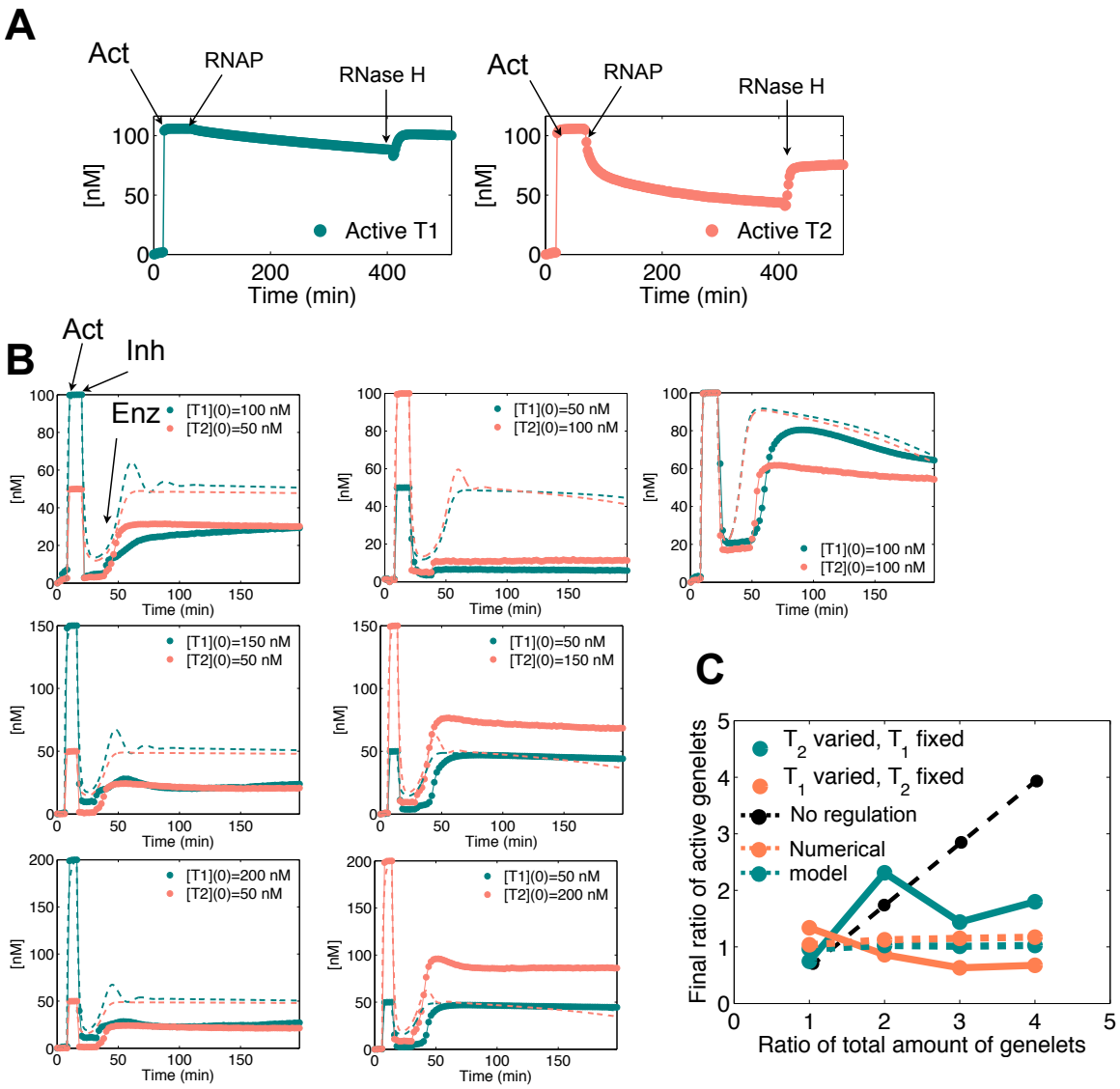


Figure 2.17: A. The current design for the transcriptional cross-activation scheme suffers from a self-inhibition side reaction. The two circuits were here considered in isolation. Self-inhibition is particularly strong for the genelet T_2, right panel. Note that adding RNase H to the active genelet T_2 does not allow recovery of the fully-on state; we do not have a good explanation for this behavior yet. B. Fluorescent traces showing the behavior of the two genelets simultaneously in solution, for different total amounts of templates. $[A_i^{tot}] = 300$ nM, $[I_i^{tot}] = 1$ μM in all the experimental traces. Dashed lines represent numerically computed traces, using model (2.13) and the adjusted parameters in Table 2.4. C. Plot showing the steady-state ratio versus total ratio of the genelets for the data collected at B.

protocol was purchased from Sigma Aldrich, Cat. n. I1891-100UN.

Transcription protocol

The transcription buffer mix was prepared prior to each experiment run (for four samples), mixing reagents to the following final concentrations: 1x Transcription Buffer and 10 mM dithiothreitol (DTT) (Epicentre Biotechnologies, Cat. n. BP1001), 7.5 mM each rNTP (Epicentre Biotechnologies, Cat. n. RN02825), 35 mM $MgCl_2$, and 0.015 U/μl pyrophosphatase (resuspended in Tris HCl 20 mM, pH 7.2, 50% glycerol (v/v)). The templates were annealed in 1× Epicentre transcription buffer from 90°C to 37°C for 1 h 30 min at a concentration 5–10× the target concentration. The annealed templates, DNA activators, and inhibitors were added to the transcription buffer mix and incubated at 37°C. Each transcription experiment for fluorescence spectroscopy was prepared for a total target volume of 60 μl. Enzymes were pre-mixed before each experimental run (two or four samples), with a volume ratio of 5:1, and 4 μL of the pre-mix was added to each cuvette.

Data acquisition

All fluorescence experiments were performed on a Horiba/Jobin Yvon Fluorolog 3 system, using 45 μL sample chamber quartz cuvettes. Fluorescence was measured at 37° C every two minutes. Excitation/emission for TYE563 (T_1) were set to the maxima 549–563 nm, those for TYE665 (T_2) at 645–665 nm, as recommended by the manufacturer IDT DNA. Slit widths were set to 2 nM for excitation and 4 nM for emission. The raw measured data were converted to estimated switch activity by normalizing with respect to maximum fluorescence $\Phi_{\max}$ (measured before adding activators, inhibitors and enzymes) and to minimum fluorescence $\Phi_{\min}$ (measured after adding activators and before adding inhibitors and enzymes):

$$[\mathrm{T_iA_i}](\mathrm{t}) = [\mathrm{T_i^{tot}}] \cdot \left(1 - \frac{\mathrm{v(t)} \cdot \Phi(\mathrm{t}) - \Phi_{\min}}{\Phi_{\max} - \Phi_{\min}}\right),$$

where $v(t)$ is a correction factor that takes into account the percent volume dilution introduced by the addition of activators and inhibitors. For the data in Figure 2.17 A, $v = 1$. For the data in Figure 2.17 B and C, after the addition of activators and inhibitors, $v(t)$ is taken to be 1.15; after addition of enzymes $v(t)$=1.2. (Dilution was very high due to the low concentration of stock aliquots and to the high target concentration of activators and inhibitors.)

Numerical simulations

The system was numerically analyzed using MATLAB (The MathWorks). Differential equations were solved using the `ode23` routine. The initial numerical studies to obtain some insight on the circuit behavior were performed using the parameters in Table 2.3, which are consistent with [63] and the references cited therein. After data collection, a subset of the parameters was tuned. The adjusted parameters are the rates $k_{T_iA_i}$, $k_{A_iI_i}$, $k_{T_iA_iI_i}$, $k_{R_iA_iI_i}$, $k_{R_iA_i}$, $k_{R_iR_j}$, $k_{R_iT_j}$, and the parameters $k_{catON_{ii}}$, k_{catOFF_i}, $k_{catOFF_{ij}}$; their values are in Table 2.4. The total amount of RNAP was fixed at

80 nM, the total amount of RNase H at 11 nM (to reflect on the higher volume of RNase H used in this set of experiments, relative to the negative auto-regulation project).

Oligonucleotides sequences

Due to technical constraints of the supplier IDT DNA, T_1-nt and T_2-nt were shortened (with respect to the nominal design) to a length of 100 bases. These modifications do not alter the regulatory domains of the transcripts. The full length of the main transcription products was not affected, as verified by gel electrophoresis (data not shown). A_2 and I_2, V1 were used for the experiment in Figure 2.17 A. A_2 and I_2, V2 were instead used for the all the experiments in Figure 2.17 B and C.

T_1-nt Full (121-mer) 5'-CAT TAG TGT CGT TCG TTC ATA ATA CGA CTC ACT ATA GGG AGA AGT GGT TAA GGT ATA GTT AGA TAG GTA AGG CAT GTT CAT TAG TGT CGT TGT GTA GTG TTG CTG ACT AAA AGT CAG AAA A-3' (not synthesized)

T_1-nt Short (99-mer) 5'-TYE563-CAT TAG TGT CGT TCG TTC ATA ATA CGA CTC ACT ATA GGG AGA AGT GGT TAA GGT ATA GTT AGA TAG GTA AGG CAT GTT CAT TAG TGT CGT TGT GTA GTG -3'

T_1-t (97-mer) 5'-TTT TCT GAC TTT TAG TCA GCA ACA CTA CAC AAC GAC ACT AAT GAA CAT GCC TTA CCT ATC TAA CTA TAC CTT AAC CAC TTC TCC CTA TAG TGA GTC G-3'

T_2-nt Full (116-mer) 5'- CCT TAC CTA TCT AAC TAT ATA ATA CGA CTC ACT ATA GGG AGA CAA CAC TAC ACT GAA CGA ACG ACA CTA ATG AAC ATG CCT TAC CTA TCT ACC TTA ACC ACT TGA CAA AGT CAA AA-3' (not synthesized)

T_2-nt Short (99-mer) 5'- TYE665-CCT TAC CTA TCT AAC TAT ATA ATA CGA CTC ACT ATA GGG AGA CAA CAC TAC ACT GAA CGA ACG ACA CTA ATG AAC ATG CCT TAC CTA TCT ACC TTA ACC -3'

T_2-t (92-mer) 5'-TTT TGA CTT TGT CAA GTG GTT AAG GTA GAT AGG TAA GGC ATG TTC ATT AGT GTC GTT CGT TCA GTG TAG TGT TGT CTC CCT ATA GTG AGT CG-3'

A_1 (30-mer) 5'-TAT TAT GAA CGA ACG ACA CTA ATG AAC TAC-IOWA black RQ-3'

I_1 (36-mer) 5'-GTA GTT CAT TAG TGT CGT TCG TTC AGT GTA GTG TTG-3'

A_2 V1 (30-mer) 5'-TAT TAT ATA GTT AGA TAG GTA AGG CAT TAG-IOWA black RQ-3'

I_2 V1 (36-mer) 5'-CTA ATG CCT TAC CTA TCT AAC TAT ACC TTA ACC ACT-3'

A_2 V2 (33-mer) 5'-TAT TAT ATA GTT AGA TAG GTA AGG CAT TAG TAG -IOWA black RQ-3'

I_2 V2 (39-mer) 5'-CTA CTA ATG CCT TAC CTA TCT AAC TAT ACC TTA ACC ACT -3'

2.4 Discussion

Regulating the flux of reagents in a molecular network is a relevant problem in the context of synthetic biology. Here, a model problem for matching the flow of two reagents has been formulated using simple ODEs. The results reported suggest that flux-matching can be achieved by programming this small network to include feedback loops. Self-repression is useful to eliminate excess of unused reagents, while cross-activation is a design more suited to maximizing both flows.

Reactions analogous to the model problem (2.1) are typically used to model the small regulatory RNA (sRNA) pathway (considered in this thesis in Section 4.3.2). The main feature of this pathway (present in many bacteria and eukaryotes) is that a single regulatory small RNA can bind to and down-regulate the translation of several other target mRNA species. This gene expression control strategy is advantageous when cells starve on specific nutrients or minerals, and all non-essential pathways relying on the scarce resource should be quickly shut down. Several studies have been recently published to compare the quantitative features of this pathway to protein-mediated transcriptional control [70, 83, 85]. In particular, advantages of RNA-based control of gene expression are speed and ultrasensitivity.

In the sRNA pathway, the complex formed by the regulatory RNA and its target is simply degraded, blocking translation. To my knowledge, the schemes proposed in this chapter have no correspondence with *in vivo* RNA-mediated feedback loops that can regulate RNA transcription [51]. However, it is conceivable that similar self-repression and cross-activation loops could be implemented *in vivo* by substituting the promoter structural alteration characteristic of transcriptional circuits, with a post-transcriptional control mechanism, such as cleavage through ribozymes or alteration of the ribosome binding site [50, 43]. These schemes could be useful when large synthetic pathways are integrated into a host: it is likely that minimization of the translation burden of unnecessary enzymes (negative feedback) or maximization the output of the pathway (positive feedback) would be possible without tuning promoter strengths in open loop. An *in vivo* version of these schemes might be robust with respect to promoter strength or plasmid copy number. Clearly, the main challenges to such implementation are posed by the diffusion limits, spatial anisotropy, and nonspecific binding sites in the cellular environment.

I will make a few further remarks regarding the negative feedback scheme. In his pioneering work on design of genetic networks, M. Savageau indicated that negative auto-regulation is the best control strategy for genes in low demand. The results presented here suggest that negative auto-regulation by means of stoichiometric binding is useful to adjust the production rate of a species to a level that is close to the flux effectively needed by the compartment. A formalization and generalization of this concept is currently being investigated. Catalytic self-repression has been shown to speed up the rise time of transcriptional networks [101], and in general to increase the robustness of gene

expression profiles at the population level [16]. A thorough mathematical comparison between the properties of stoichiometric and catalytic negative regulation will be considered in the future.

The positive feedback scheme presented here is useful to maximize output formation; therefore, this scheme would be appropriate if circuit design specifications required a high amount of product. This is consistent with M. Savageau's theory in the context of control of gene expression: positive regulation is the best strategy for genes in high demand [108]. Positive auto-regulation in transcriptional control of gene expression is a motif exhibiting a slow response time [8]. In feedforward loops, which are perhaps more similar to our cross-activation structure, positive regulation introduces a delay in the network; this delay is programmable by defining the activation thresholds [8]. Based on numerical analysis not shown in this thesis, tuning the concentration of DNA inhibitor in an activated genelet achieves the same tunable delay property. One of the main drawbacks of our cross-activation design is the undesired self-inhibition pathway; this issue might be solved using DNA translation gates [120] or a pair of decoupling genelets within the activation pathway.

In summary, this chapter was dedicated to the analysis of two different feedback schemes to achieve flux matching for two biochemical species interacting to form a product. The modeling and experimental results suggest that feedback confers robustness to the system with respect to the open loop production rate of the reagents. Ongoing and future work on this project are outlined in Section 5.1.

Chapter 3

Modular biochemical networks: timing molecular devices with a transcriptional clock

3.1 Introduction

Circadian clocks, which time the functionalities of a whole organism [133], are probably the best example of a biochemical oscillatory system. To explore the features and properties of biochemical oscillators, in the last decade several synthetic clock systems with a reduced number of components have been constructed *in vivo* [13, 25, 33, 44, 122, 128]. However, most of these artificial oscillators are still relatively complex and difficult to understand quantitatively, since they interact with the complex and uncertain environment of their host, and rely on the full transcriptional machinery of the cell. At the other extreme, inorganic oscillators can be quite robust, but difficult to systematically couple to a wide range of downstream processes [67, 71].

Recently, molecular oscillatory circuits have been designed and implemented in the *in vitro* context [64, 86], opening the possibility to use *in vitro* clocks to drive other circuits or devices.

In this chapter, the transcriptional oscillator proposed in [64] is used to time downstream molecular processes and explore the general challenges arising in coupled biochemical reaction networks. In particular, the focus is on how a downstream process can perturb the dynamic behavior of the "core oscillator", and how this "retroactivity" phenomenon [106, 27, 39] can be reduced.

As a simple "load" for the biochemical oscillator, a DNA nanomechanical device called "DNA tweezers" [139] was designed to selectively bind to a specific oligonucleotide component of the oscillator. These tweezers are built with two double-helical domains connected by a hinge, and have two single-stranded "hands". The hands can be designed to bind to a targeted single-stranded nucleic acid and thereby close the tweezers. The perturbation introduced by the chemical interactions between the oscillator and the tweezers are proportional to the total amount of tweezers present in

solution. To increase the robustness of clock with respect to this molecular load, an "insulator circuit" was developed to put only a small load on the oscillator, amplify the measured signal, and drive larger downstream loads. The project presented in this chapter was developed in close collaboration with J. Kim, E. Winfree and the group of Prof. Friedrich Simmel at the Technical University in Munich. The data presented here were collected by me, unless otherwise noted. Figures are adapted from [41].

3.2 Problem formulation

To highlight the challenges that arise when coupling molecular systems, I will start by illustrating a model problem. First, I will introduce a simple molecular oscillator model, which has been previously proposed in [64] to describe the transcriptional clock used in this project. Then, such model will be extended to include coupling of the oscillator to an additional molecular species, with the general purpose of transmitting its oscillatory signal to a downstream device. The objective here is to emphasize how the structure of the resulting biochemical reaction network is perturbed, and how such perturbation can be reduced by suitably modifying the coupling mechanism. We will consider the following equations, proposed in [64], as a simple model for our oscillator:

$$\begin{aligned}
\frac{\mathrm{d[rA1]}}{\mathrm{dt}} &= \mathrm{k_p[SW12] - k_d[rA1]},\\
\tau\frac{\mathrm{d[SW21]}}{\mathrm{dt}} &= \mathrm{[SW21^{tot}]}\frac{\frac{\mathrm{[rA1]^m}}{\mathrm{KA^m}}}{1+\frac{\mathrm{[rA1]^m}}{\mathrm{KA^m}}} - \mathrm{[SW21]},\\
\frac{\mathrm{d[rI2]}}{\mathrm{dt}} &= \mathrm{k_p[SW21] - k_d[rI2]},\\
\tau\frac{\mathrm{d[SW12]}}{\mathrm{dt}} &= \mathrm{[SW12^{tot}]}\frac{1}{1+\frac{\mathrm{[rI2]^n}}{\mathrm{KI^n}}} - \mathrm{[SW12]}.
\end{aligned}$$

The species rA1 and rI2 are RNA molecules that interact through two genelet switches that produce them, respectively, SW12 and SW21. In particular, rA1 is an activator for SW21, while rI2 is an inhibitor for SW12 (Figure 3.1 A). The effectiveness of the RNA species in activating or repressing the switches is modulated by the thresholds KA and KI, and by the Hill coefficients m and n. The relaxation constant τ scales the speed of the switches dynamics. The concentration of each species oscillates for suitable choices of the parameters. For this system, the existence of an oscillatory solution can be shown to be equivalent to the instability of the sole equilibrium point for the system (see Appendix, Section 3.7.1). By linearizing the dynamics around such equilibrium, and by examining the eigenvalues of the Jacobian, it is possible to numerically assess the oscillatory domain. In particular, this domain is defined by the Hill coefficients m and n and by the lumped parameters $\beta = (\mathrm{kp/kd})[\mathrm{SW21^{tot}}]/\mathrm{KI}$ and $\alpha = (\mathrm{kp/kd})[\mathrm{SW12^{tot}}]/\mathrm{KA}$. Figure 3.1 D shows the numerically

computed oscillatory domain when $m = n$ and $\alpha = \beta$. Unless otherwise noted, from now on the operating point of this oscillator model is defined by the parameters $k_p = 0.05/s$, $k_d = 0.002/s$, KA=KI=0.5 μM, $[SW21^{tot}] = [SW12^{tot}] = 100\,nM$, m=n=5, and $\tau = 500\,s$.

We want to use one of the oscillator component species to bind to a "load" molecule L, driving the periodic formation of an "active" complex L^a. We assume that $[L^{tot}] = [L] + [L^a]$. We will make a distinction on whether the mass of the oscillator component driving the load is consumed or not by binding to the load. Such a distinction is relevant with respect to our experimental system, as we will remark later. Without loss of generality, we will consider the case where the species rI2 is coupled to the load L. Additional chemical reactions are now present in the system. The active form of the load is produced according to the second-order reaction: $rI2 + L \xrightarrow{k_f} L^a$. The active load decays to its inactive form: $L^a \xrightarrow{k_r} L$. If the mass of the oscillator species is not consumed, the previous reaction is replaced by: $L^a \xrightarrow{k_r} rI2 + L$. In both cases, the concentration dynamics for L are:

$$\frac{d[L^a]}{dt} = -k_r \cdot [L^a] + k_f \cdot [L][rI2]. \tag{3.1}$$

The rI2 concentration dynamics are perturbed by the new reactions:

$$\frac{d[rI2]}{dt} = k_p \cdot [SW21] - k_d \cdot [rI2] \underbrace{+k_r \cdot [L^a] \overbrace{-k_f \cdot [L][rI2]}^{\text{consumptive}}}_{\text{non-consumptive}}, \tag{3.2}$$

where the braces highlight the additional terms appearing in the consumptive and non-consumptive coupling cases.

Let us for now ignore the perturbation introduced by the presence of L on the oscillator, and assume that [rI2(t)] is unaffected by L. Then, we can approximate the solution of equation (3.1) with the quasi-steady-state expression:

$$\widehat{[L^a]}(t) = [L^{tot}] \left(1 - \frac{k_r}{k_r + k_f[rI2(t)]}\right). \tag{3.3}$$

The above approximation is satisfactory when the load binding rates are faster than the timescale of the oscillator, as shown in detail in the Appendix, Section 3.7.1.2. Here, for illustrative purposes, we choose $k_r \approx k_d = 0.006/s$. Therefore, the load dynamics are always faster than the dynamics of rI2, and converge to a periodic orbit forced by the oscillatory input (this is demonstrated in the Appendix, Section 3.7.1.2.

Referring to equation (3.3), we can make some considerations on the "signal propagation" from the oscillator to the load. Suppose $[rI2](t) \approx A_0 + A_1 \sin \omega t$: then, the ratio of k_r and k_f [rI2] influences

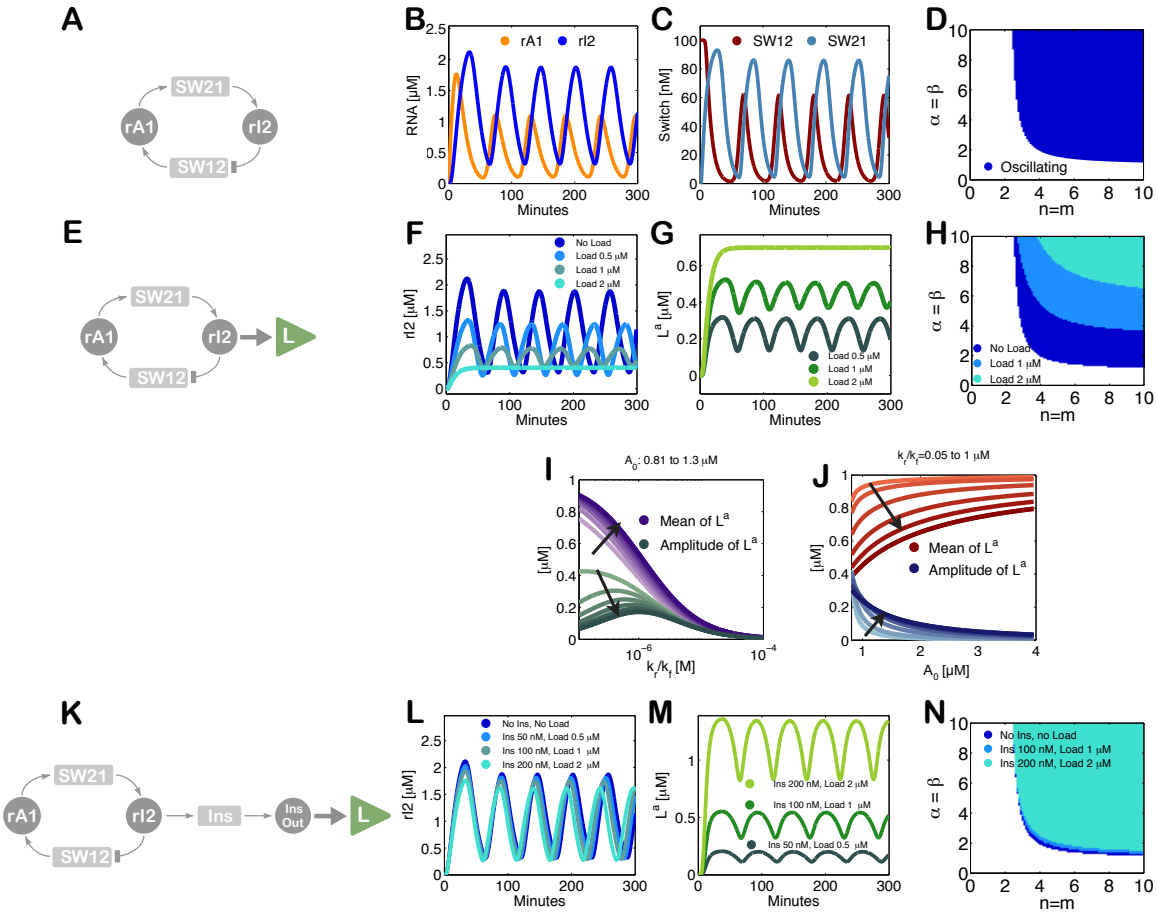


Figure 3.1: Circuits and simulations for a simple oscillator system coupled to a load. Unless otherwise noted, the parameters used for all simulations in this panel are: $k_p = 0.05/s$, $k_d = 0.002/s$, KA=KI=0.5 μM, $[SW21^{tot}] = [SW12^{tot}] = 100\,nM$, m=n=5, $\tau = 500\,s$, $k_r = 0.006/s$, $k_f = 7.9 \cdot 10^3/M/s$. For the insulating gene, the RNA output production rate is $k_p^i = 0.15/s$, and the RNA degradation rate is $k_d^i = 0.006/s$. The consumptive binding rates of the insulator and rI2 are chosen as $k_r = 0.006/s$ and $k_f = 7.9 \cdot 10^3/M/s$. The binding rates of the insulator RNA output and the load are chosen as $k_r^i = 0.006/s$ and $k_f^i = 6 \cdot 10^3/M/s$. A. Diagram for the simple model for the oscillator. B. Time traces for the oscillator species rA1 and rI2. C. Time traces for the oscillator species SW12 and SW21. D. Oscillatory domain of the simple model as a function of the non-dimensional parameters $\alpha = \beta$ and m = n. E. Oscillator scheme with consumptive load coupled to rI2. F and G. Time traces for the oscillator and load for consumptive coupling on rI2. H. The oscillatory domain shrinks as a function of $[L^{tot}]$ for the consumptive coupling to rI2. I. Mean and amplitude of the active load $[L^a]$ as a function of the ratio of k_r and k_f, when the driving input is $rI2 = A_0 + A_1 \sin \omega t$, with A_0 varying between 0.81 (light color) and 1.3 μM, and $A_1 = .8\mu M$, $\omega = 0.001\,rad/s$. J. Mean and amplitude of the active load signal $[L^a]$ as a function of the baseline A_0 for the input oscillating signal, for ratios k_r/k_f varying between 0.05 and 1 μM. K. Oscillator scheme with consumptive insulating circuit and consumptive load. L and M. Time traces for the oscillator and load when the insulating genelet is used to amplify rI2. N. The perturbation of the oscillatory domain is reduced by using a small amount of an additional genelet (insulator) that amplifies the oscillatory signal. Figure reproduced from [41]

the amplitude of the load signal as shown in Figure 3.1 I and J. In particular, we can derive the ratio $k_r/k_f = \sqrt{(A_0^2 - A_1^2)}$, which maximizes the amplitude of $[\widehat{L^a}]$. Therefore, for $A_0 \approx 1.1\,\mu$M and $A_1 \approx 0.8\,\mu$M, if we choose $k_r = 0.006 \approx 3 \cdot k_d$, then $k_f \approx 7.9 \cdot 10^3$ is the optimal binding rate. However, a larger k_f will increase the mean value of $[L^a]$. Another observation is that a high mean A_0 results in a lower load amplitude (Figure 3.1 I and J).

Under the assumption that the load dynamics are well approximated with their stationary solution, we can write new expressions for the perturbed dynamics of rI2. For the consumptive case we have:

$$\frac{d[\widehat{rI2}]}{dt} = k_p \cdot [SW21] - k_d \cdot [\widehat{rI2}] - \boxed{k_r \cdot [L^{tot}] \frac{k_f\,[\widehat{rI2}]}{k_r + k_f [\widehat{rI2}]}}, \tag{3.4}$$

where the box highlights the stationary perturbation term. This term is bounded by the constant $k_r[L^{tot}]$, and converges to it for large values of $k_f[\widehat{rI2}]$. Loosely speaking, adding the load is similar to introducing in the rI2 dynamics a new degradation term, directly proportional to the total load amount. While the approximated trajectory (3.4) provides qualitative insight on the system behavior, we report the full numerical simulations of the five ordinary differential equations describing the oscillator with load in Figure 3.1, which shows the rI2 and load trajectories for increasing $[L^{tot}]$. The oscillatory domain of the system is consequently altered as shown in Figure 3.1 H. Numerically simulated time traces of the oscillator and of the load are shown in Figure 3.1 F and G.

For the non-consumptive case, it is easy to see that the sum of the approximated perturbation terms is equal to zero. Therefore, we can conclude that after a transient the dynamics of the oscillator are unaffected by the presence of the load. Numerical simulations that testify this result are reported in the Appendix, Figure 3.10.

It is important to emphasize that the non-consumptive case has been previously analyzed in [27]. By assuming rigorous time-scale separation of the load dynamics relative to the driving chemical species, the authors were able to derive a general expression for the "retroactivity" caused by the load. Although not derived under the same assumptions on the parameters, our conclusions are consistent with the results reported in that work, where the retroactivity can be minimized by choosing appropriately fast binding rates and by reducing the total load amount.

In practical cases it may be impossible to couple non-consumptively a signaling molecule to the desired load. It may also not be possible to adjust the binding rates arbitrarily to provide small retroactivity and good signal transmission. If we fall in the consumptive load coupling case with limited freedom in tuning k_f and k_r, expression (3.4) shows that the only way to bound the perturbation on rI2 is to reduce $[L^{tot}]$. We can overcome this limit by using rI2 to activate another genelet, whose RNA output amplifies the oscillator signal and can drive larger amounts of $[L^{tot}]$.

The genelet effectively acts as an "insulator" and will be denoted as Ins. We assume that the genelet Ins binds to rI2 consumptively: $\mathrm{rI2} + \mathrm{Ins} \xrightarrow{k_f} \mathrm{Ins^a}$, $\mathrm{Ins^a} \xrightarrow{k_r} \mathrm{Ins}$. The active genelet $\mathrm{Ins^a}$ produces an RNA output similarly to the oscillator switches: $\mathrm{Ins^a} \xrightarrow{k_p^i} \mathrm{Ins^a} + \mathrm{InsOut}$. We finally assume that the RNA output, which in practice amplifies the oscillatory signal, in turn activates the desired load by the usual consumptive binding mechanism: $\mathrm{InsOut} + \mathrm{L} \xrightarrow{k_f^i} \mathrm{L^a}$, $\mathrm{L^a} \xrightarrow{k_r^i} \mathrm{L}$. The RNA output is also degraded as the other RNA species in the system: $\mathrm{InsOut} \xrightarrow{k_d^i} \emptyset$. The full set of dynamic equations are reported in the Appendix, Section 3.7.1.3. For illustrative purposes we assume that $k_p^i = 0.16\,/\mathrm{s}$, $k_r^i = 0.006\,/\mathrm{s}$, and $k_f^i = 6 \cdot 10^3\,/\mathrm{M/s}$. As shown in Figure 3.1 L and M, using a small amount of insulator genelet it is possible to drive large amounts of load introducing negligible perturbations. The oscillatory domain of the system is not significantly affected, as shown in Figure 3.1 N.

3.3 Experimental results

3.3.1 Synthesis of a molecular oscillator using transcriptional circuits

Two transcriptional circuits can be interconnected through their RNA transcripts according to the scheme considered in our problem formulation. Figure 3.2 A shows the design specifications to achieve activation of the switch SW21 through the RNA activator rA1 (output of template T12), and inhibition of SW12 through the RNA inhibitor rI2 (output of template T21). This design has been proposed in [64], and builds up on earlier work on transcriptional circuits [62, 63, 124]. The DNA species A1, A2, and dI1 generate ultrasensitive responses in the activation and inhibition of the switches [21, 81]. This feature is consistent with the oscillatory response of the simple model considered in Section 3.2, where high Hill coefficients are needed to guarantee the existence of a stable periodic orbit (see Figure 3.1 D). dI1 and A2 are added in excess over their complementary counterparts A1 and T12, respectively. As shown in Figure 3.2 B, such high concentrations represent tunable thresholds that influence the dynamic behavior of the oscillator.

The total concentration of templates and thresholds, together with the concentration of enzymes, determine the stable or oscillatory behavior of the system. We choose the operating point of the oscillator as 4–6 large amplitude oscillations within 15 hours, as shown in Figure 3.2 C. Specific experiments highlighting the influence of enzymes and thresholds on the oscillatory regime are shown in the Appendix, Sections 3.7.16 and 3.7.17. Loosely speaking, we could compare the concentration ratio of RNAP/RNase H (production/degradation ratio) to the "loop gain" of the system [123], and the thresholds as parameters introducing a delay in the system (the RNA species have to build up for some time and exceed the thresholds before they perform their activatory/inhibitory action on the switches). This analogy is inspired by the theoretical results in [11], and is currently being developed in further detail.

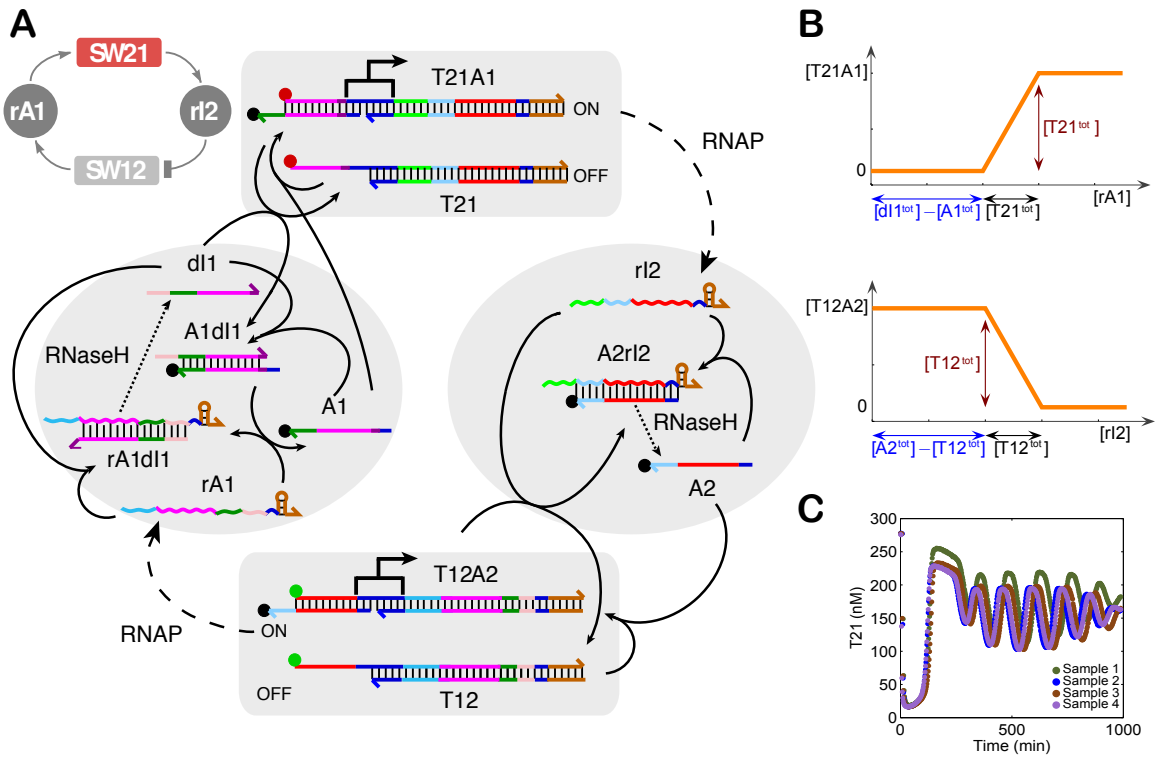


Figure 3.2: A. Operation scheme of the transcriptional oscillator system. Colors indicate complementary DNA and RNA domains. Sequences are given in Section 3.6 and detailed reaction diagrams can be found in Figures 3.13–3.14 of the Appendix, Section 3.7.2. When switch SW21 is turned on, RNA polymerase (RNAP) transcribes regulatory RNA (rI2) from the genelet template T21. RNA strand rI2 inhibits transcription from switch SW12 by removal of DNA strand A2 from template T12, resulting in an incomplete promoter region. On the other hand, RNA species rA1, which is transcribed from SW12, activates transcription from SW21 by releasing A1 from the A1·dI1 complex. RNA levels in the system are controlled by RNase H-mediated RNA degradation. By fluorescently labeling strand T21 with Texas Red or TYE665 (red dot), strand T12 with TAMRA or TYE563 (green dot), and activation strands A1 and A2 with Iowa Black RQ quenchers (black dots), the genelet states can be monitored by fluorescence measurements–high signals correspond to low transcription activity. B. Thresholds are set by adding threshold strands dI1 and A2 in excess over A1 and T12, respectively. In a typical experiment, the concentrations were $[\mathrm{T21^{tot}}] = 250\,\mathrm{nM}$, $[\mathrm{A1^{tot}}] = 250\,\mathrm{nM}$, $[\mathrm{dI1^{tot}}] = 700\,\mathrm{nM}$, $[\mathrm{T12^{tot}}] = 120\,\mathrm{nM}$, $[\mathrm{A2^{tot}}] = 500\,\mathrm{nM}$. C. Oscillator traces showing T21 levels for typical oscillations obtained in several, separate experiments. Note the good reproducibility of the oscillations. T12 has lower amplitude oscillations and is not shown. Figure adapted from [41]. All data shown were taken by me at Caltech.

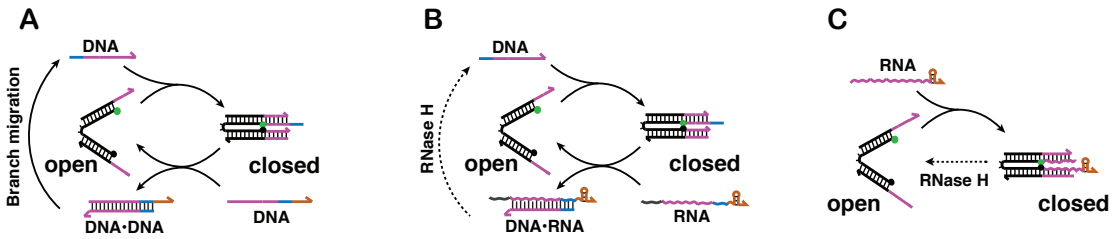


Figure 3.3: In the context of transctiptional circuits, the open–closed conformation of the molecular tweezers can be cycled through three basic reaction pathways. A. The tweezers can be closed by binding to a DNA species: if the closing strand is provided with a toehold, it can then be branch migrated from its complement, thereby opening the tweezers. The closing strand may be released in solution by another branch migration reaction (displacing species not shown) and recycled to close the tweezers. B. The DNA closing strand can be branch migrated by an RNA species, opening the tweezers, and released in solution after RNAse H degradation. C. An RNA species closes the tweezers, which are then opened by RNase H degradation.

The active or inactive state of the oscillator templates T12 and T21 can be measured using fluorescent probes attached to the 5' end of their non-template strand. When an activator strand (A1 or A2, respectively), 3' labeled with a quencher, binds to the template (forming transcriptionally active complexes T12·A2 or T21·A1), fluorescence is quenched. Low fluorescence therefore corresponds to an active genelet, whereas high fluorescence corresponds to an inactive genelet. Here, only the fluorescence traces of T21, shown in Figure 3.2 C, will be utilized. The oscillatory signal of T12 has a very limited amplitude and is therefore not considered for data processing; Section 3.7.6 in the Appendix is dedicated to this topic.

3.3.2 A simple load mechanism: molecular tweezers

As a load for the oscillator, we used the DNA tweezers system first described in [139]. The tweezers are a structure composed of three DNA strands, which bind to form two 18 base long double-stranded "arms" connected by a 4 base "hinge". The double-stranded arms can be extended with single-stranded "hands" of variable length, whose sequences can be freely designed to bind a desired target oligonucleotide. When bound to their target, the tweezers are in a closed conformation. As in previous works [139, 29], the tweezers are also labeled with a distinct fluorophore-quencher pair: open tweezers display a high fluorescence, while closed tweezers exhibit low fluorescence. The tweezer state is measured simultaneously with the T21 genelet state.

Figure 3.3 shows how the tweezers can be opened and closed in the context of transcriptional circuits. The tweezers can be closed by binding of a DNA strand, and opened through branch migration operated by DNA (A) or RNA (B). This can be considered a non-consumptive binding mechanism with respect to the closing molecule, which will be eventually released back in solution. Alternatively, the tweezers can be closed by RNA binding and opened through RNAse H degradation.

This mechanism can be considered consumptive, as the tweezers' target is effectively depleted. We have used the oligonucleotides of our molecular oscillator to open and close the tweezers with the three mechanisms shown in Figure 3.3, as described in detail in the next section.

3.3.3 Coupling the oscillator to the tweezers load: signal transmission and back-action

The transcriptional oscillator shown in Figure 3.2 is constituted by two switches and by several single-stranded oligonucleotide species: A1, dI1, rI1, A2, and rI2. Each of these single-stranded species has been used directly as a closing strand for the molecular tweezers, by redesigning the tweezers' arms. We can thus define a tweezers driving "mode" for each species being targeted. Table 3.1 summarizes the different direct coupling modes that have been tested. Detailed strand schemes and the reaction pathways that occur between the oscillator and the different tweezers modes are shown in the Appendix, Figures 3.15–3.18.

Table 3.1: Summary of Directly Coupled Tweezers Modes

.

Mode:	mode I	mode II	mode II*	mode III	mode IV
Target species:	dI1	A1	A2	rI2	rI1

We can evaluate the performance of each mode with respect to two criteria: (1) The amplitude of the oscillations transmitted to the tweezers. (2) The perturbation on the oscillator dynamics introduced by the tweezers. In the brief analysis that follows, I will refer to the features of the simple model considered in Section 3.2. Such model in fact provides several insights on the behavior of the implementation with transcriptional circuits. Clearly, the experimental system is far more complex than the intuitive model: therefore, all analogies must be drawn with some caution.

Signal transmission

Figure 3.4 shows the schemes and the experimental results for modes I, II, and III. Modes I and II were the most successful modes in terms of load amplitude, while mode III represents an example of a "failure" mode, given the modest amplitude of the tweezers' oscillations. Modes II* and IV suffered from the same amplitude failure, and are therefore not shown in this section: the corresponding data sets are in the Appendix, Figures 3.43 and 3.45.

The phase relationship (Figure 3.4 C and H) between the oscillator T21 fluorescent signal and the tweezers signal can be explained easily. In mode I, the tweezers are opened by rA1 binding to the dI1 toehold. In the oscillator circuit, rA1 binds to the complex releases A1·dI1, thereby releasing A1 which can activate (quench) SW21. Thus, high tweezers fluorescence (open state) corresponds to low SW21 fluorescence, and vice versa, resulting in a phase shift by a half period of the oscillator. In

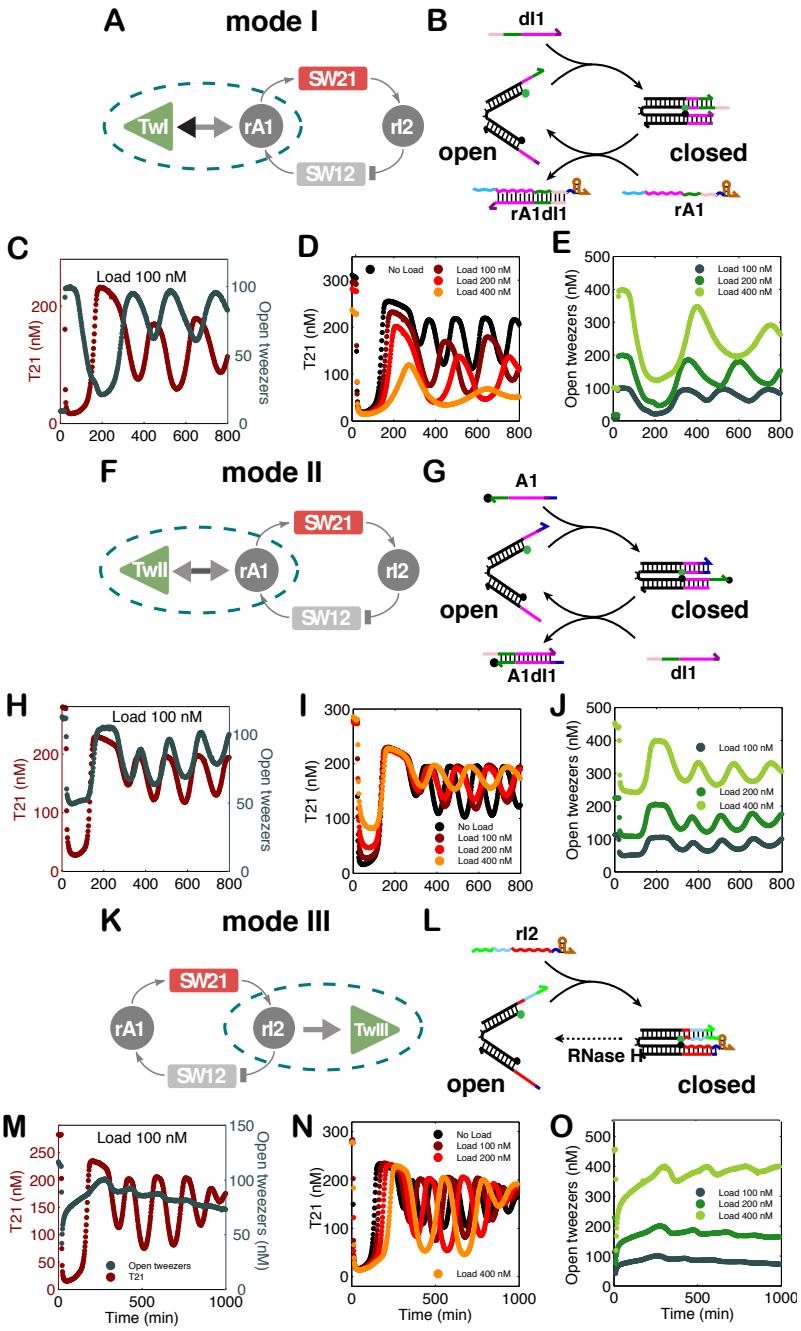


Figure 3.4: Three different ways of coupling a load to the oscillator. A. In the simple model scheme, mode I couples to the rA1 node. B. Molecularly, mode I uses dI1 to close the DNA tweezers, and rA1 to open them. C. Oscillator traces (T21 levels) and mode I tweezers oscillations superimposed for a load of 100 nM tweezers. D. Load dependence of the core oscillator (load 0–400 nM) E. Corresponding oscillations of the tweezers load. F. In the simple model scheme, mode II also couples to the rA1 node. G. On the molecular level, mode II uses A1 to close the tweezers and dI1 to open them. H. Oscillator traces and mode II tweezers oscillations superimposed for 100 nM load. I and J. Oscillations of the core oscillator and the tweezers load for different load concentrations. K. In the simple model, mode III couples to the rI2 node. L. Mode III uses rI2 to close the tweezers and RNaseH to open them. M. Oscillator traces and mode III tweezers oscillations superimposed for 100 nM load. N. Oscillations of the core oscillator for different load concentrations. O. Variation of the mode III tweezers signal for different load concentrations. Figure adapted from [41]. All data shown were taken by me at Caltech.

mode II, A1 molecules are also released by rA1 in the same way as in mode I. In this case, however, A1 is used to close the tweezers (quenched state): therefore, the fluorescent signals of the oscillator and the tweezers are in phase.

The simplified oscillator model can provide a high-level interpretation of these results. Refer to the quasi-steady-state approximation of the load dynamics (3.3):

$$\widehat{[\mathrm{L^a}]}(\mathrm{t}) = [\mathrm{L^{tot}}] \left(1 - \frac{\mathrm{k_r}}{\mathrm{k_r} + \mathrm{k_f} \cdot \mathrm{u(t)}}\right),$$

where u(t) is the input concentration. Assuming that $\mathrm{u(t)} = \mathrm{A_0} + \mathrm{A_1} \sin \omega \mathrm{t}$, we found that the amplitude of the load oscillations depends strongly on the mean $\mathrm{A_0}$ and amplitude $\mathrm{A_1}$ and on the effective coupling constant $\mathrm{k_r/k_f}$ (see Figure 3.1 I and J). In this respect, mode I is driven by the available dI1 concentration, which can have a maximum amplitude $\mathrm{A_1}$ of around 700 nM, while in mode II the amplitude of the activator A1 is at most 250 nM. Indeed, mode I appears to work better than mode II.

Regarding the failure modes II*, III, and IV, in all cases the tweezers remained almost fully open. For mode II*, we know that at our operating point rI2 concentrations are high, presumably sequestering A2 most of the time (in fact, on average only 20 nM of SW12 are active, see Section 3.7.6 in the Appendix). Hence, the closing strand concentrations are simply too low in this mode to actuate tweezers efficiently (low $\mathrm{A_0}$ and $\mathrm{A_1}$). For modes III and IV, I examined the interactions between enzymes and tweezers in the absence of the core oscillator (see Section 3.7.15 in the Appendix). We were concerned that RNase H could only partially degrade RNA fuel strands when complexed with the DNA device, resulting in "poisoned" DNA tweezers that are always open. Surprisingly, several (if not all) tweezers designs served as substrates for promiscuous RNAP activity that caused tweezers opening. In the case of modes I and II, RNase H eliminated this effect and restored function to the DNA closing strand. However, the RNA closing strands for modes III and IV were not fully effective, even in the presence of RNase H, helping to explain the poor performance of these modes when driven by the oscillator circuit. In addition to partial degradation of RNA closing strands, it was found in [64] that even in the core oscillator, RNase H leaves partial degradation products that can accumulate to reach micromolar concentrations; these are predicted to have sequences complementary to one hand of the tweezers for both modes III and IV (see Figures 3.17 and 3.18). Carrying on the simple model analogy, a high mean RNA concentration (high $\mathrm{A_0}$) relative to the tweezers concentration also contributes to the observed small oscillations in the load.

Retroactivity

Referring to Figure 3.4, increasing the load generally affects both amplitude and frequency of the oscillator. The influence on the oscillations is smaller for coupling mode II (Figure 3.4 I) than for coupling mode I (Figure 3.4 D), both in terms of amplitude and period perturbation: this is

quantified more specifically in Figure 3.7. Making another analogy with our simplified model, one could classify modes I and II with respect to the release speed of their input species. In mode I, strand dI1 is sequestered by the tweezers: before being released again it must be displaced by rA1 and subsequently processed by RNase H within the dI1·rA1 complex (one branch migration reaction, and one enzymatic degradation step). In mode II, strand A1 bound to the tweezers is first displaced by dI1, and subsequently released by rA1 (two branch migration reactions). Although not rigorous, we can make an analogy with the simple model in terms of consumptive versus non-consumptive load. Loosely speaking, if the oscillator species is released slowly, in the limit we fall into the consumptive case, which has a high back-action compared to the non-consumptive case. Presumably, the two branch migration steps necessary to release A1 in mode II are faster than the branch migration and enzymatic processing required to release dI1 in mode I. Hence the smaller back-action of mode II relative to mode I.

Even though coupling modes III and IV do not lead to a satisfactory oscillatory actuation of the tweezers, the oscillator dynamics are strongly affected by the presence of the DNA device (Figure 3.4 N, and Figures 3.44 and 3.45 in the Appendix). This is consistent with the hypothesis that partially-degraded transcripts bind to one hand, keeping the tweezers open, while the other hand is still active in binding new transcripts and serving as a substrate for RNase H (pathway shown in Figure 3.17 and 3.18 of the Appendix). Finally, RNase H degradation results in a consumptive processes that permanently removes the RNA oscillator species from the system.

As explained in detail in the Appendix, Section 3.7.12, due to the specific concentration dynamics of the core oscillator strands, mode IV affected the oscillations more drastically than all other modes. For a similar reason, mode II* had only a negligible effect, but also resulted in very inefficient load coupling.

Most of the behavior of the oscillator under load can also be heuristically understood in terms of changes of threshold strand concentrations [A2] and [dI1] caused by the reaction pathways with the tweezers. For instance, in mode I, tweezers are closed by dI1 and opened in a strand displacement reaction by rA1 (see Figure 3.4 B). As can be seen from the core oscillator reaction scheme in Figure 3.2 A, strand A1 is similarly bound by dI1 and freed by a strand displacement reaction with rA1. An increase in the concentration of mode I tweezers therefore is roughly equivalent to an increase in [A1], which in turn corresponds to an effective reduction of the threshold set by [dI1]. By contrast, in mode II (Figure 3.4 G) a fraction of activator strands A1 is bound to the tweezers rather than to template T21. The effective reduction in [A1] should therefore be analogous to an increase in threshold [dI1]. Finally, an increase of tweezers concentrations in mode III (Figure 3.4 L) is similar in effect as an increase in [A2], as rI2 interacts with A2 in the core oscillator. We experimentally confirmed this heuristic argument by changing the concentrations of threshold strands in the core oscillator and found that indeed this reproduces most of the general trends in amplitude and period

observed in Figures 3.4 D, I, and N (see Appendix, Section 3.7.17). Similar reasoning can be applied to all other coupling modes.

Regarding retroactivity of the load in terms of the simplified model, an effective change of thresholds as discussed above will change parameters α, β, and also the steepness of the response functions (see Figure 3.2 B) that define the oscillatory domain of the system.

3.3.4 Implementing an insulation component

The data in Figure 3.4 emphasize that the back-action, or retroactivity, caused by the tweezers is proportional to the total amount of load present. Following the idea presented in our simple model, it is possible to reduce such retroactivity effects by the isolation of source components from downstream loads using buffering and amplification stages. A simple implementation of this strategy for our oscillator system is displayed in Figure 3.5, and has been termed "mode V".

A third genelet, which we call "insulator" genelet, is operated in parallel with oscillator switch SW12, i.e., it is activated by A2 and deactivated by rI2 (Figure 3.5 A). The insulator is designed to produce a new RNA species InsOut, which in turn displaces the tweezers mode V closing strand, termed TwCls. Figure 3.5 C–D shows load experiments where tweezers and insulator genelet are added in a ratio of 4:1. Transcription from the insulator genelet acts as an amplifier: a small increase in the concentration of the insulator genelet (which incurs a small disruption of the core oscillator dynamics) results in a large increase in the RNA available to drive the tweezers. Furthermore, this design effectively decouples the tweezers design and operation from the oscillator. Even when there are more tweezers than can be effectively driven, the absence of specific interactions between the tweezers and the core oscillator strands leaves the core oscillator dynamics relatively intact (see Figure 3.46 in the Appendix). These features allow mode V to drive much larger loads than the direct coupling modes.

As shown in Figure 3.5 C, even excessive loading of the circuit with 800 nM tweezers (and, hence, 200 nM insulator genelet) does not affect the oscillator dynamics significantly. At the same time, the amplitude of the insulated DNA tweezers is satisfactory (Figure 3.5 D). In contrast, tweezers in modes I and II (Figure 3.4 D–E and I–J), strongly affected the load at much lower concentrations. The overall comparison of direct and insulated modes in terms of load amplitude and perturbation of oscillator period and amplitude is shown in Figure 3.7.

A second insulation pathway was tested, where the insulator genelet was operated in parallel to SW21. This mode was termed mode V* (see Appendix, Figures 3.19 and 3.47), and was successful in reducing the period perturbation; however, the amplitude of the core oscillator was increased, even at very small insulator amounts. One possible explanation for this unexpected behavior is the high leak and transcription rate exhibited by the mode V* genelet (Section 3.7.13 of the Appendix). Higher affinity for RNAP and very large amounts of InsOut RNA bound to its target TwCls (additional

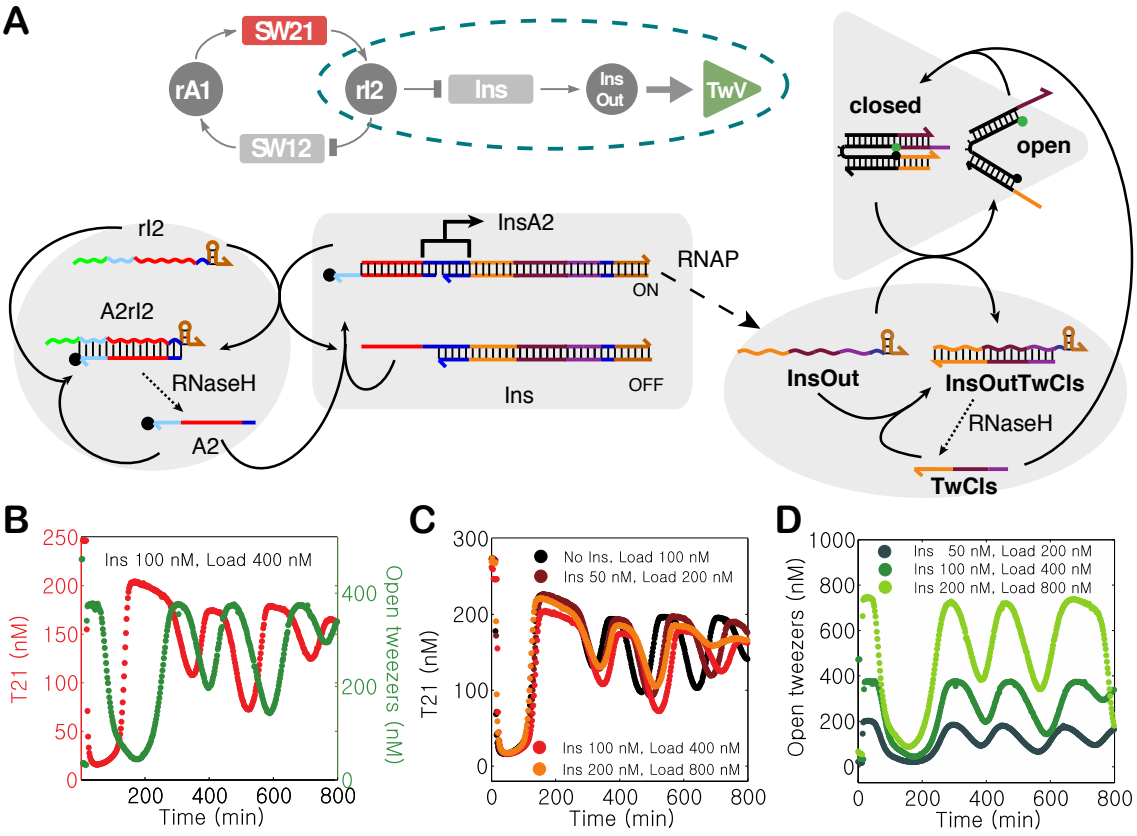


Figure 3.5: An insulator circuit (mode V coupling). A. Insulator genelet Ins is operated in parallel with SW12. It is activated by A2 and deactivated by rI2. Transcription of Ins results in RNA signal InsOut which opens tweezers previously closed by DNA strand TwCls. ("Load" for mode V is defined as closed tweezers with a 50 nM excess of TwCls, in contrast to modes I–IV where the load consists only of open tweezers.) The RNA part of hybrid duplex TwCls·InsOut is degraded by RNase H, resulting in free TwCls. This operation principle is analogous to mode I tweezers. B. Oscillator (red) and tweezers (green) traces for 100 nM insulator genelet and 400 nM tweezers load. C. Core oscillator traces for 0 nM Ins and 100 nM tweezers load (black), and 200 nM (dark red), 400 nM (red), and 800 nM (orange) tweezers load and a 4:1 ratio of tweezers:Ins. D. Tweezers signal for 200 nM (dark green), 400 nM (green), and 800 nM (light green) tweezers load. Figure adapted from [41]. All data shown were taken by me at Caltech.

substrate for RNase H) could represent a significant burden for the enzymes shared in the circuit. Therefore, it is plausible that the pathways introduced by mode V* deplete the overall enzyme levels and introduce a different type of "global" retroactivity.

3.4 Modeling

The simple model proposed in Section 3.2 is a useful representation of the main features of a chemical load coupled to an oscillating system. However, as noted before, the simple model cannot capture the complexity of the system's experimental implementation nor all the data trends observed.

A more detailed model for the oscillator and tweezers can be constructed by listing the most significant biochemical pathways occurring in the system, and systematically deriving the corresponding set of ODEs. Such equations can be then numerically solved and their parameters fitted using standard software tools. The detailed numerical models' derivation and corresponding data fits were carried out by J. Kim and R. Jungmann.

First, a model was generated and fitted for the transcriptional oscillator in isolation, to capture the trends observed when varying the DNA thresholds concentrations. Then, the data for modes I, II, and V were also fitted in a similar manner, clearly including additional parameters describing the load and insulator related reactions. In the Appendix, the captions for Figures 3.13–3.21 lists the reaction pathways used to construct the detailed numerical models for the system.

The detailed models overall semi-quantiatively reproduce the experimental data, as shown in Figure 3.6. One notable discrepancy is that frequency change introduced by the load in modes I and II is comparable in simulations, while experiments show that such change is much greater in mode I than in mode II. Sources of uncertainty for the model fits are mainly: 1) Unmodeled side reactions, including spurious hybridization, incomplete transcription and degradation products, and dimerization of the tweezers [139]. 2) Uncertainty and variability of the enzymatic activity parameters. For instance, the rate of RNAP synthesis varies over time, showing an initial burst stage [56]; oxidation of Cys residues produces a continuous decrease in activity over time [23, 65]. Moreover, after termination the enzyme has to revert to a competent state ("recycling") [35].

Further details regarding the challenges encountered in the detailed models fits can be found in [41] and its Supplementary Appendix, Section 24 and following.

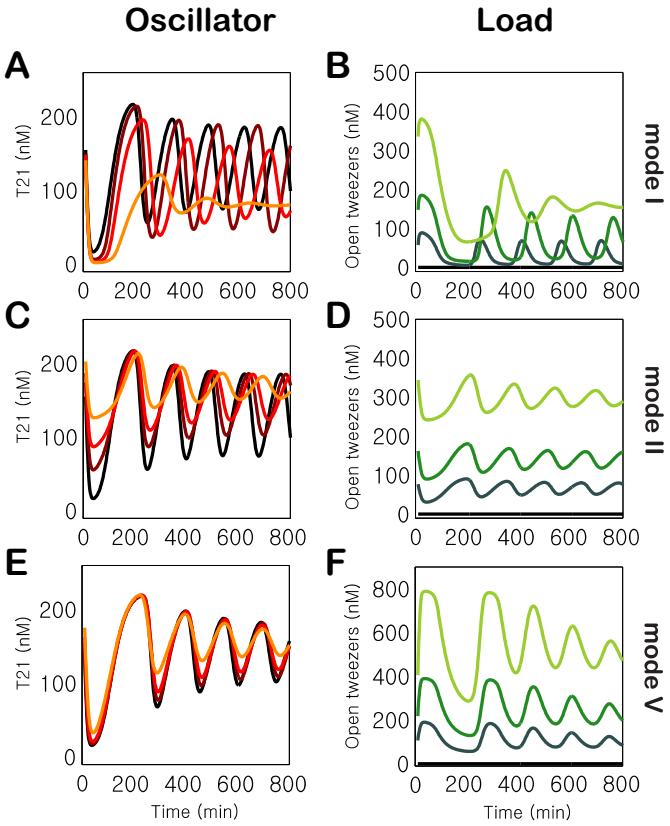


Figure 3.6: Simulations of the core oscillator and oscillator driving loads, using the mechanistic mass-action model described in the SI Appendix, Sections 24–33, for initial DNA concentrations identical to those in several experiments. A and B. See Figure 3.4 D and E. C and D. See Figure 3.4 I and J. E and F. See Figure 3.5 C and D. Figure adapted from [41]. These simulations were done by J. Kim.

3.5 Discussion

In this work, a synthetic transcriptional oscillator has been used to drive conformational changes of a DNA nanodevice. Such an oscillator is constructed from two genelets interconnected through their transcripts, with an overall negative feedback loop. The output rI2 of switch SW21 inhibits SW12, while the output rA1 of SW12 activates SW21. The DNA species A1, A2, and dI1 represent activation/inhibition thresholds. This transcriptional oscillator has been directly coupled to the well-known DNA tweezers system [139] through all of its single-stranded components, yielding coupling mode I (coupling to dI1), mode II, mode III (rI2), mode II* (A2), and mode IV (rA1). Each connection channel resulted in different efficiency of the transmission of the oscillatory signal and different perturbations (or retroactivity) of the oscillator dynamics. With respect to signal propagation, mode I is the best. The retroactivity of the different modes is summarized in Figure 3.7: in all coupling modes the oscillations tend to get slower with increasing load (Figure 3.7 A). This might be due to the fact that the parameters of the oscillator in isolation were initially optimized for fast oscillations. Hence, any perturbation typically moves the dynamics away from these optimum settings. For modes I and II the amplitude of the oscillations decreases with increasing load, while for mode III the amplitude increases (Figure 3.7 B). Mode I seems to affect the oscillator dynamics most strongly; presumably this is related to the fact that in this mode a larger fraction of the load is driven than in mode II (compare also amplitudes in Figure 3.4 E and J). The period and amplitude change were therefore plotted with respect to the "effective load" – the maximum amplitude swing induced in the tweezers. When only the influence of the effective load is considered, modes I and II affect the oscillator similarly (Figure 3.7 C, D). One of the most important contributions of this work is the implementation of an insulator genelet (mode V). The insulator acts as an amplifier that diverts a small amount of an oscillator species and amplifies it to drive downstream load processes. As shown in Figure 3.7, the insulator renders the system almost insensitive with respect to load.

Several features of this system can be understood on the basis of the simple theoretical model proposed at the beginning of this chapter. For example, this simple model relates the load coupling efficiency to the mean value and amplitude of the input oscillating species, it predicts the reduction of the oscillatory space in the presence of a load process, and it predicts that retroactivity can be remedied by an insulator concept. The simple model, however, does not provide a quantitative description of the experiments. To this purpose, a mechanistic numerical model was formulated that includes most of the occurring chemical reactions. This detailed model semi-quantitatively reproduced all of the experimental data with a single set of physically acceptable parameters.

Our results represent a step forward in two directions:

1. We improved our general understanding of the challenges arising when coupling biochemical systems: if mass is conserved, transmitting a signal means simultaneously introducing a perturbation

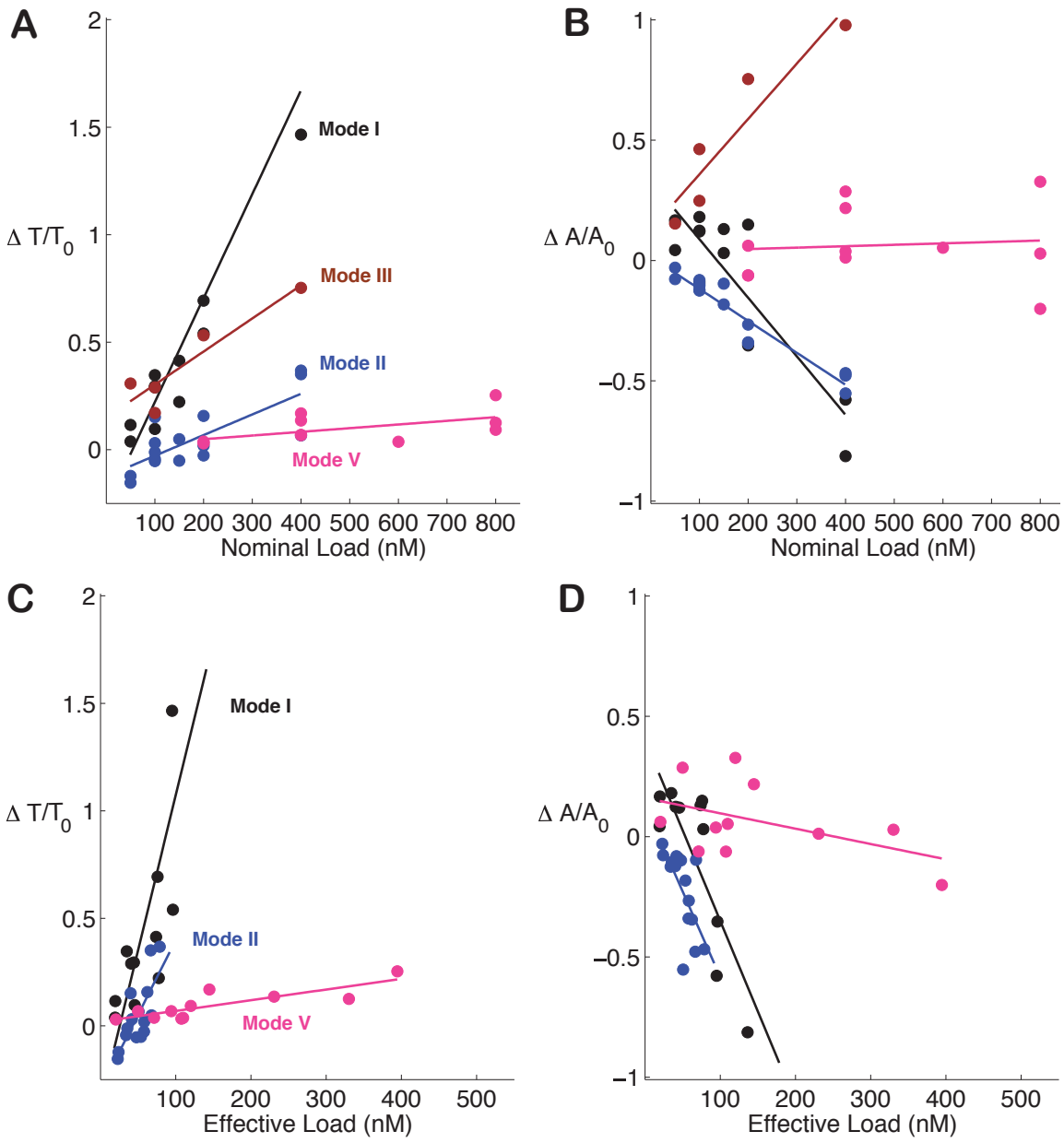


Figure 3.7: Analysis of the influence of load on the oscillation amplitude and period. A and B. Relative period change as a function of the nominal (A) and effective (B) load concentrations. C and D. Relative amplitude change as a function of the nominal (C) and effective (D) load concentrations. Figure adapted from [41]. The data points shown combine experimental results collected both at Caltech and TUM on modes I, II, and III. A complete overview of all coupling modes data is found in the Appendix, Section 3.7.18. The equivalent of Figure 3.7, including all other modes presented in this thesis, is also in the Appendix, Figure 3.32.

in the original chemical source of such signal. Biochemical devices that overcome mass conservation through amplification of their input are a possible solution to this problem.

2. A programmable clock has been tested for potential use in the context of orchestrating synthetic molecular processes and artificial cells [91, 72]. Our synthetic gene regulatory system with only two enzymes and a handful of DNA oligonucleotides already results in complex molecular interactions. Brute-force attempts to design, analyze, and experimentally characterize molecular systems may not be successful. In synthetic biology, there is a need for systems design principles to achieve robustness, fault tolerance, and avoidance of undesired dynamics. In this study, a first step towards such robustness has been demonstrated by the insulator circuit. The next steps include the construction of an improved system that is not sensitive to enzyme activities, that can provide a stable rhythm regardless of the source or age of the enzyme batch used and that is separately tunable in frequency and amplitude. Future insulator subcircuits should be characterized to drive dynamically changing loads.

In summary, our synthetic oscillator coupled to a load represents one of the first realizations of an *in vitro* biochemical clock that is used to drive other processes. This could serve as a model system for the study of modularity, coupling of subcircuits, and robustness in biochemical networks.

3.6 Materials and methods

DNA oligonucleotides and enzymes

The DNA sequences were ordered from IDT DNA (Coralville, IA, USA). (For the experiments run at TUM, strands were ordered from IDT DNA (Belgium), IBA (Göttingen, Germany), or biomers.net (Ulm, Germany).) RNA polymerase was purchased from Ambion, Austin, TX, as part of the T7 Megashortscript kit (#1354), or from Epicentre Biotechnologies, Cat. n. TM910K (200 U/μl). *E. coli* RNase H was purchased from Ambion (#2292). Inorganic lyophilized pyrophosphatase was purchased from Sigma Aldrich, Cat. n. I1891-100UN.

Transcription protocol

Transcription was run at 37°C, either using the T7 Megashortscript kit (#1354), Ambion, Austin, TX (which includes T7 RNA polymerase enzyme mix, transcription buffer and rNTPs), or a transcription mix composed of: 1x Transcription Buffer and 10 mM dithiothreitol (DTT), (Epicentre Biotechnologies, Cat. n. BP1001), 7.5 mM each rNTP (Epicentre Biotechnologies, Cat. n. RN02825), 35 mM $MgCl_2$ and 0.015 U/μl pyrophosphatase (resuspended in Tris HCl 20 mM, pH 7.2, 50% glycerol (v/v)), and T7 RNAP purchased from Epicentre Biotechnologies, Cat. n. TM910K (200 U/μl). More details on the transcription protocol and sample preparation are given in the Appendix, Section 3.7.4.

Data acquisition

Concentrations of nucleic acids were determined by absorption measurements (Nanodrop 2000c, Thermo Scientific, DE, USA), using sequence-dependent extinction coefficients.

All fluorescence experiments were performed on a Horiba/Jobin Yvon Fluorolog 3 system, using 45 μL sample chamber quartz cuvettes. Fluorescence emission of the dye/quencher pair labelled DNA strands was recorded every two minutes. A sample temperature of 37°C was maintained using a water circulation thermostat. Excitation/emission for TYE563 (T12) was set to its maxima at 549–563 nm, those for TYE665 (T21) at 645–665 nm, and Rhodamine Green (Tweezers) emission and excitation maxima were set at 504 nm–531 nm, according to the recommendation of the supplier IDT DNA. The raw fluorescence measurements recorded with the Horiba Fluorolog 3 were converted into molar concentrations according to the procedure described in the Appendix, Section 3.7.5.

Denaturing polyacrylamide gels were either cast in house, using 8%–10% 19:1 acrylamide:bis and 7 M urea in TBE buffer, 100 mM Tris, 90 mM boric acid, 1 mM EDTA, or purchased pre-cast from Invitrogen. Gels were run at 23°C for 45–50 min with 10 V/cm in TBE buffer. Samples were loaded using Xylene Cyanol FF dye. For quantitation, denaturing gels were stained with SYBR Gold (Molecular Probes, Eugene, OR; #*S*-11494). In the control lane a 10-base DNA ladder (Invitrogen, Carlsbad, CA; #1082-015) was utilized. Gels were scanned using the Molecular Imager FX (Biorad, Hercules, CA).

Numerical simulations

The numerical analysis on the simple model considered in Section 3.2 was performed using several scripts written by me in MATLAB. The experimental data sets were fitted by J. Kim and R. Jungmann, using ordinary differential equation models derived from first principles. Specifically, the DNA and RNA hybridization reactions, branch migration reactions, and Michaelis–Menten enzyme reactions for the core oscillator are identical to those used in the extended model of the Design I oscillator in [64]. The standard mass action kinetics and Michaelis–Menten approximations were used to convert these reaction equations to a set of ordinary differential equations as outlined in [64]. In the Appendix, Section 3.7.2, a list of the relevant interactions among nucleotides is reported, specifying which reactions were used to build numerical fits. More details on the numerical simulations can be found in the Supplementary Appendix of [41].

Oligonucleotide sequences

Oscillator sequences: The oscillator sequences are taken from [61] and [64].

T12-t 5'-TTT CTG ACT TTG TCA GTA TTA GTG TGT AGT AGT AGT TCA TTA GTG TCG TTC GTT CTT TGT TTC TCC CTA TAG TGA GTC G

T12-nt 5'-TYE563-AAG CAA GGG TAA GAT GGA ATG ATA ATA CGA CTC ACT ATA GGG AGA AAC AAA GAA CGA ACG ACA CTA ATG AAC TAC TAC TAC ACA CTA ATA CTG

ACA AAG TCA GAA A

T21-t 5'-TTT CTG ACT TTG TCA GTA TTA TCA TTC CAT CTT ACC CTT GCT TCA ATC CGT TTT ACT CTC CCT ATA GTG AGT CG

T21-nt 5'-TYE665-CAT TAG TGT CGT TCG TTC ACA GTA ATA CGA CTC ACT ATA GGG AGA GTA AAA CGG ATT GAA GCA AGG GTA AGA TGG AAT GAT AAT ACT GAC AAA GTC AGA AA

dI1 5'-GTG TGT AGT AGT AGT TCA TTA GTG TCG TTC GTT CAC AG

A1 5'-TAT TAC TGT GAA CGA ACG ACA CTA ATG AAC TAC TAC-Iowa Black RQ

A2 5'-TAT TAT CAT TCC ATC TTA CCC TTG CTT CAA TCC GT-Iowa Black RQ

rA1 (RNA) 5'-GGG AGA AAC AAA GAA CGA ACG ACA CUA AUG AAC UAC UAC UAC ACA CUA AUA CUG ACA AAG UCA GAA A

rI2 (RNA) 5'-GGG AGA GUA AAA CGG AUU GAA GCA AGG GUA AGA UGG AAU GAU AAU ACU GAC AAA GUC AGA AA

For the data collected in sets 1, 2 and 3, collected at TUM (for details on the data sets, refer to Section 3.7.3), T12-nt was labeled with the dye TAMRA instead of TYE563 and T21-nt with Texas Red instead of TYE665.

For set 6, due to technical constraints of the supplier IDT DNA, T21-nt and T12-nt were, respectively, shortened 1 and 6 bases from the 3' end, to have a length of 100 bases. These modifications do not alter the regulatory domains of the transcripts. Also the full length of the main transcription products was not affected, as verified by gel electrophoresis.

Tweezers subsystem (mode I–mode IV):

TW A 5'-RhodamineGreen-TGC CTT GTA AGA GCG ACC ATC AAC CTG GAA TGC TTC GGA T-BHQ1

TW B (I) 5'-CTG TGA ACG AAC GAC ATC CGA AGC ATT CCA GGT

TW C (I) 5'-GGT CGC TCT TAC AAG GCA ACT AAT GAA CTA CTA

TW B (II) 5'-GTT CAT TAG TGT CGT ATC CGA AGC ATT CCA GGT

TW C (II) 5'-GGT CGC TCT TAC AAG GCA TCG TTC ACA GTA ATA

TW B (II*) 5'-TGA AGC AAG GGT AA ATC CGA AGC ATT CCA GGT

TW C (II*) 5'-GGT CGC TCT TAC AAG GCA GAT GGA ATG ATA ATA

TW B (III) 5'-TAT CAT TCC ATC TTA CCC TAT CCG AAG CAT TCC AGG T

TW C (III) 5'-GGT CGC TCT TAC AAG GCA TGC TTC AAT CCG TTT TAC T

TW B (IV) 5'-GTA GTA GTT CAT TAG ATC CGA AGC ATT CCA GGT

TW C (IV) 5'-GGT CGC TCT TAC AAG GCA TGT CGT TCG TTC TTT GTT T

Tweezers insulating subsystem (mode V):

Ins*-nt 5'-CAT TAG TGT CGT TCG TTC ACA GTA ATA CGA CTC ACT ATA GGG AGA TCA AAT TTA CAA CGC AAC TAA CAT ATA ATC GAA GAC TTA ATA CTG ACA AAG TCA

Ins*-t 5'-TTT CTG ACT TTG TCA GTA TTA AGT CTT CGA TTA TAT GTT AGT TGC GTT GTA AAT TTG ATC TCC CTA TAG TGA GTC G

Ins-nt 5'-AAG CAA GGG TAA GAT GGA ATG ATA ATA CGA CTC ACT ATA GGG AGA TCA AAT TTA CAA CGC AAC TAA CAT ATA ATC GAA GAC TTA ATA CTG ACA AAG TC

Ins-t 5'-TTT CTG ACT TTG TCA GTA TTA AGT CTT CGA TTA TAT GTT AGT TGC GTT GTA AAT TTG ATC TCC CTA TAG TGA GTC G

InsOut (RNA) 5'-GGG AGA UCA AAU UUA CAA CGC AAC UAA CAU AUA AUC GAA GAC UUA AUA CUG ACA AAG UCA GAA A

TwCls 5'-AAG TCT TCG ATT ATA TGT TAG TTG CGT TGT AAA TTT GA

TW B (V) 5'-TCA AAT TTA CAA CGC ATC CGA AGC ATT CCA GGT

TW C (V) 5'-GGT CGC TCT TAC AAG GCA AAC TAA CAT ATA ATC

Note that Ins*-t and Ins-t are identical. Mode I tweezers are constructed by hybridizing the arms of the central strand (TW A) with the two flanking strands TW B (I) and TW C (I). Mode II - V tweezers are constructed analogously. The 3' ends of the insulators non-template strands are a few bases shorter than their template strand. This is shown in Figures 3.19 and 3.20. This did not affect their regulatory domains nor the length of their transcription products.

3.7 Appendix

3.7.1 Simple model for the oscillator: load coupling and insulation

In this section, we will provide further details on the simple oscillator model presented in the main paper.

3.7.1.1 A simple model for the transcriptional oscillator and its non-dimensional version

Equations (3.5)–(3.8) were proposed in [64] as a simple model for our transcriptional oscillator:

$$\frac{\mathrm{d[rA1]}}{\mathrm{dt}} = \mathrm{k_p[SW12] - k_d[rA1]}, \tag{3.5}$$

$$\tau\frac{\mathrm{d[SW21]}}{\mathrm{dt}} = [\mathrm{SW21^{tot}}]\frac{\frac{[\mathrm{rA1}]^{\mathrm{m}}}{\mathrm{KA^m}}}{1+\frac{[\mathrm{rA1}]^{\mathrm{m}}}{\mathrm{KA^m}}} - \mathrm{[SW21]}, \tag{3.6}$$

$$\frac{\mathrm{d[rI2]}}{\mathrm{dt}} = \mathrm{k_p[SW21] - k_d[rI2]}, \tag{3.7}$$

$$\tau\frac{\mathrm{d[SW12]}}{\mathrm{dt}} = [\mathrm{SW12^{tot}}]\frac{1}{1+\frac{[\mathrm{rI2}]^{\mathrm{n}}}{\mathrm{KI^n}}} - \mathrm{[SW12]}. \tag{3.8}$$

A scheme representing the above equations is shown in Figure 3.8 A. The species rA1 and rI2 are RNA molecules that interact through two "genelet" switches that produce them, respectively, SW12 and SW21. In particular, rA1 is an activator for SW21, while rI2 is an inhibitor for SW12. The effectiveness of the RNA species in activating or repressing the switches is modulated by the thresholds KA and KI, and by the Hill coefficients m and n. The relaxation constant τ scales the speed of the switches' dynamics. Unless otherwise noted, from now on the operating point of this oscillator model is defined by the parameters $\mathrm{k_p} = 0.05\,/\mathrm{s}$, $\mathrm{k_d} = 0.002\,/\mathrm{s}$, KA=KI=.5 μM, $[\mathrm{SW21^{tot}}] = [\mathrm{SW12^{tot}}] = 100\,\mathrm{nM}$, m=n=5, $\tau = 500\,\mathrm{s}$. Figure 3.8 B and C show the system trajectories generated using the MATLAB `ode23` routine from initial conditions $[\mathrm{rA1}](0) = 0\,\mu\mathrm{M}$, $[\mathrm{SW21}](0) = 0\,\mathrm{nM}$, $[\mathrm{rI2}](0) = 0\,\mu\mathrm{M}$, $[\mathrm{SW12}](0) = 100\,\mathrm{nM}$.

Non-dimensional model

The above model can be mapped to a set of non-dimensional differential equations as follows:

Non-dimensional variables

$$\mathrm{x} = \frac{[\mathrm{rA1}]}{\mathrm{KA}} \quad \mathrm{v} = \frac{[\mathrm{SW21}]}{[\mathrm{SW21^{tot}}]}$$
$$\mathrm{y} = \frac{[\mathrm{rI2}]}{\mathrm{KI}} \quad \mathrm{u} = \frac{[\mathrm{SW12}]}{[\mathrm{SW12^{tot}}]}$$

Time rescaling

$$\tilde{\mathrm{t}} = \mathrm{t}/\tau$$

Non-dimensional parameters

$$\alpha = \frac{\mathrm{k_p}\,[\mathrm{SW12^{tot}}]}{\mathrm{k_d}\,\mathrm{KA}} \quad \beta = \frac{\mathrm{k_p}\,[\mathrm{SW21^{tot}}]}{\mathrm{k_d}\,\mathrm{KI}}$$
$$\gamma = \frac{1}{\mathrm{k_d}\,\tau}$$

Non-dimensional equations in $\tilde{t}$

$$\gamma\,\dot{\mathrm{x}} = \alpha\mathrm{u} - \mathrm{x}$$
$$\dot{\mathrm{v}} = \frac{\mathrm{x^m}}{1+\mathrm{x^m}} - \mathrm{v}$$
$$\gamma\,\dot{\mathrm{y}} = \beta\mathrm{v} - \mathrm{y}$$
$$\dot{\mathrm{u}} = \frac{1}{1+\mathrm{y^n}} - \mathrm{u}$$

Given our choice of the parameters, $\gamma = 1$. A value of $\gamma \approx 1$ is required to achieve oscillations, as found in [64].

Existence of periodic orbits

It is convenient to investigate the existence of periodic orbits by taking into consideration the non-dimensional model. We will start by making the following observations:

1. The equilibrium of each variable x, v, y, and u depends monotonically on its input. For example, given a fixed input $\bar{\mathrm{u}}_1$, the equilibrium of $\bar{\mathrm{x}}$ is $\bar{\mathrm{x}}_1 = \alpha\bar{\mathrm{u}}_1$. For any $\bar{\mathrm{u}}_2$, $\bar{\mathrm{u}}_2 > \bar{\mathrm{u}}_1$, then $\bar{\mathrm{x}}_2 > \bar{\mathrm{x}}_1$.

2. The trajectories of this system are always bounded. In fact, the switches concentration is bounded: u, v $\in [0, 1]$. The dynamics of x and y are exponentially stable, given a constant bounded input.

3. The system admits a unique equilibrium. In fact, by setting to zero the non-dimensional dynamics, we can derive the following expressions for the system nullclines:

$$\bar{\mathrm{x}} = \alpha\frac{1}{1+\bar{\mathrm{y}}^{\mathrm{n}}}, \qquad \bar{\mathrm{y}} = \beta\frac{\bar{\mathrm{x}}^{\mathrm{m}}}{1+\bar{\mathrm{x}}^{\mathrm{m}}}.$$

   The above curves are monotonic and intersect in a single point, as shown in Figure 3.8 D. Therefore, the system admits a single equilibrium.

Based on the observations above, we can invoke the Mallet-Paret theorem [76]. This theorem is the extension of the Poincaré-Bendixson theorem to dimension higher than 2, and is valid when the

system dynamics are monotonic and cyclic, as in our case.

Based on the theorem, if observations 1, 2, and 3 are true, and if the unique admissible equilibrium of the system is unstable, then the system must admit a periodic orbit. We can verify the stability properties of the equilibrium by inspecting the eigenvalues of the Jacobian matrix at the equilibrium:

$$\mathrm{J} = \begin{bmatrix} \frac{\partial \mathrm{f}_1}{\partial \mathrm{x}} & \frac{\partial \mathrm{f}_1}{\partial \mathrm{v}} & \frac{\partial \mathrm{f}_1}{\partial \mathrm{y}} & \frac{\partial \mathrm{f}_1}{\partial \mathrm{u}} \\ \frac{\partial \mathrm{f}_2}{\partial \mathrm{x}} & \cdots & \cdots & \frac{\partial \mathrm{f}_2}{\partial \mathrm{u}} \\ \vdots & \cdots & \cdots & \vdots \\ \frac{\partial \mathrm{f}_4}{\partial \mathrm{x}} & \cdots & \cdots & \frac{\partial \mathrm{f}_4}{\partial \mathrm{u}} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & \alpha \\ \frac{\mathrm{m}\bar{\mathrm{x}}^{(\mathrm{m}-1)}}{(1+\bar{\mathrm{x}}^{\mathrm{m}})^2} & -1 & 0 & 0 \\ 0 & \beta & -1 & 0 \\ 0 & 0 & -\frac{\mathrm{n}\bar{\mathrm{y}}^{(\mathrm{n}-1)}}{(1+\bar{\mathrm{y}}^{\mathrm{n}})^2} & -1 \end{bmatrix}.$$

The entries of the above matrix are evaluated at the unique equilibrium, which is the intersection of the nullclines shown in Figure 3.8 D. Note that the nullclines and the corresponding equilibrium depend on the choice of the paramters. Decreasing β, for example, causes the equilibrium to move towards higher values of x and lower values of y. If β is too small, the periodic orbit is lost (Figure 3.8 D). If we assume for simplicity that $\alpha = \beta$ and $\mathrm{m} = \mathrm{n}$, we can numerically assess the eigenvalues of the Jacobian as shown in Figure 3.8 E: eigenvalues with positive real part, and therefore a periodic orbit, are found in the blue region.

3.7.1.2 Oscillator coupled to a molecular load and stationary approximation

We want to transmit the oscillations to a downstream molecule L. Without loss of generality, we will assume we can couple rI2 to L (the same analysis can be easily carried out for SW21, SW12 and rA1.) Following the main text, we will consider two different cases:

- rI2 is consumed by the load. Chemical reactions: $\mathrm{rI2} + \mathrm{L} \xrightarrow{\mathrm{k_f}} \mathrm{L^a} \xrightarrow{\mathrm{k_r}} \mathrm{L}$.
- rI2 is not consumed by the load. Chemical reactions: $\mathrm{rI2} + \mathrm{L} \xrightarrow{\mathrm{k_f}} \mathrm{L^a} \xrightarrow{\mathrm{k_r}} \mathrm{rI2} + \mathrm{L}$.

The overall model is as follows:

$$\frac{\mathrm{d[rA1]}}{\mathrm{dt}} = \mathrm{k_p[SW12]} - \mathrm{k_d[rA1]}, \tag{3.9}$$

$$\tau \frac{\mathrm{d[SW21]}}{\mathrm{dt}} = [\mathrm{SW21^{tot}}] \frac{\frac{[\mathrm{rA1}]^{\mathrm{m}}}{\mathrm{KA^m}}}{1 + \frac{[\mathrm{rA1}]^{\mathrm{m}}}{\mathrm{KA^m}}} - \mathrm{[SW21]}, \tag{3.10}$$

$$\frac{\mathrm{d[rI2]}}{\mathrm{dt}} = \mathrm{k_p \cdot [SW21]} - \mathrm{k_d \cdot [rI2]} \underbrace{+\mathrm{k_r \cdot [L^a]}}_{\mathrm{non-consumptive}} \overbrace{-\mathrm{k_f \cdot [L][rI2]}}^{\mathrm{consumptive}}, \tag{3.11}$$

$$\tau \frac{\mathrm{d[SW12]}}{\mathrm{dt}} = [\mathrm{SW12^{tot}}] \frac{1}{1 + \frac{[\mathrm{rI2}]^{\mathrm{n}}}{\mathrm{KI^n}}} - \mathrm{[SW12]}, \tag{3.12}$$

$$\frac{\mathrm{d[L^a]}}{\mathrm{dt}} = -\mathrm{k_r \cdot [L^a]} + \mathrm{k_f \cdot [L][rI2]}, \tag{3.13}$$

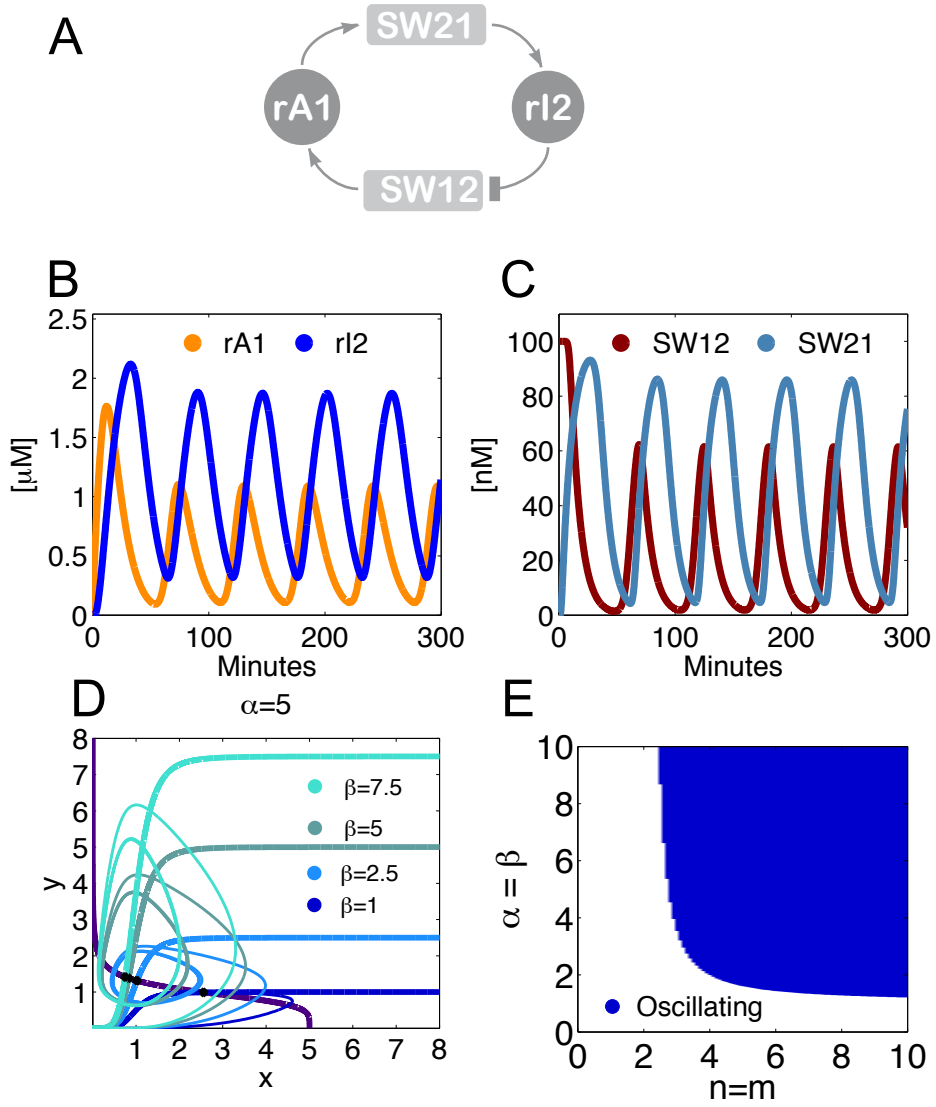


Figure 3.8: A. Scheme representing the simple oscillator model. B. Simulated rI2 and rA1 trajectories. C. Simulated SW12 and SW21 trajectories. D. Nullclines and corresponding trajectories (starting from zero initial conditions) for the non-dimensional model of the oscillator, plotted for varying β and $\alpha = 5$. E. Oscillatory domain calculated for the non-dimensional model when for simplicity we assume $\alpha = \beta$ and m=n. The nominal parameters used for these numerical simulations are chosen for illustrative purposes as: $k_p = 0.05\,/s$, $k_d = 0.002\,/s$, KA=KI=.5 μM, $[SW21^{tot}] = [SW12^{tot}] = 100\,nM$, m=n=5, $\tau = 500\,s$ (except for panel D, where β is varied).

and a schematic representation is given in Figure 3.9 A. For illustrative purposes, we choose $k_r = 0.006\,/s$ and $k_f = 7.9 \cdot 10^3\,/M/s$.

The dynamics of $[rI2^{tot}]$ are independent from the load in the consumptive coupling under pseudo-steady-state conditions

If we assume that $k_r \approx k_d$, we can show that the behavior of the total amount of rI2 is independent from the load, when we look at timescale shorter than the oscillation period:

$$\begin{aligned}\frac{d[rI2^{tot}]}{dt} &= \frac{d[rI2]}{dt} + \frac{d[L^a]}{dt} = k_p \cdot [SW21] - k_d \cdot [rI2] - k_f \cdot [L][rI2] - k_r \cdot [L^a] + k_f \cdot [L][rI2], \\ &= k_p \cdot [SW21] - k_d \cdot ([rI2] + [L^a]), \\ &= k_p \cdot [SW21] - k_d \cdot [rI2^{tot}].\end{aligned}$$

Since $[rI2^{tot}]$ is independent from [L], it is legitimate to solve separately the dynamics of $[L^a]$ in the short time scale:

$$\begin{aligned}\frac{d[L^a]}{dt} &= -k_r \cdot [L^a] + k_f \cdot ([L^{tot}] - [L^a])([rI2^{tot}](t) - [L^a]), \\ &= k_f \cdot [L^{tot}][rI2^{tot}](t) - [L^a]\{k_r + k_f \cdot ([L^{tot}] + [rI2^{tot}](t))\} + k_f \cdot [L^a]^2.\end{aligned}$$

The above differential equation is Lipschitz continuous and has no finite escape time, therefore its solution is unique at all times. If $[rI2^{tot}]$ is a positive forcing input to the system, the equation is an inhomogeneous ordinary differential equation which is driven by a periodic input.

It is possible to demonstrate that the solution to the above ordinary differential equation converges to a periodic orbit, whose period is determined by that of the input. An elegant way to prove this is to use the so called "contractivity" theory, and follow the theorems proposed in [105]. In short, it is sufficient to verify that the linearization of the differential equation is bounded by a negative constant and is therefore contractive. Since our system evolves on a compact and convex set, such property is global inside such set, and for any initial condition the system will converge to the periodic solution. If $d[L^a]/dt = f([L^a], [rI2^{tot}])$, we have:

$$\begin{aligned}\frac{\partial f([L^a], [rI2^{tot}])}{\partial [L^a]} &= -(k_r + k_f \cdot [rI2^{tot}](t) + k_f \cdot [L^{tot}]) + 2k_f \cdot [L^a], \\ &= -\left(k_r + k_f \cdot \underbrace{([rI2^{tot}](t) - [L^a])}_{\geq 0} + k_f \cdot \underbrace{([L^{tot}] - [L^a])}_{\geq 0} \right) \leq -c^2,\end{aligned}$$

with $c = \sqrt{k_r} > 0$. This verifies the condition of contractivity, and therefore we know that the load dynamics always converge to a periodic solution, having the same period as the input $[rI2^{tot}]$.

If we indicate $[\overline{L^a}]$ as the stationary solution, we can estimate the convergence speed by looking at the dynamics of the error $e = [L^a] - [\overline{L^a}]$:

$$\frac{de}{dt} = -k_r \cdot e - k_f \cdot ([rI2^{tot}] + [L^{tot}]) \cdot e + k_f \cdot e([\overline{L^a}] + [L^a]).$$

Take $V = e^2$ as a Lyapunov function for the system:

$$\begin{aligned}\frac{dV}{dt} = \frac{dV}{de}\frac{de}{dt} &= 2e \cdot \left(-k_r - k_f \cdot ([rI2^{tot}](t) + [L^{tot}] - [L^a] - [\overline{L^a}])\right) \cdot e, \\ &= 2 \cdot \left(-k_r - k_f \cdot \underbrace{([rI2^{tot}](t) - [L^a])}_{\geq 0} - k_f \cdot \underbrace{([L^{tot}] - [\overline{L^a}])}_{\geq 0} \right) \cdot e^2, \\ &= -2 \cdot Q \cdot e^2,\end{aligned}$$

where Q is a positive coefficient. Therefore, the dynamics of $[L^a]$ converge exponentially to their stationary solution, and the speed is driven by the coefficient $Q > k_r \approx k_d$.

To sum up, we can state that the equation:

$$\frac{d[L^a]}{dt} = -k_r \cdot [L^a] + k_f \cdot ([L^{tot}] - [L^a])([rI2^{tot}](t) - [L^a])$$

converges exponentially to the stationary solution with a timescale that is faster than $1/k_r$.

Quasi-steady-state approximation for the load dynamics

We have just shown that the dynamics of the load converge to the stationary solution with a speed $1/k_r$: based on our choice of $k_r = 0.006\,/s$, we know that the speed of convergence is on the order of 160 s. The nominal oscillator period for our simple model is around 1 hour, or 3600 s. Therefore, it is legitimate to approximate the load dynamics with the quasi-steady-state expression:

$$[\widehat{L^a}](t) \approx [L^{tot}] \left(1 - \frac{k_r}{k_r + k_f[rI2](t)}\right). \tag{3.14}$$

The validity of this approximation is illustrated in Figure 3.11 A and B, assuming that $k_r = 0.006\,/s$ and $k_f = 7.9 \cdot 10^3\,/M/s$.

Coupling efficiency

Assuming that [rI2](t) is a sinusoidal signal, we can use the static load approximation to evaluate the efficiency of the signal transmission. In particular, we can compute the amplitude of the load as a function of the oscillator amplitude. We will assume that $[\mathrm{rI2}](\mathrm{t}) = \mathrm{A}_0 + \mathrm{A}_1 \sin \omega \mathrm{t}$, where $\mathrm{A}_0, \mathrm{A}_1 > 0$ and $\mathrm{A}_0 > \mathrm{A}_1$. Define $\kappa = \mathrm{k_r}/\mathrm{k_f}$. The amplitude of the load oscillations is then given by:

$$\mathrm{A_L} = \frac{\mathrm{L^{tot}}}{2}\left(\frac{\kappa}{\kappa + (\mathrm{A}_0 - \mathrm{A}_1)} - \frac{\kappa}{\kappa + (\mathrm{A}_0 + \mathrm{A}_1)}\right).$$

By taking the derivative of $\mathrm{A_L}$ with respect to κ, and setting the derivative to zero, we can calculate the value of κ that maximizes $\mathrm{A_L}$:

$$\kappa_{\max} = \sqrt{\mathrm{A}_0^2 - \mathrm{A}_1^2}.$$

For instance, take $\mathrm{A}_0 \approx 1.1\,\mu\mathrm{M}$ and $\mathrm{A}_1 \approx 0.8\,\mu\mathrm{M}$ as in the nominal oscillations for [rI2]. Then, if we assume $\mathrm{k_r} = 0.006\,/\mathrm{s}$, the value of $\mathrm{k_f}$ that maximizes the load amplitude is $\mathrm{k_f} \simeq 7.9 \cdot 10^3\,/\mathrm{M/s}$. In the numerical simulations shown in the main paper, and reported here for the readers' convenience, we chose $\omega = 1 \cdot 10^{-3}\,\mathrm{rad/s}$, which is a good approximation of the nominal oscillation frequency of the system in the absence of load.

Perturbation of the oscillator caused by the load:

We can use the quasi-steady-state approximation of the $[\mathrm{L^a}](\mathrm{t})$ dynamics in the differential equation modeling [rI2]. This will give us a simpler expression to gain insight into the perturbation (or retroactivity) effect of the load on the oscillator dynamics. We will again consider the two separate cases of consumptive and non-consumptive coupling.

- **Consumptive coupling:** If we plug the load stationary solution into the consumptive dynamics of [rI2], we find:

$$\frac{\mathrm{d}[\widehat{\mathrm{rI2}}]}{\mathrm{dt}} = \mathrm{k_p} \cdot [\mathrm{SW21}] - \mathrm{k_d} \cdot [\widehat{\mathrm{rI2}}]\boxed{-\mathrm{k_f} \cdot [\widehat{\mathrm{rI2}}][\mathrm{L^{tot}}]\left(\frac{\mathrm{k_r}}{\mathrm{k_r} + \mathrm{k_f}[\widehat{\mathrm{rI2}}]}\right),} \tag{3.15}$$

$$[\widehat{\mathrm{L^a}}](\mathrm{t}) = [\mathrm{L^{tot}}]\left(1 - \frac{\mathrm{k_r}}{\mathrm{k_r} + \mathrm{k_f}[\widehat{\mathrm{rI2}}](\mathrm{t})}\right), \tag{3.16}$$

  where the box highlights the quasi-steady-state approximated perturbation term. Loosely speaking, the total amount of load linearly modulates an additional, bounded degradation term. (In fact, the perturbation term converges to $\mathrm{k_r} \cdot [\mathrm{L^{tot}}]$ for high values of [rI2].) The differential equations above were solved for varying amounts of $[\mathrm{L^{tot}}]$ numerically using MATLAB `ode23` routine; the results are shown in Figure 3.9 B and C and Figure 3.11 A and B. Initial conditions were chosen as: $[\mathrm{rA1}](0) = 0\,\mu\mathrm{M}$, $[\mathrm{SW21}](0) = 0\,\mathrm{nM}$, $[\mathrm{rI2}](0) = 0\,\mu\mathrm{M}$, $[\mathrm{SW12}](0) = 100\,\mathrm{nM}$, $[\mathrm{L^a}](0) = 0\,\mu\mathrm{M}$.

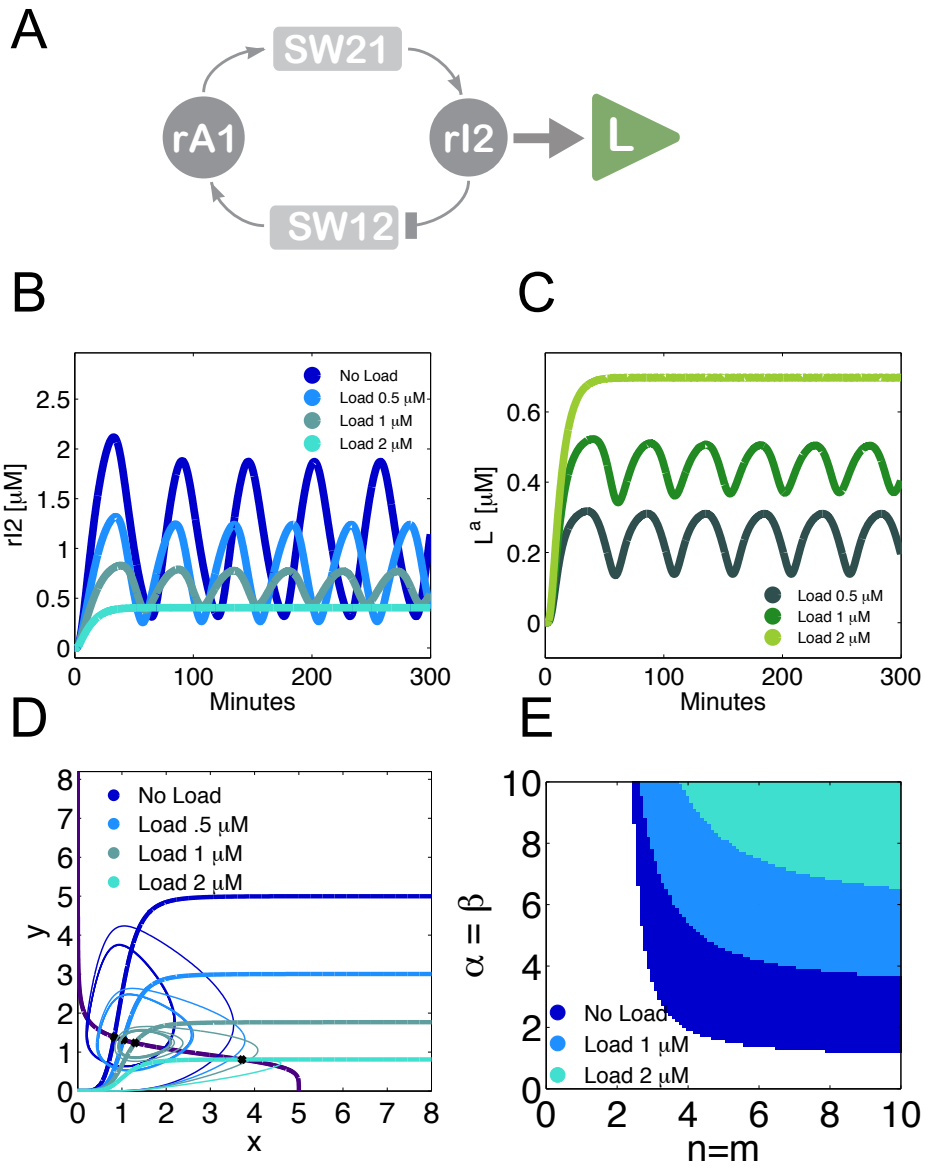


Figure 3.9: A. Schematic representation for the model problem oscillator coupled to a load. B. Trajectories of the rI2 species as a function of the total amount of load present in solution. C. Corresponding load trajectories. D. Nullclines and trajectories for the x and y variables of the non-dimensional model of the oscillator coupled to a load, plotted for variable amounts of total load. E. Oscillatory domain of the non-dimensional model, as a function of the total amount of load. The parameters are chosen as: $k_p = 0.05\,/s$, $k_d = 0.002\,/s$, KA=KI=.5 μM, $[SW21^{tot}] = [SW12^{tot}] = 100\,nM$, m=n=5, $\tau = 500\,s$, $k_r = 0.006\,/s$, $k_f = 7.9 \cdot 10^3\,/M/s$.

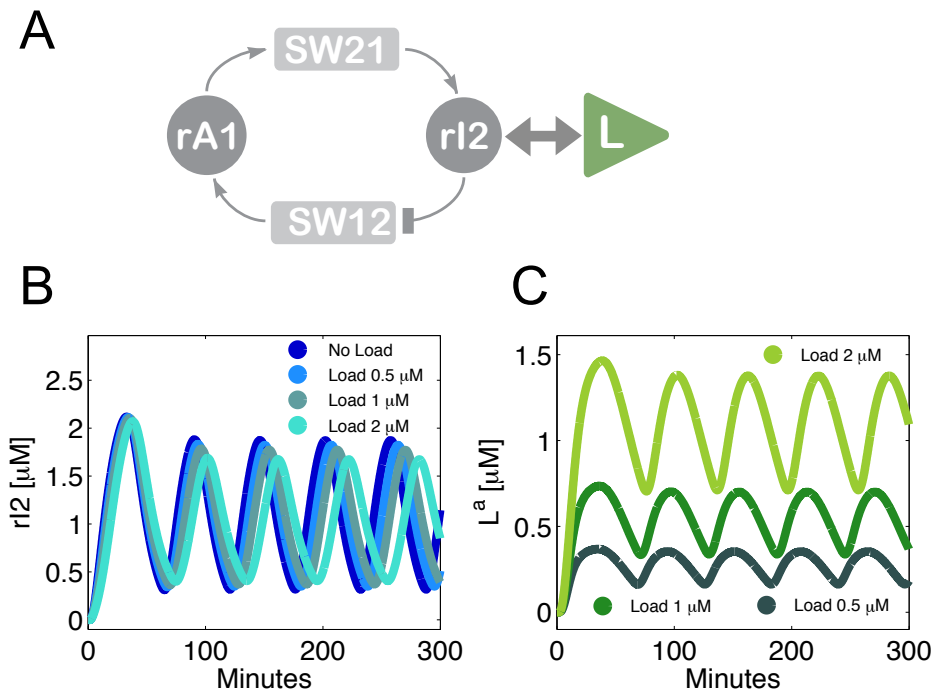


Figure 3.10: A. Scheme for the oscillator and load trajectories in the non-consumptive coupling mode. B. Trajectories of the rI2 species as a function of the total amount of load present in solution. C. Corresponding load trajectories. The parameters are chosen as: $k_p = 0.05\,/s$, $k_d = 0.002\,/s$, KA=KI=.5 μM, $[SW21^{tot}] = [SW12^{tot}] = 100\,nM$, m=n=5, $\tau = 500\,s$, $k_r = 0.006\,/s$, $k_f = 7.9 \cdot 10^3\,/M/s$.

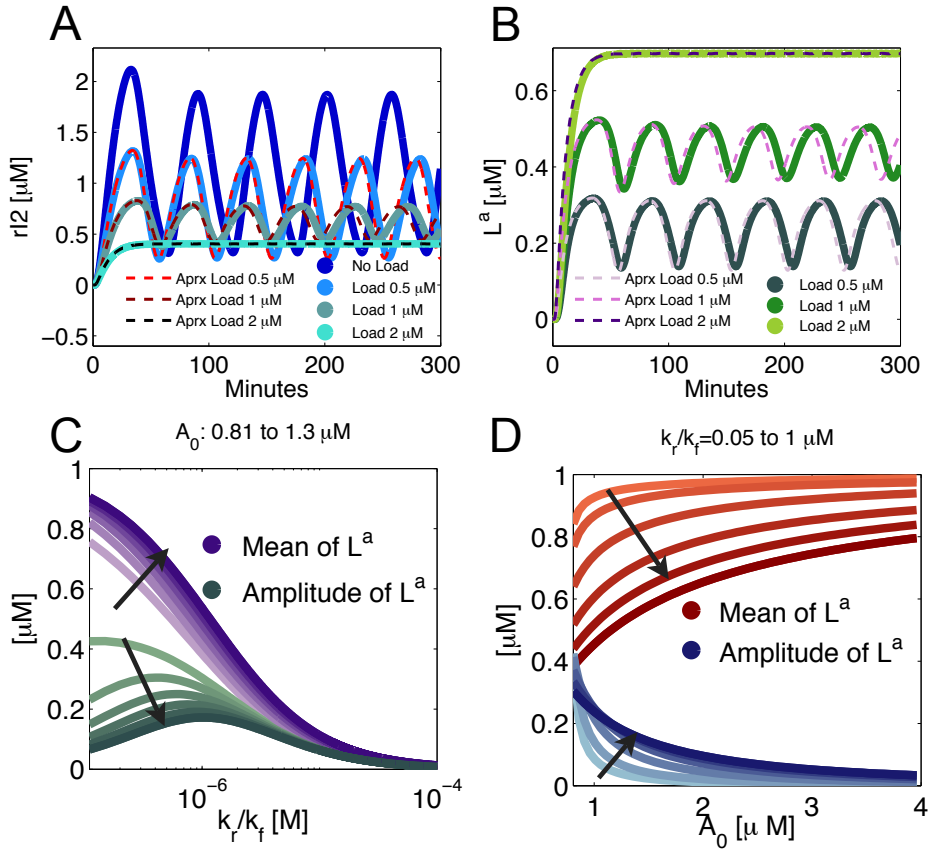


Figure 3.11: A. Numerical simulation comparing the full solution for the ordinary differential equations (3.9)–(3.13) (solid lines) to the solution when the load dynamics are approximated with the quasi-steady-state expression (3.14) (dashed lines). B. Load trajectories, comparison between the full solution (solid lines) and the quasi-steady-state approximation (dashed lines). C. Mean and amplitude of the active load $[L^a]$ as a function of the ratio of k_f and k_r, when the driving input is $rI2 = A_0 + A_1 \sin \omega t$, with A_0 varying between 0.81 (light color) and 1.3 μM, and $A_1 = 0.8\mu$M, $\omega = 0.001$ rad/s. D. Mean and amplitude of the active load signal $[L^a]$ as a function of the baseline A_0 for the input oscillating signal, for ratios of k_f and k_r varying between 0.05 and 1 μM. For A and B the parameters are chosen as: $k_p = 0.05$ /s, $k_d = 0.002$ /s, KA=KI=.5 μM, $[SW21^{tot}] = [SW12^{tot}] = 100$ nM, $[L^{tot}] = 1\,\mu$M, m=n=5, $\tau = 500$ s, $k_r = 0.006$ /s, $k_f = 7.9 \cdot 10^3$ /M/s.

- **Non-consumptive coupling:** When we plug the stationary approximation of $[L^a]$ into the non-consumptive version of equation (3.11), the resulting perturbation term is zero. This suggests that when $[L^a]$ converges faster than the oscillator to stationary dynamics, the stationary perturbation on the oscillator nominal trajectories is negligible in the non-consumptive case. However, this does not provide information on the perturbation magnitude produced on the transient dynamics of the oscillator. Figure 3.10 B and C shows the oscillator and load trajectories simulated in the non-consumptive coupling case. Comparing these plots with those of Figure 3.9 B and C, we can notice that the perturbation on rI2 is negligible, and therefore the oscillating signal is better propagated to the load.

The non-consumptive case has been considered in [27], where the authors derive an analytical expression for the retroactivity induced by the load. Such derivation is based on time-scale separation arguments requiring arbitrarily fast rates k_r and k_f. We highlight that we are not making this type of assumption in our analysis. Here, we will concisely summarize the results of [27] in the context of our system, referring the reader to the original paper for more technical details.

Following the reasoning in [27], suppose that k_r is much faster than k_d, and that the second-order binding rate k_f has a resulting speed comparable to the kinetics of k_r. It is then legitimate to assume that equation (3.13) reaches steady-state very fast and can be equated to zero. We can then reason that the total RNA amount $[rI2^{tot}] = [rI2] + [L^a]$ is the slow variable in the system, and we can rewrite equation (3.13) as a function of $[rI2^{tot}]$. By setting such equation to zero, we can find $[L^a]_s = g([rI2^{tot}])$, i.e., we can express the dynamics of $[L^a]$ on the slow manifold of the system. We can write:

$$\begin{aligned}\frac{d[rI2]_s}{dt} =& \frac{d[rI2^{tot}]_s}{dt} - \frac{d[L^a]_s}{dt},\\ =& \frac{d[rI2^{tot}]_s}{dt} - \frac{d\,g([rI2^{tot}]_s)}{d\,[rI2^{tot}]_s}\frac{d[rI2^{tot}]_s}{dt},\\ =& \frac{d[rI2^{tot}]_s}{dt}\left(1 - \frac{d\,g([rI2^{tot}]_s)}{d\,[rI2^{tot}]_s}\right).\end{aligned}$$

The term $d\,g([rI2^{tot}]_s)/d\,[rI2^{tot}]_s$ is called *retroactivity*, and it expresses the effect that the load has on the dynamics of the molecule it binds to, after a fast transient. Follwing [27], this term can be evaluated using the implicit function theorem. The final expression for the variable $[rI2]_s$ is:

$$\frac{d[rI2]_s}{dt} = \frac{d[rI2^{tot}]_s}{dt}\left(1 - \frac{1}{1 + \frac{\kappa}{[L^{tot}]}\left(1 + \frac{[rI2]_s}{\kappa}\right)^2}\right),$$

where $\kappa = k_r/k_f$. Based on our choice of parameter values we cannot carry out a rigorous

timescale separation. However, verifying the resulting retroactivity magnitude is still a useful exercise. Plugging into the above expression the numerical values: $\mathrm{k_r} = 6 \cdot 10^{-3}\,/\mathrm{s}$, $\mathrm{k_f} = 7.9 \cdot 10^{3}\,/\mathrm{M/s}$, we get $\kappa \approx 0.75\,\mu\mathrm{M}$. Let us assume that $[\mathrm{L^{tot}}] \approx 1\,\mu\mathrm{M}$. Also, hypothesize that $[\mathrm{rI2}]_\mathrm{s}$ is on the order of $1\,\mu\mathrm{M}$: then, $(1 + \frac{[\mathrm{rI2}]_\mathrm{s}}{\kappa})^2 \approx 5$. Finally, since $\frac{\kappa}{[\mathrm{L^{tot}}]} \approx 0.75$, we can conclude that in the presence of the load, the dynamics of [rI2], approximated on the slow manifold, are scaled by a factor 0.8 with respect to the load free trajectory (i.e., when $[\mathrm{rI2^{tot}}] = [\mathrm{rI2}]$). However, if we were to operate at κ either much larger or much smaller than $1\mu M$, the retroactivity would rapidly approach zero. This would be consistent with our approximate result saying that a non-consumptive coupling causes negligible perturbations on the source of chemical signal.

We remark that we do not invoke a formal timescale separation argument in our stationary approximation, and it is therefore not possible to rigorously compare our results to those in [27]. (However, we do justify the validity of our quasi-steady-state approximation of the load dynamics by comparing their convergence speed to the oscillator speed.)

Consumptive coupling: non-dimensional analysis of the oscillatory domain

The differential equations modeling the oscillator consumptively coupled to the load can be non-dimensionalized following the same procedure shown earlier.

Non-dimensional variables

$$\mathrm{x} = \frac{[\mathrm{rA1}]}{\mathrm{KA}} \qquad \mathrm{v} = \frac{[\mathrm{SW21}]}{[\mathrm{SW21^{tot}}]}$$

$$\mathrm{y} = \frac{[\mathrm{rI2}]}{\mathrm{KI}} \qquad \mathrm{u} = \frac{[\mathrm{SW12}]}{[\mathrm{SW12^{tot}}]}$$

$$\mathrm{w} = \frac{[\mathrm{L}]}{[\mathrm{L^{tot}}]}$$

Time rescaling

$$\tilde{\mathrm{t}} = \mathrm{t}/\tau$$

Non-dimensional parameters

$$\alpha = \frac{\mathrm{k_p}\,[\mathrm{SW12^{tot}}]}{\mathrm{k_d}\,\mathrm{KA}} \qquad \beta = \frac{\mathrm{k_p}\,[\mathrm{SW21^{tot}}]}{\mathrm{k_d}\,\mathrm{KI}}$$

$$\delta = \frac{\mathrm{k_r}}{\mathrm{k_d}} \qquad \theta = \frac{\mathrm{k_f}[\mathrm{L^{tot}}]}{\mathrm{k_d}} \qquad \phi = \frac{\mathrm{k_f KI}}{\mathrm{k_d}}$$

Non-dimensional equations in $\tilde{t}$

$$\gamma\,\dot{\mathrm{x}} = \alpha\mathrm{u} - \mathrm{x} \tag{3.17}$$

$$\dot{\mathrm{v}} = \frac{\mathrm{x^m}}{1 + \mathrm{x^m}} - \mathrm{v} \tag{3.18}$$

$$\gamma\,\dot{\mathrm{y}} = \beta\mathrm{v} - \mathrm{y}(1 + \theta\mathrm{w}) \tag{3.19}$$

$$\dot{\mathrm{u}} = \frac{1}{1 + \mathrm{y^n}} - \mathrm{u} \tag{3.20}$$

$$\gamma\,\dot{\mathrm{w}} = \delta(1 - \mathrm{w}) - \phi\,\mathrm{wy} \tag{3.21}$$

Note that $\gamma = 1$. Equilibria will be indicated as $\bar{x}$, $\bar{u}$, $\bar{y}$, $\bar{u}$ and $\bar{w}$:

$$(3.17) = 0 \Longrightarrow \bar{\mathrm{x}} = \alpha\bar{\mathrm{u}}$$

$$(3.18) = 0 \Longrightarrow \bar{\mathrm{v}} = \frac{\bar{\mathrm{x}}^{\mathrm{m}}}{1+\bar{\mathrm{x}}^{\mathrm{m}}}$$

$$(3.19)\,\&\,(3.21) = 0 \Longrightarrow \bar{\mathrm{y}} \quad \text{is the positive solution of} \quad \phi \mathrm{y}^2 + \mathrm{y}(\delta(\theta+1) - \phi\beta\bar{\mathrm{v}}) - \delta\beta\bar{\mathrm{v}} = 0$$

$$(3.20) = 0 \Longrightarrow \bar{\mathrm{u}} = \frac{1}{1+\bar{\mathrm{y}}^{\mathrm{n}}}$$

$$(3.21) = 0 \Longrightarrow \bar{\mathrm{w}} = \frac{\delta}{\delta + \phi\bar{\mathrm{y}}}.$$

The system Jacobian is:

$$\mathrm{J_y} = \begin{bmatrix} -1 & 0 & 0 & \alpha & 0 \\ \frac{\mathrm{m}\bar{\mathrm{x}}^{(\mathrm{m}-1)}}{(1+\bar{\mathrm{x}}^{\mathrm{m}})^2} & -1 & 0 & 0 & 0 \\ 0 & \beta & -1-\theta\bar{\mathrm{w}} & 0 & -\theta\bar{\mathrm{y}} \\ 0 & 0 & -\frac{\mathrm{n}\bar{\mathrm{y}}^{(\mathrm{n}-1)}}{(1+\bar{\mathrm{y}}^{\mathrm{n}})^2} & -1 & 0 \\ 0 & 0 & -\phi\bar{\mathrm{w}} & 0 & -\delta-\phi\bar{\mathrm{y}} \end{bmatrix}.$$

The nullclines of the system are represented in Figure 3.9 D. Increasing the load induces similar changes in the nullclines as decreasing β in the absence of load (Figure 3.8 D), which is equivalent to increasing the degradation rate for [rI2].

Figure 3.9 E shows how the oscillatory domain of the system shrinks when the total amount of load is increased. The figure is obtained by checking the eigenvalues of the Jacobian matrix $\mathrm{J_y}$.

3.7.1.3 Insulation

Consider the case where the load is coupled consumptively to the oscillator. How can the perturbation on the oscillator be reduced? When it is not practical to modify the binding rates that introduce the coupling, the only way to reduce perturbation is to use a minimal amount of load. We can overcome this restriction by coupling the oscillatory signal to a small amount of another molecular device, whose output is capable of amplifying the oscillator signal and driving large amounts of load. We will call this device an insulator, following the analysis proposed in [27]. A schematic representation of this idea is shown in Figure 3.12 A.

An insulating device can be implemented easily as a small amount of a third switch, Ins, which is directly coupled to the oscillator. The RNA output from the insulating switch, InsOut, is used to drive the load.

The set of chemical reactions representing the insulator and load are:

$$\mathrm{rI2} + \mathrm{Ins} \xrightarrow{k_f} \mathrm{Ins^a}, \qquad \mathrm{Ins^a} \xrightarrow{k_r} \mathrm{Ins},$$

$$\mathrm{Ins^a} \xrightarrow{k_p^i} \mathrm{Ins^a} + \mathrm{InsOut},$$

$$\mathrm{InsOut} \xrightarrow{k_d^i} \emptyset,$$

$$\mathrm{InsOut} + \mathrm{L} \xrightarrow{k_f^i} \mathrm{L^a}, \qquad \mathrm{L^a} \xrightarrow{k_r^i} \mathrm{L},$$

$$\mathrm{Ins^{tot}} = \mathrm{Ins^a} + \mathrm{Ins}, \qquad \mathrm{L^{tot}} = \mathrm{L} + \mathrm{L^a}.$$

The differential equations corresponding to the oscillator and the insulated load are:

$$\begin{aligned}
\frac{\mathrm{d[rA1]}}{\mathrm{dt}} &= \mathrm{k_p} \cdot \mathrm{[SW12]} - \mathrm{k_d} \cdot \mathrm{[rA1]}, \\
\tau \frac{\mathrm{[SW21]}}{\mathrm{dt}} &= \mathrm{[SW21^{tot}]} \left(\frac{\left(\frac{\mathrm{[rA1]}}{\mathrm{KA}}\right)^{\mathrm{m}}}{1 + \left(\frac{\mathrm{[rA1]}}{\mathrm{KA}}\right)^{\mathrm{m}}} \right) - \mathrm{[SW21]}, \\
\frac{\mathrm{d[rI2]}}{\mathrm{dt}} &= \mathrm{k_p} \cdot \mathrm{[SW21]} - \mathrm{k_d} \cdot \mathrm{[rI2]} - \mathrm{k_f} \cdot \mathrm{[Ins][rI2]}, \\
\tau \frac{\mathrm{[SW12]}}{\mathrm{dt}} &= \mathrm{[SW12^{tot}]} \left(\frac{1}{1 + \left(\frac{\mathrm{[rI2]}}{\mathrm{KI}}\right)^{\mathrm{n}}} \right) - \mathrm{[SW12]}, \\
\frac{\mathrm{d[Ins]}}{\mathrm{dt}} &= \mathrm{k_r} \cdot \mathrm{[Ins^a]} - \mathrm{k_f} \cdot \mathrm{[Ins][rI2]}, \\
\frac{\mathrm{d[InsOut]}}{\mathrm{dt}} &= \mathrm{k_p^i} \cdot \mathrm{[Ins^a]} - \mathrm{k_d^i} \cdot \mathrm{[InsOut]} - \mathrm{k_f^i} \cdot \mathrm{[InsOut][L]}, \\
\frac{\mathrm{d[L]}}{\mathrm{dt}} &= \mathrm{k_r^i} \cdot (\mathrm{[L^{tot}]} - \mathrm{[L]}) - \mathrm{k_f^i} \cdot \mathrm{[InsOut][L]}.
\end{aligned}$$

The parameters chosen for the numerical analysis of the system are: $\mathrm{k_p} = 0.05\,/\mathrm{s}$, $\mathrm{k_d} = 0.002\,/\mathrm{s}$, KA=KI=.5 μM, $\mathrm{[SW21^{tot}]} = \mathrm{[SW12^{tot}]} = 100\,\mathrm{nM}$, m=n=5, $\tau = 500\,\mathrm{s}$, $\mathrm{k_r} = 0.006\,/\mathrm{s}$, $\mathrm{k_f} = 7.9 \cdot 10^3\,/\mathrm{M/s}$, $\mathrm{k_p^i} = 0.15\,/\mathrm{s}$, $\mathrm{k_d^i} = 0.006\,/\mathrm{s}$, $\mathrm{k_r^i} = 0.006\,/\mathrm{s}$, and $\mathrm{k_f^i} = 6 \cdot 10^3\,/\mathrm{M/s}$. (Note that the oscillator parameters have not been changed from those used in Section 3.7.1.1.) All the parameters have been chosen for illustrative purposes. The above differential equations have been solved numerically using the MATLAB `ode23` solver, and are shown in Figure 3.12 B and C. Initial conditions were chosen as: $\mathrm{[rA1]}(0) = 0\,\mu\mathrm{M}$, $\mathrm{[SW21]}(0) = 0\,\mathrm{nM}$, $\mathrm{[rI2]}(0) = 0\,\mu\mathrm{M}$, $\mathrm{[SW12]}(0) = 100\,\mathrm{nM}$, $\mathrm{[Ins^a]}(0) = 0\,\mathrm{nM}$, $\mathrm{[InsOut]}(0) = 0\,\mu\mathrm{M}$, $\mathrm{[L^a]}(0) = 0\,\mu\mathrm{M}$.

Non-dimensional model

This model can be rendered non-dimensional with the same procedure adopted before:

Non-dimensional variables

$$\mathrm{x} = \frac{[\mathrm{rA1}]}{\mathrm{KA}} \quad \mathrm{v} = \frac{[\mathrm{SW21}]}{[\mathrm{SW21^{tot}}]}$$
$$y = \frac{[\mathrm{rI2}]}{\mathrm{KI}}, \quad u = \frac{[\mathrm{SW12}]}{[\mathrm{SW12^{tot}}]}$$
$$\mathrm{h} = \frac{[\mathrm{Ins}]}{[\mathrm{Ins^{tot}}]} \quad \mathrm{z} = \frac{[\mathrm{InsOut}]}{[\mathrm{L^{tot}}]}$$
$$w = \frac{[\mathrm{L}]}{[\mathrm{L^{tot}}]}$$

Time rescaling

$$\tilde{\mathrm{t}} = \mathrm{t}/\tau$$

Non-dimensional parameters

$$\alpha = \frac{\mathrm{k_p}\,[\mathrm{SW12^{tot}}]}{\mathrm{k_d}\,\mathrm{KA}} \quad \beta = \frac{\mathrm{k_p}\,[\mathrm{SW21^{tot}}]}{\mathrm{k_d}\,\mathrm{KI}}$$
$$\delta = \frac{\mathrm{k_r}}{\mathrm{k_d}} \qquad \theta = \frac{\mathrm{k_f}[\mathrm{Ins^{tot}}]}{\mathrm{k_d}} \qquad \phi = \frac{\mathrm{k_f KI}}{\mathrm{k_d}}$$
$$\lambda = \frac{\mathrm{k_p^i}}{\mathrm{k_d}}\frac{[\mathrm{Ins^{tot}}]}{[\mathrm{L^{tot}}]} \quad \rho = \frac{\mathrm{k_f^i}}{\mathrm{k_d}}[\mathrm{L^{tot}}] \quad \xi = \frac{\mathrm{k_r^i}}{\mathrm{k_d}} \quad \psi = \frac{\mathrm{k_d^i}}{\mathrm{k_d}} \quad \gamma = \frac{1}{\mathrm{k_d}\,\tau}$$

Non-dimensional equations in $\tilde{t}$

$$\gamma\,\dot{\mathrm{x}} = \alpha\mathrm{u} - \mathrm{x} \tag{3.22}$$
$$\dot{\mathrm{v}} = \frac{\mathrm{x^m}}{1+\mathrm{x^m}} - \mathrm{v} \tag{3.23}$$
$$\gamma\,\dot{\mathrm{y}} = \beta\mathrm{v} - \mathrm{y}(1{+}\theta\mathrm{h}) \tag{3.24}$$
$$\dot{\mathrm{u}} = \frac{1}{1+\mathrm{y^n}} - \mathrm{u} \tag{3.25}$$
$$\gamma\,\dot{\mathrm{h}} = \delta(1-\mathrm{h}) - \phi\,\mathrm{hy} \tag{3.26}$$
$$\gamma\,\dot{\mathrm{z}} = \lambda(1-\mathrm{h}) - \mathrm{z}(\psi + \rho\,\mathrm{w}) \tag{3.27}$$
$$\gamma\,\dot{\mathrm{w}} = \xi(1-\mathrm{w}) - \rho\,\mathrm{zw}, \tag{3.28}$$

where $\gamma = 1$.

Equilibria will be indicated as $\bar{x}$, $\bar{u}$, $\bar{y}$, $\bar{u}$, $\bar{h}$, $\bar{z}$, and $\bar{w}$:

$(3.22) = 0 \Longrightarrow \bar{\mathrm{x}} = \alpha\bar{\mathrm{u}}$

$(3.23) = 0 \Longrightarrow \bar{\mathrm{v}} = \dfrac{\bar{\mathrm{x}}^{\mathrm{m}}}{1+\bar{\mathrm{x}}^{\mathrm{m}}}$

$(3.24) = 0\,\&\,(3.26) = 0 \Longrightarrow \bar{\mathrm{y}}$ is the positive solution of $\phi\mathrm{y}^2 + \mathrm{y}(\delta(\theta+1) - \phi\beta\bar{\mathrm{v}}) - \delta\beta\bar{\mathrm{v}} = 0$

$(3.25) = 0 \Longrightarrow \bar{\mathrm{u}} = \dfrac{1}{1+\bar{\mathrm{y}}^{\mathrm{n}}}$

$(3.26) = 0 \Longrightarrow \bar{\mathrm{h}} = \dfrac{\delta}{\delta + \phi\bar{\mathrm{y}}}$

$(3.27) = 0\&\,(3.28) = 0 \Longrightarrow \bar{\mathrm{z}}$ is the positive solution of $\rho\psi\mathrm{z}^2 + \mathrm{z}(\xi(\rho+\psi) - \rho\lambda(1-\bar{\mathrm{h}})) = \xi\lambda(1-\bar{\mathrm{h}})$

$(3.28) = 0 \Longrightarrow \bar{\mathrm{w}} = \dfrac{\xi}{\xi + \rho\bar{\mathrm{z}}}.$

The Jacobian of this set of equations is:

$$J_{yIns} = \begin{bmatrix} -1 & 0 & 0 & \alpha & 0 & 0 & 0 \\ \frac{m\bar{x}^{(m-1)}}{(1+\bar{x}^m)^2} & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta & -1-\theta\bar{h} & 0 & -\theta\bar{y} & 0 & 0 \\ 0 & 0 & -\frac{n\bar{y}^{(n-1)}}{(1+\bar{y}^n)^2} & -1 & 0 & 0 & 0 \\ 0 & 0 & -\phi\bar{h} & 0 & -\delta-\phi\bar{y} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\lambda & -\psi-\rho\bar{w} & -\rho\bar{z} \\ 0 & 0 & 0 & 0 & 0 & -\rho\bar{w} & -\xi-\rho\bar{z} \end{bmatrix}.$$

The nullclines of the system are represented in Figure 3.12 D. Because the amount of insulator is small, the perturbation introduced in the oscillator dynamics is negligible: the oscillatory domain of the system is almost coincident with that of the oscillator in the absence of load. Figure 3.12 E shows the oscillatory domain corresponding to different amounts of total load. The figure is obtained by numerically checking the eigenvalues of the Jacobian matrix J_{yIns}.

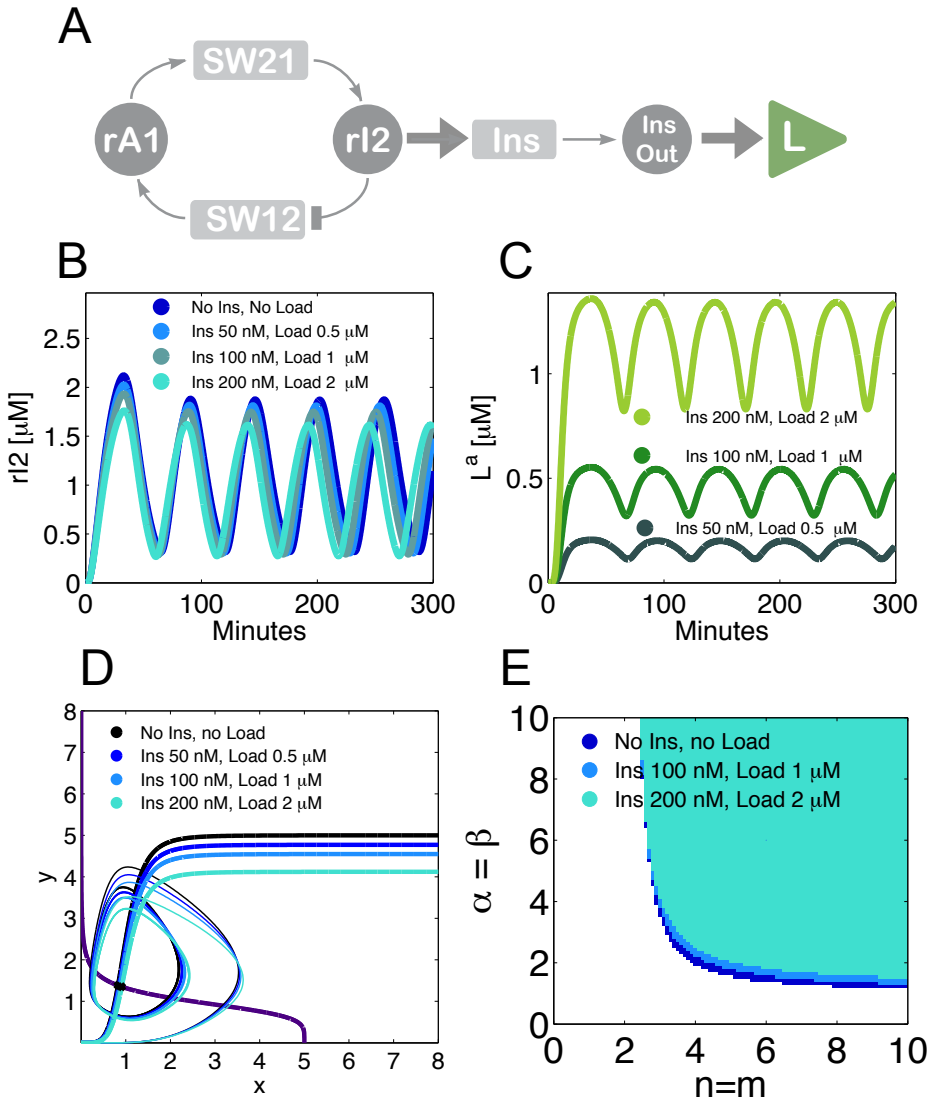


Figure 3.12: A. Scheme for the oscillator coupled to a load through an insulating switch. B. Trajectories of the oscillator rI2 species as a function of the total amount of insulating genelet and load. C. Corresponding load trajectories. D. Nullclines and trajectories for the non-dimensional oscillator, x and y variables (corresponding to rA1 and rI2). E. Oscillatory domain of the oscillator as a function of the total amount of insulator and load. The parameters are chosen as: $k_p = 0.05\,/s$, $k_d = 0.002\,/s$, KA=KI=.5 μM, $[SW21^{tot}] = [SW12^{tot}] = 100\,nM$, m=n=5, $\tau = 500\,s$, $k_r = 0.006\,/s$, $k_f = 7.9 \cdot 10^3\,/M/s$, $k_p^i = 0.15\,/s$, $k_d^i = 0.006\,/s$, $k_r^i = 0.006\,/s$ and $k_f^i = 6 \cdot 10^3\,/M/s$.

3.7.2 Relevant sequence interactions

This section contains a series of schematic figures, which represent the most relevant predicted interactions among the nucleic acids composing the oscillator and tweezers system. The color coding for the different domains follows the one chosen for Figures 3.2, 3.4, and 3.5. These schemes have an illustrative purpose and are not an exhaustive list of all secondary structures that can occur in the system. Toehold-mediated branch migration reactions will be listed and analyzed when the exposed toeholds are longer than 4 bases. We will neglect reactions involving toeholds 4 bases or shorter, under the assumption that the corresponding time scales exceed the oscillator dynamics.

As evidenced in [64] through gel electrophoresis, a number of short RNA species accumulates over time during an oscillator experiment. Such short species have lengths between 5–30 bases and are the product of abortive RNAP transcription and incomplete degradation by RNase H. It is known that RNase H may fail to degrade up to 7 RNA bases on the 3' end of DNA in an RNA-DNA duplex. In this section we will only highlight potential interactions of the oscillator and tweezer strands with incomplete degradation products of maximum length. The detailed reactions involving incomplete degradation products can be found in [41], Supplementary Appendix.

SW12

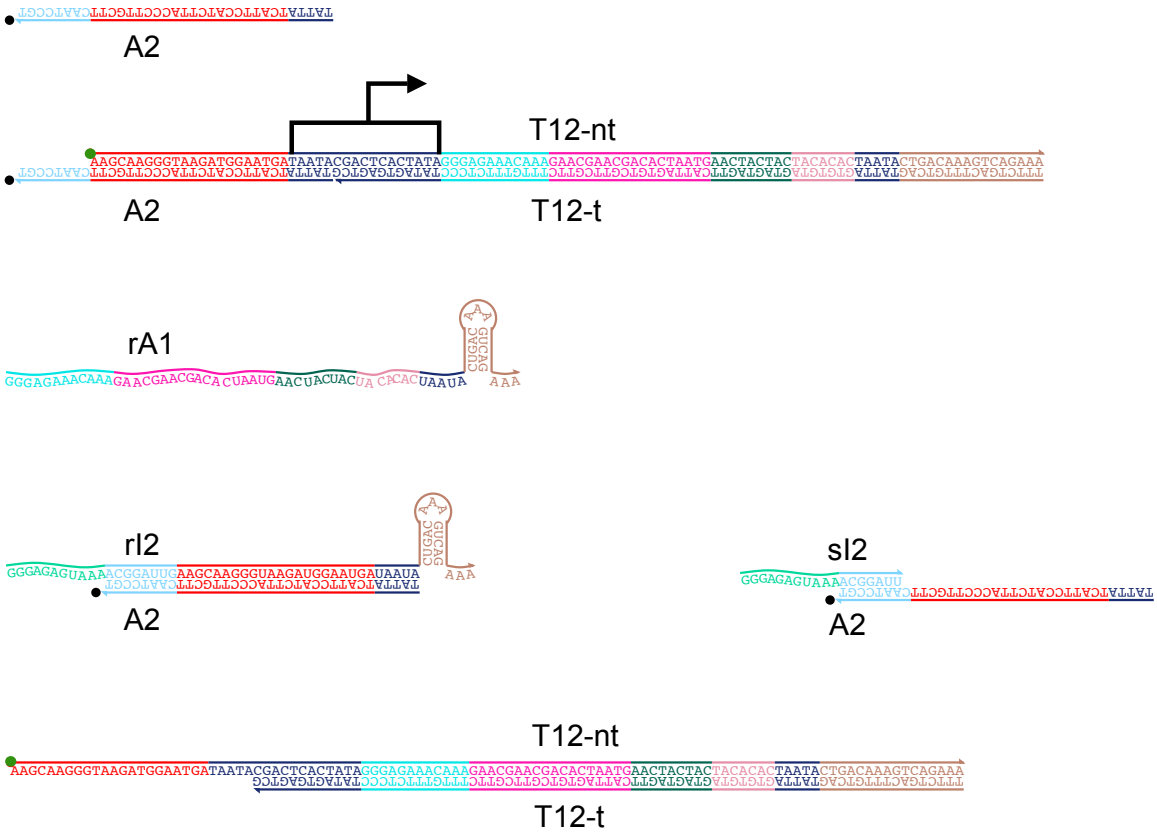


Figure 3.13: Scheme of relevant interactions for SW12. From top to bottom: activator A2; on state SW12; output rA1 of SW12; activator A2 sequestered by the RNA input rI2; incomplete RNase H degradation product sI2, binding to A2; off-state SW12.

	Reaction Pathways
Activation:	$A2 + T12 \rightharpoonup T12 \cdot A2$
Inhibition:	$rI2 + T12 \cdot A2 \rightharpoonup rI2 \cdot A2 + T12$
Annihilation:	$A2 + rI2 \rightharpoonup rI2 \cdot A2$
Transcription:	$T12 \cdot A2 + RNAP \rightleftharpoons T12 \cdot A2 \cdot RNAP \rightharpoonup rA1 + T12 \cdot A2 + RNAP$
	$T12 + RNAP \rightleftharpoons T12 \cdot RNAP \rightharpoonup rA1 + T12 + RNAP$
Degradation:	$rI2 \cdot A2 + RNaseH \rightleftharpoons rI2 \cdot A2 \cdot RNaseH \rightharpoonup sI2 \cdot A2 + RNaseH$

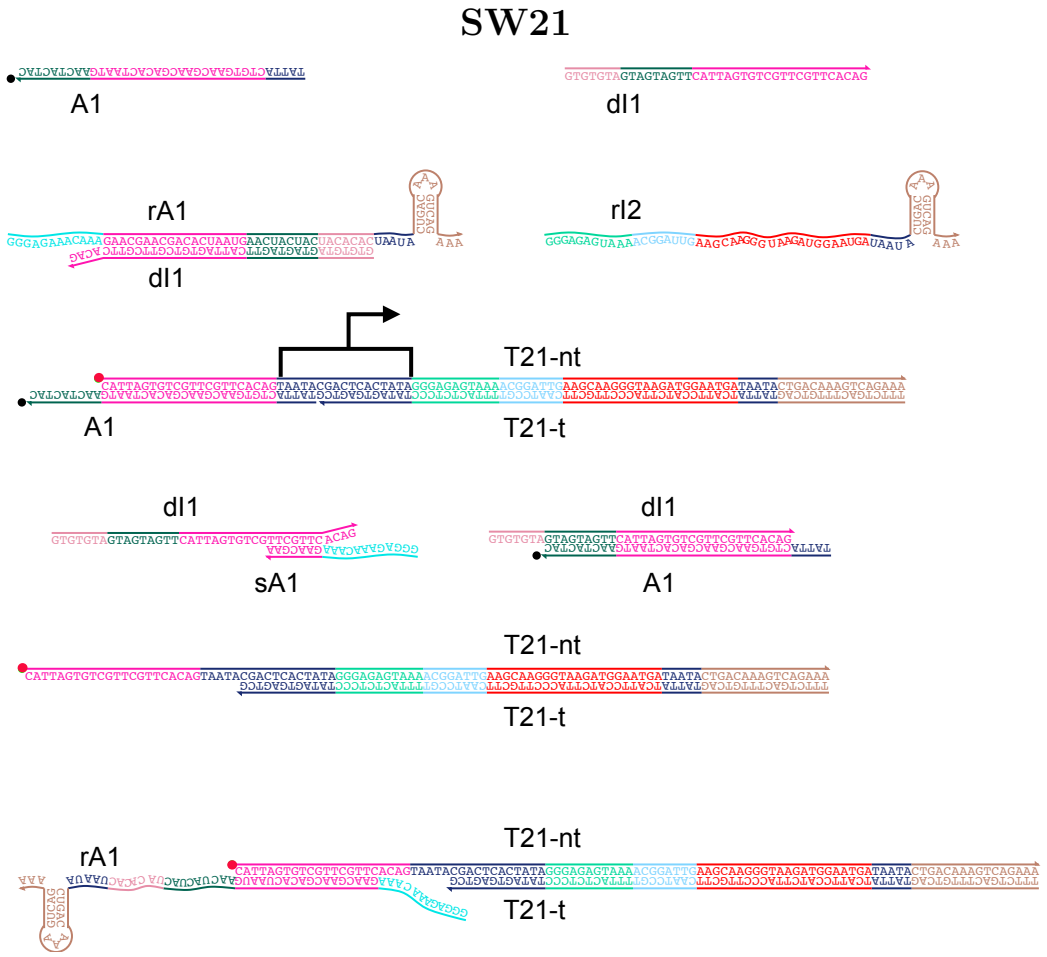


Figure 3.14: Relevant interactions for SW21. From top to bottom: activator A1 and inhibitor dI1; dI1 sequestered by RNA input rA1 and output rI2 of SW21; on-state SW21; incomplete RNase H degradation product sA1 binding to dI1 and dI1 sequestering the activator A1; off-state SW21; finally, unwanted interaction between RNA input rA1 and off-state SW21. The latter complex is a substrate for RNase H; moreover rA1 has a 16-base toehold for initiation of strand displacement by inhibitor dI1, and a 9-base domain exposed for A1 invasion. The T21·rA1 substrate does not represent a suitable binding site for RNA polymerase (Section 3.7.14).

Reaction Pathways	
Activation:	$A1 + T21 \rightharpoonup T21 \cdot A1$
Inhibition:	$dI1 + T21 \cdot A1 \rightharpoonup dI1 \cdot A1 + T21$
Release:	$rA1 + dI1 \cdot A1 \rightharpoonup rA1 \cdot dI1 + A1$
Annihilation:	$dI1 + A1 \rightharpoonup dI1 \cdot A1$
	$dI1 + rA1 \rightharpoonup dI1 \cdot rA1$
Transcription:	$T21 \cdot A1 + RNAP \rightleftharpoons T21 \cdot A1 \cdot RNAP \rightharpoonup rI2 + T21 \cdot A1 + RNAP$
	$T21 + RNAP \rightleftharpoons T21 \cdot RNAP \rightharpoonup rI2 + T21 + RNAP$
Degradation:	$rA1 \cdot dI1 + RNaseH \rightleftharpoons rA1 \cdot dI1 \cdot RNaseH \rightharpoonup dI1 + RNaseH$

Unmodeled Reactions	
Interfering:	$rA1 + T21 \rightharpoonup T21 \cdot rA1$
Recapturing:	$dI1 + T21 \cdot rA1 \rightharpoonup rA1 \cdot dI1 + T21$
Recovering:	$A1 + T21 \cdot rA1 \rightharpoonup A1 \cdot T21 + rA1$
Degradation:	$T21 \cdot rA1 + RNaseH \rightleftharpoons T21 \cdot rA1 \cdot RNaseH \rightharpoonup T21 + RNaseH$

Tweezers mode I

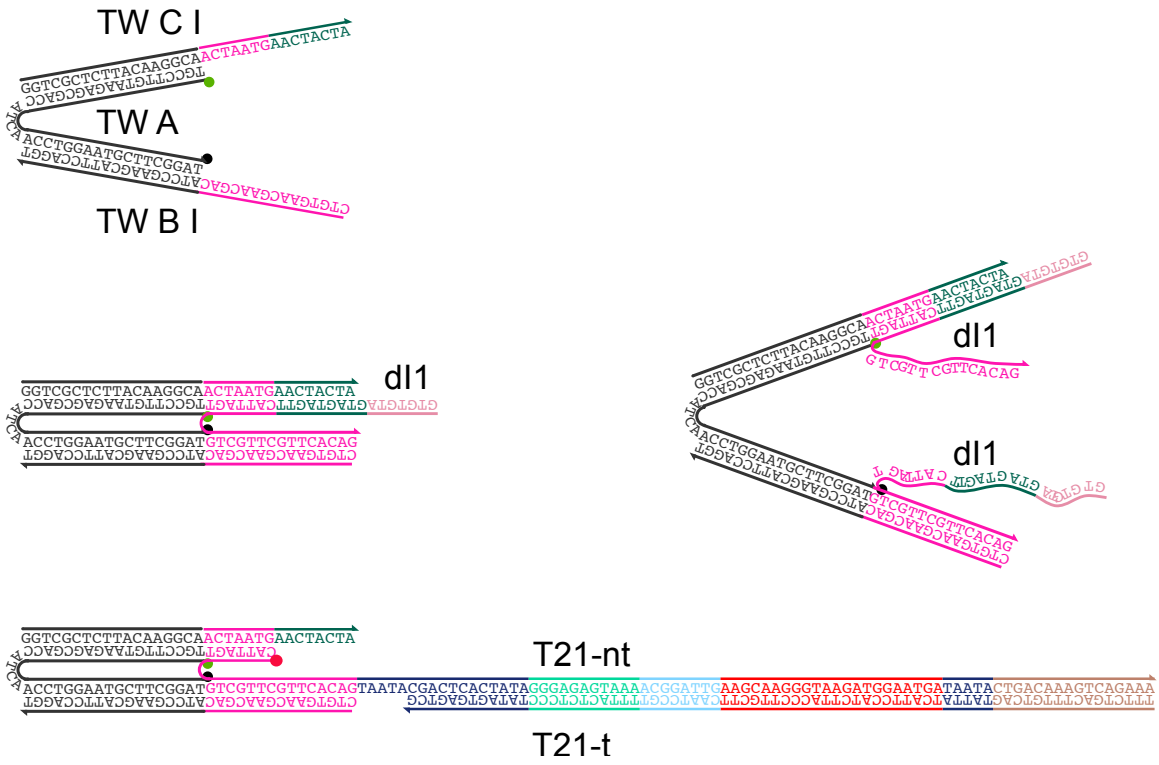


Figure 3.15: Relevant interactions for tweezers mode I. From top to bottom: tweezers mode I in open state; tweezers in closed state bound to their target dI1. Finally, unwanted interaction between open tweezers mode I and off-state SW21. Activator A1 can invade this undesired complex, binding to the exposed 5-base `TAATA` promoter domain, thereby displacing the tweezers.

Reaction Pathways

Closing:	$\mathrm{TwI} + \mathrm{dI1} \rightharpoonup \mathrm{TwI} \cdot \mathrm{dI1}$
Opening:	$\mathrm{TwI} \cdot \mathrm{dI1} + \mathrm{rA1} \rightharpoonup \mathrm{dI1} \cdot \mathrm{rA1} + \mathrm{TwI}$

Unmodeled reactions

Interfering:	$\mathrm{TwI} + \mathrm{T21} \rightharpoonup \mathrm{TwI} \cdot \mathrm{T21}$
Recovering:	$\mathrm{TwI} \cdot \mathrm{T21} + \mathrm{dI1} \rightharpoonup \mathrm{T21} + \mathrm{TwI} \cdot \mathrm{dI1}$
Recapturing:	$\mathrm{TwI} \cdot \mathrm{T21} + \mathrm{A1} \rightharpoonup \mathrm{TwI} + \mathrm{T21} \cdot \mathrm{A1}$
Double binding:	$\mathrm{TwI} + \mathrm{dI1} + \mathrm{dI1} \rightharpoonup \mathrm{TwI} \cdot \mathrm{dI1} \cdot \mathrm{dI1}$
Clearing:	$\mathrm{TwI} \cdot \mathrm{dI1} \cdot \mathrm{dI1} + \mathrm{rA1} + \mathrm{rA1} \rightharpoonup \mathrm{dI1} \cdot \mathrm{rA1} + \mathrm{dI1} \cdot \mathrm{rA1} + \mathrm{TwI}$

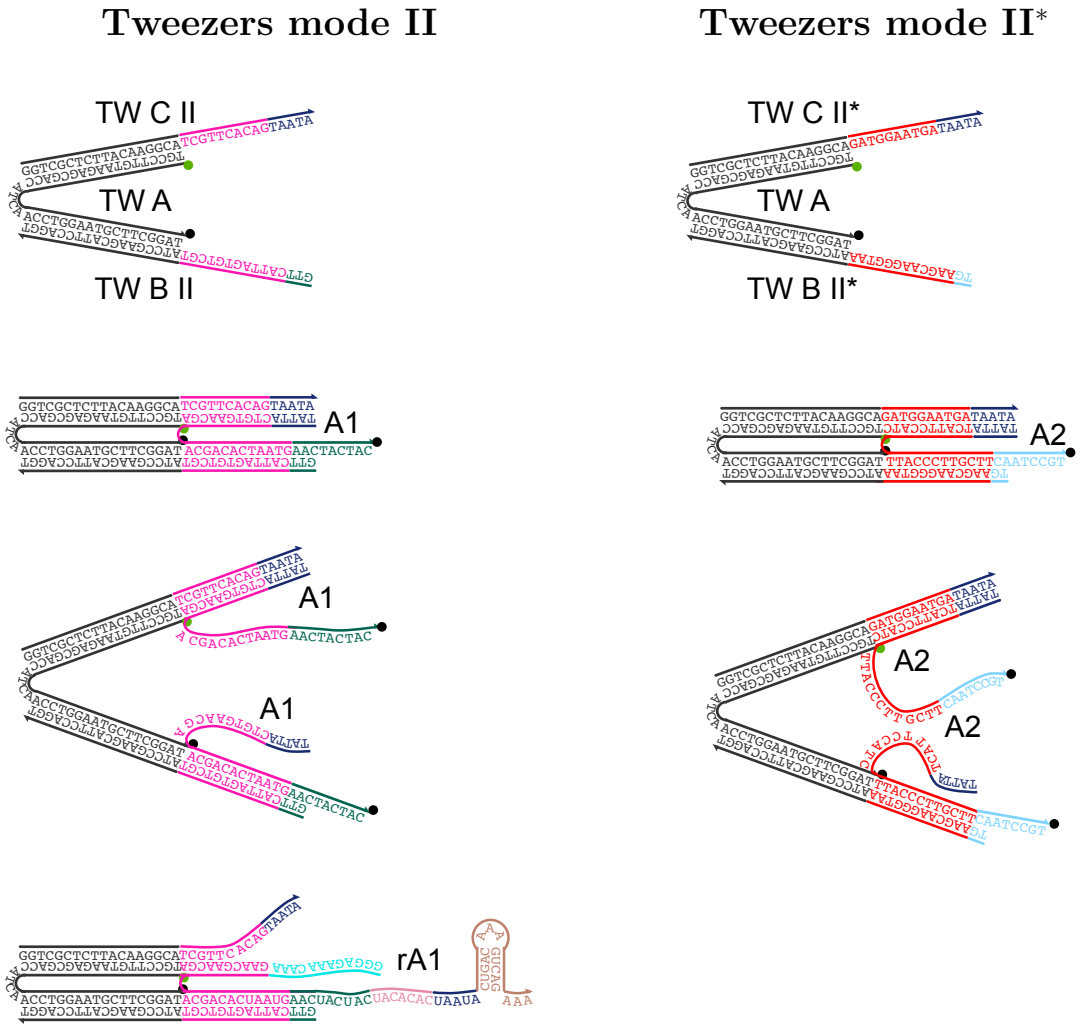


Figure 3.16: Relevant interactions for tweezers mode II. Left: mode II, with A1 input. Right: mode II*, with A2 input. Left, top to bottom: open tweezers; target A1 closing the tweezers; unwanted interaction with rA1. The latter complex is a substrate for RNase H, and there is a 9-base toehold for displacement of the tweezers by A1. Right, top to bottom: open tweezers; target A2 closing the tweezers.

Reaction Pathway Examples		
	Tweezers mode II	Tweezers mode II*
Closing:	$\mathrm{TwII} + \mathrm{A1} \rightharpoonup \mathrm{TwII} \cdot \mathrm{A1}$	$\mathrm{TwII^*} + \mathrm{A2} \rightharpoonup \mathrm{TwII^*} \cdot \mathrm{A2}$
Opening:	$\mathrm{TwII} \cdot \mathrm{A1} + \mathrm{dI1} \rightharpoonup \mathrm{TwII} + \mathrm{A1} \cdot \mathrm{dI1}$	$\mathrm{TwII^*} \cdot \mathrm{A2} + \mathrm{rI2} \rightharpoonup \mathrm{TwII^*} + \mathrm{A2} \cdot \mathrm{rI2}$

Unmodeled reactions	
	Tweezers mode II
Interfering:	$\mathrm{TwII} + \mathrm{rA1} \rightharpoonup \mathrm{TwII} \cdot \mathrm{rA1}$
	$\mathrm{TwII} + \mathrm{rA1} + \mathrm{rA1} \rightharpoonup \mathrm{TwII} \cdot \mathrm{rA1} \cdot \mathrm{rA1}$
Recapturing:	$\mathrm{TwII} \cdot \mathrm{rA1} + \mathrm{dI1} \rightharpoonup \mathrm{TwII} + \mathrm{dI1} \cdot \mathrm{rA1}$
Recovering:	$\mathrm{TwII} \cdot \mathrm{rA1} + \mathrm{A1} \rightharpoonup \mathrm{TwII} \cdot \mathrm{A1} + \mathrm{rA1}$
Degradation:	$\mathrm{TwII} \cdot \mathrm{rA1} + \mathrm{RNaseH} \rightleftharpoons \mathrm{TwII} \cdot \mathrm{rA1} \cdot \mathrm{RNaseH} \rightharpoonup \mathrm{TwII} + \mathrm{RNaseH}$
	$\mathrm{TwII} \cdot \mathrm{rA1} \cdot \mathrm{rA1} + \mathrm{RNaseH} \rightleftharpoons \mathrm{TwII} \cdot \mathrm{rA1} \cdot \mathrm{rA1} \cdot \mathrm{RNaseH} \rightharpoonup \mathrm{TwII} + \mathrm{RNaseH}$
Double binding:	$\mathrm{TwII} + \mathrm{A1} + \mathrm{A1} \rightharpoonup \mathrm{TwII} \cdot \mathrm{A1} \cdot \mathrm{A1}$
Clearing:	$\mathrm{TwII} \cdot \mathrm{A1} \cdot \mathrm{A1} + \mathrm{dI1} + \mathrm{dI1} \rightharpoonup \mathrm{dI1} \cdot \mathrm{A1} + \mathrm{dI1} \cdot \mathrm{A1} + \mathrm{TwII}$

Tweezers mode III

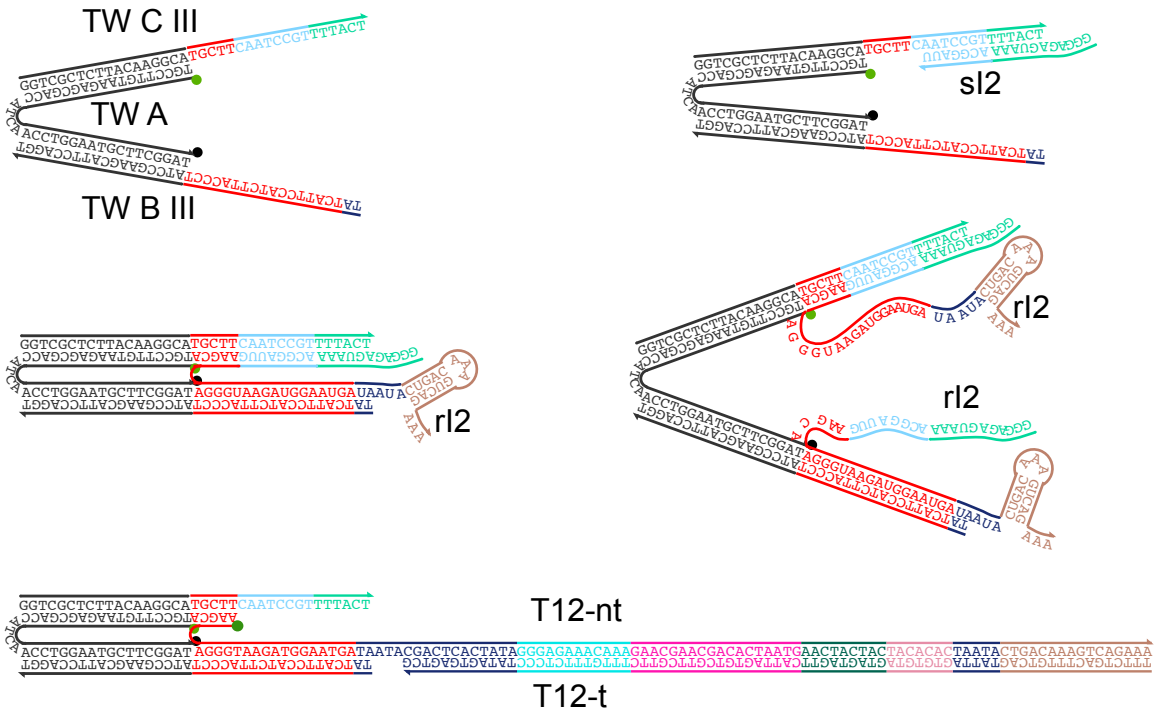


Figure 3.17: Relevant interactions for tweezers mode III. From top to bottom: open tweezers and incomplete RNase H degradation product sI2 binding to the tweezers; RNA target rI2 closing the tweezers; undesired complex TwIII·T12. The unwanted complex has a 14-base exposed toehold that can be targeted by rI2.

Reaction Pathway Examples	
Closing:	$\mathrm{TwIII} + \mathrm{rI2} \rightharpoonup \mathrm{TwIII} \cdot \mathrm{rI2}$
Double binding:	$\mathrm{TwIII} + \mathrm{rI2} + \mathrm{rI2} \rightharpoonup \mathrm{TwIII} \cdot \mathrm{rI2} \cdot \mathrm{rI2}$
Opening/Degradation:	$\mathrm{TwIII} \cdot \mathrm{rI2} + \mathrm{RNaseH} \rightleftharpoons \mathrm{TwIII} \cdot \mathrm{rI2} \cdot \mathrm{RNaseH} \rightharpoonup \mathrm{TwIII} + \mathrm{RNaseH}$
	$\mathrm{TwIII} \cdot \mathrm{rI2} \cdot \mathrm{rI2} + \mathrm{RNaseH} \rightleftharpoons \mathrm{TwIII} \cdot \mathrm{rI2} \cdot \mathrm{rI2} \cdot \mathrm{RNaseH} \rightharpoonup \mathrm{TwIII} + \mathrm{RNaseH}$
Opening/Branch migration:	$\mathrm{TwIII} \cdot \mathrm{rI2} \cdot \mathrm{rI2} + \mathrm{A2} \rightharpoonup \mathrm{TwIII} \cdot \mathrm{rI2} + \mathrm{A2} \cdot \mathrm{rI2}$
Interfering:	$\mathrm{TwIII} + \mathrm{T12} \rightharpoonup \mathrm{TwIII} \cdot \mathrm{T12}$
Recapturing:	$\mathrm{TwIII} \cdot \mathrm{T12} + \mathrm{rI2} \rightharpoonup \mathrm{TwIII} \cdot \mathrm{T12} \cdot \mathrm{rI2}$
Recovering:	$\mathrm{TwIII} \cdot \mathrm{T12} + \mathrm{A2} \rightharpoonup \mathrm{TwIII} + \mathrm{T12} \cdot \mathrm{A2}$

Tweezers mode IV

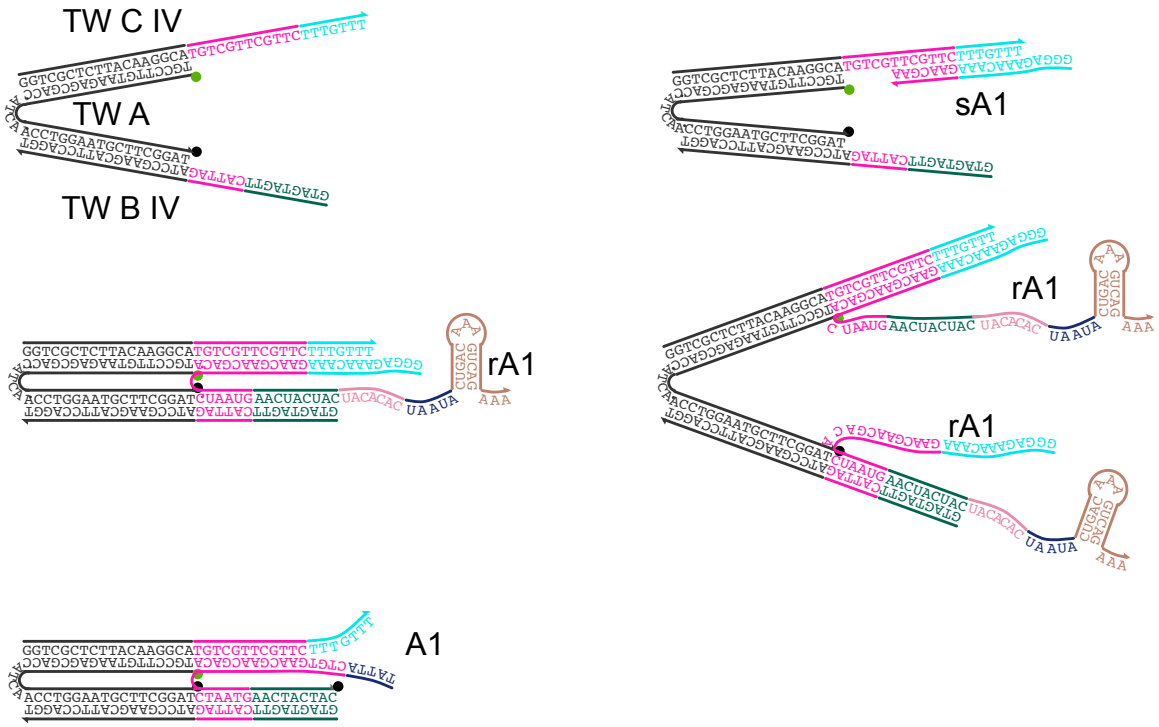


Figure 3.18: Relevant interactions for tweezers mode IV. From top to bottom: open tweezers and incomplete RNase H degradation product sA1 binding to the tweezers; RNA target rA1 closing the tweezers (this complex can be opened either by RNase H degradation or toehold-mediated strand migration by dI1 on the 7-base exposed light pink domain of rA1); undesired interaction with DNA species A1. The desired target complex TwIV·rA1 shares 7 more bases than the undesired complex TwIV· A1.

Reaction Pathway Examples	
Closing:	$\mathrm{TwIV} + \mathrm{rA1} \rightharpoonup \mathrm{TwIV} \cdot \mathrm{rA1}$
Double binding:	$\mathrm{TwIV} + \mathrm{rA1} + \mathrm{rA1} \rightharpoonup \mathrm{TwIV} \cdot \mathrm{rA1} \cdot \mathrm{rA1}$
Opening:	$\mathrm{TwIV} \cdot \mathrm{rA1} + \mathrm{dI1} \rightharpoonup \mathrm{rA1} \cdot \mathrm{dI1} + \mathrm{TwIV}$
Clearing:	$\mathrm{TwIV} \cdot \mathrm{rA1} \cdot \mathrm{rA1} + \mathrm{dI1} + \mathrm{dI1} \rightharpoonup \mathrm{rA1} \cdot \mathrm{dI1} + \mathrm{rA1} \cdot \mathrm{dI1} + \mathrm{TwIV}$
	$\mathrm{TwIV} \cdot \mathrm{A1} + \mathrm{dI1} \rightharpoonup \mathrm{TwIV} + \mathrm{dI1} \cdot \mathrm{A1}$
Degradation:	$\mathrm{TwIV} \cdot \mathrm{rA1} + \mathrm{RNaseH} \rightleftharpoons \mathrm{TwIV} \cdot \mathrm{rA1} \cdot \mathrm{RNaseH} \rightharpoonup \mathrm{TwIV} + \mathrm{RNaseH}$
	$\mathrm{TwIV} \cdot \mathrm{rA1} \cdot \mathrm{rA1} + \mathrm{RNaseH} \rightleftharpoons \mathrm{TwIV} \cdot \mathrm{rA1} \cdot \mathrm{rA1} \cdot \mathrm{RNaseH} \rightharpoonup \mathrm{TwIV} + \mathrm{RNaseH}$
Interfering:	$\mathrm{TwIV} + \mathrm{A1} \rightharpoonup \mathrm{TwIV} \cdot \mathrm{A1}$
Recovering:	$\mathrm{TwIV} \cdot \mathrm{A1} + \mathrm{rA1} \rightharpoonup \mathrm{TwIV} \cdot \mathrm{rA1} + \mathrm{A1}$

Mode V* – Insulator A1

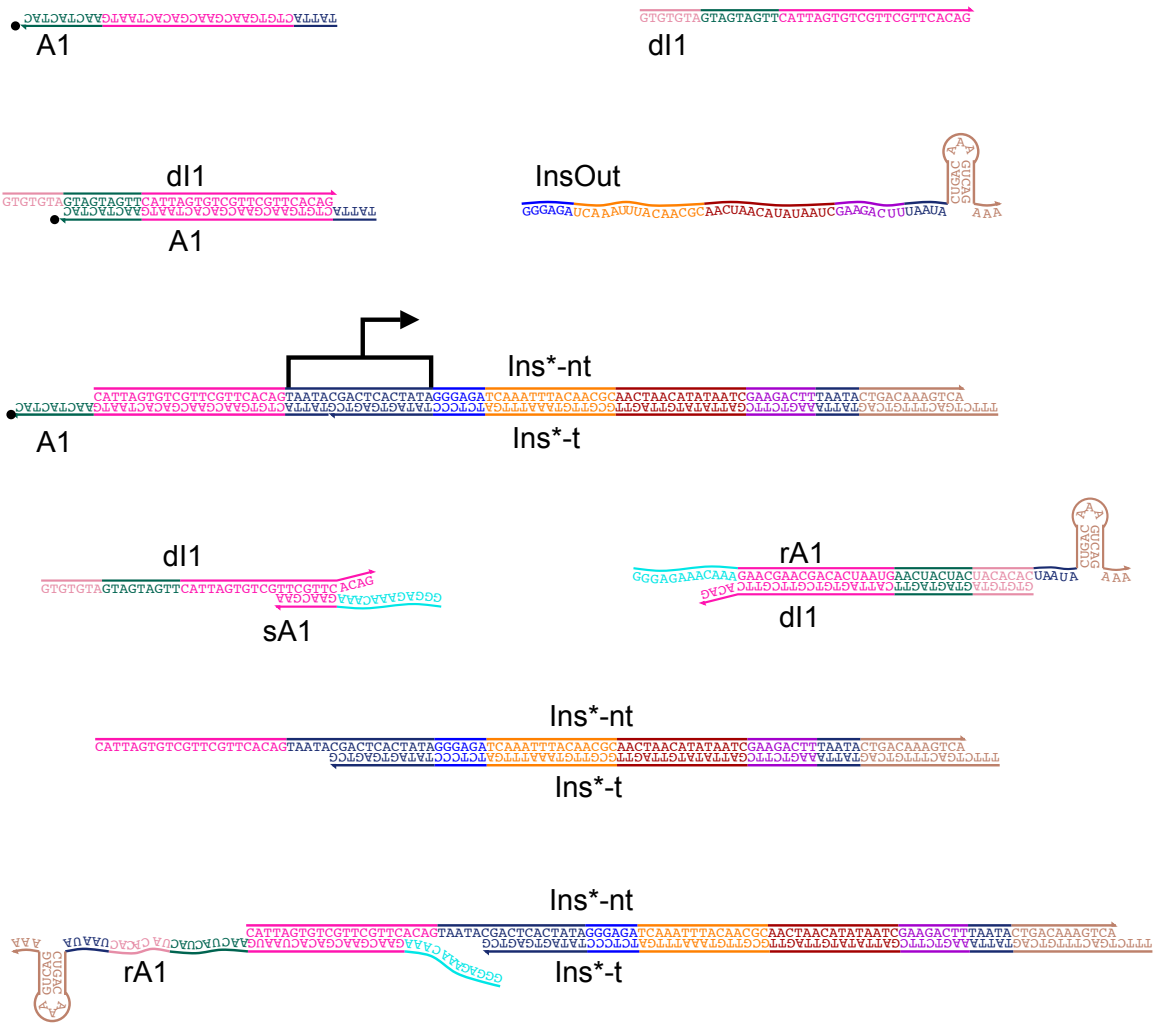


Figure 3.19: Relevant interactions for the mode V* insulator, with input A1. This insulator switch has the same input domain of SW21, and therefore the same complexes and side reactions as in Figure 3.14 are represented. The output of the insulator is denoted as InsOut.

	Reaction Pathway Examples
Activation:	$A1 + Ins^* \rightharpoonup Ins^* \cdot A1$
Inhibition:	$dI1 + Ins^* \cdot A1 \rightharpoonup dI1 \cdot A1 + Ins^*$
Release:	$rA1 + dI1 \cdot A1 \rightharpoonup rA1 \cdot dI1 + A1$
Annihilation:	$dI1 + A1 \rightharpoonup dI1 \cdot A1$
	$rA1 + dI1 \rightharpoonup rA1 \cdot dI1$
Transcription:	$Ins^* \cdot A1 + RNAP \rightleftharpoons Ins^* \cdot A1 \cdot RNAP \rightharpoonup InsOut + Ins^* \cdot A1 + RNAP$
	$Ins^* + RNAP \rightleftharpoons Ins^* \cdot RNAP \rightharpoonup InsOut + Ins^* + RNAP$
Degradation:	$rA1 \cdot dI1 + RNaseH \rightleftharpoons rA1 \cdot dI1 \cdot RNaseH \rightharpoonup dI1 + RNaseH$
	$Ins^* \cdot rA1 + RNaseH \rightleftharpoons Ins^* \cdot rA1 \cdot RNaseH \rightharpoonup Ins^* + RNaseH$
Interfering:	$rA1 + Ins^* \rightharpoonup Ins^* \cdot rA1$
Recapturing:	$dI1 + Ins^* \cdot rA1 \rightharpoonup rA1 \cdot dI1 + Ins^*$
Recovering:	$A1 + Ins^* \cdot rA1 \rightharpoonup Ins^* \cdot A1 + rA1$

Mode V – Insulator A2

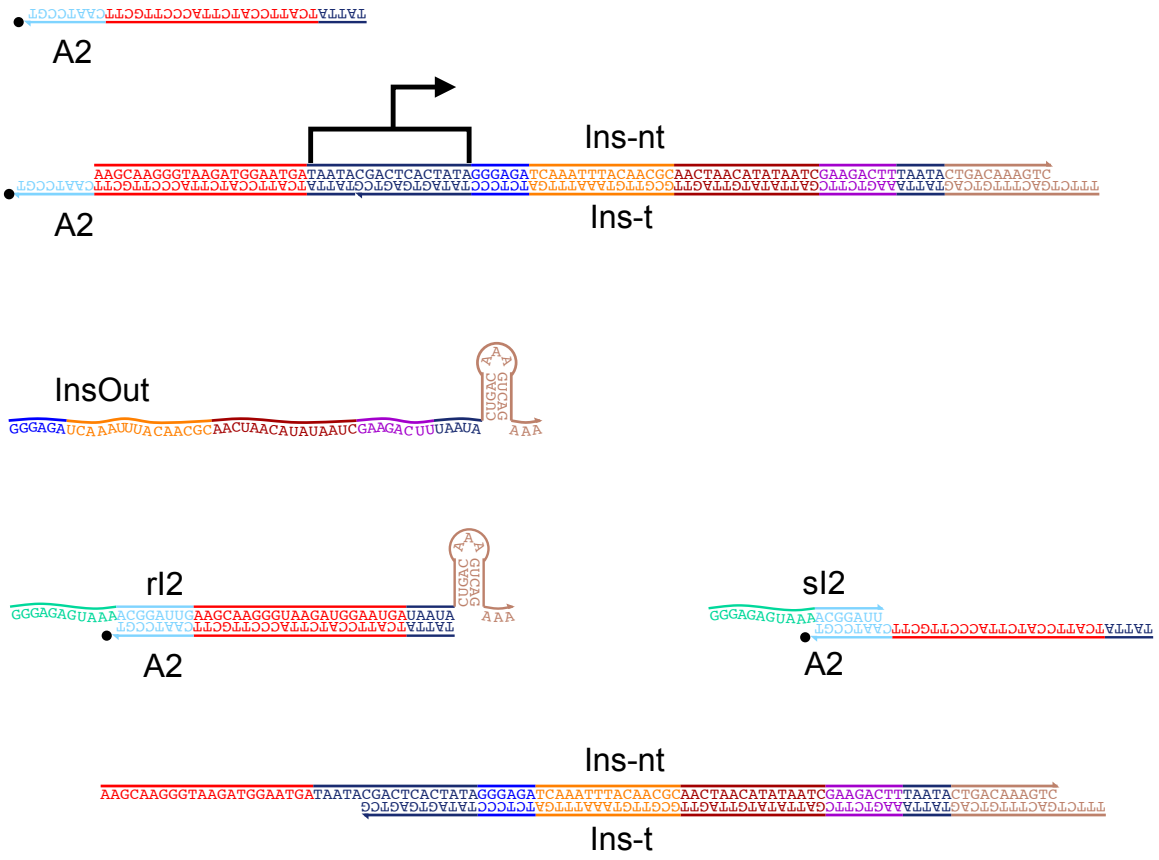


Figure 3.20: Relevant interactions for the mode V insulator, with input A2. This insulator switch has the same input domain of SW12, and therefore the same complexes as in Figure 3.13 are represented. The output of the insulator is denoted as InsOut, and is the same as in the insulator having input A1, shown in Figure 3.19.

Reaction Pathways	
Activation:	$A2 + Ins \rightharpoonup Ins \cdot A2$
Inhibition:	$rI2 + Ins \cdot A2 \rightharpoonup rI2 \cdot A2 + Ins$
Annihilation:	$rI2 + A2 \rightharpoonup rI2 \cdot A2$
Transcription:	$Ins \cdot A2 + RNAP \rightleftharpoons Ins \cdot A2 \cdot RNAP \rightharpoonup InsOut + Ins \cdot A2 + RNAP$
	$Ins + RNAP \rightleftharpoons Ins \cdot RNAP \rightharpoonup InsOut + Ins + RNAP$
Degradation:	$rI2 \cdot A2 + RNaseH \rightleftharpoons rI2 \cdot A2 \cdot RNaseH \rightharpoonup sI2 \cdot A2 + RNaseH$

Tweezers mode V

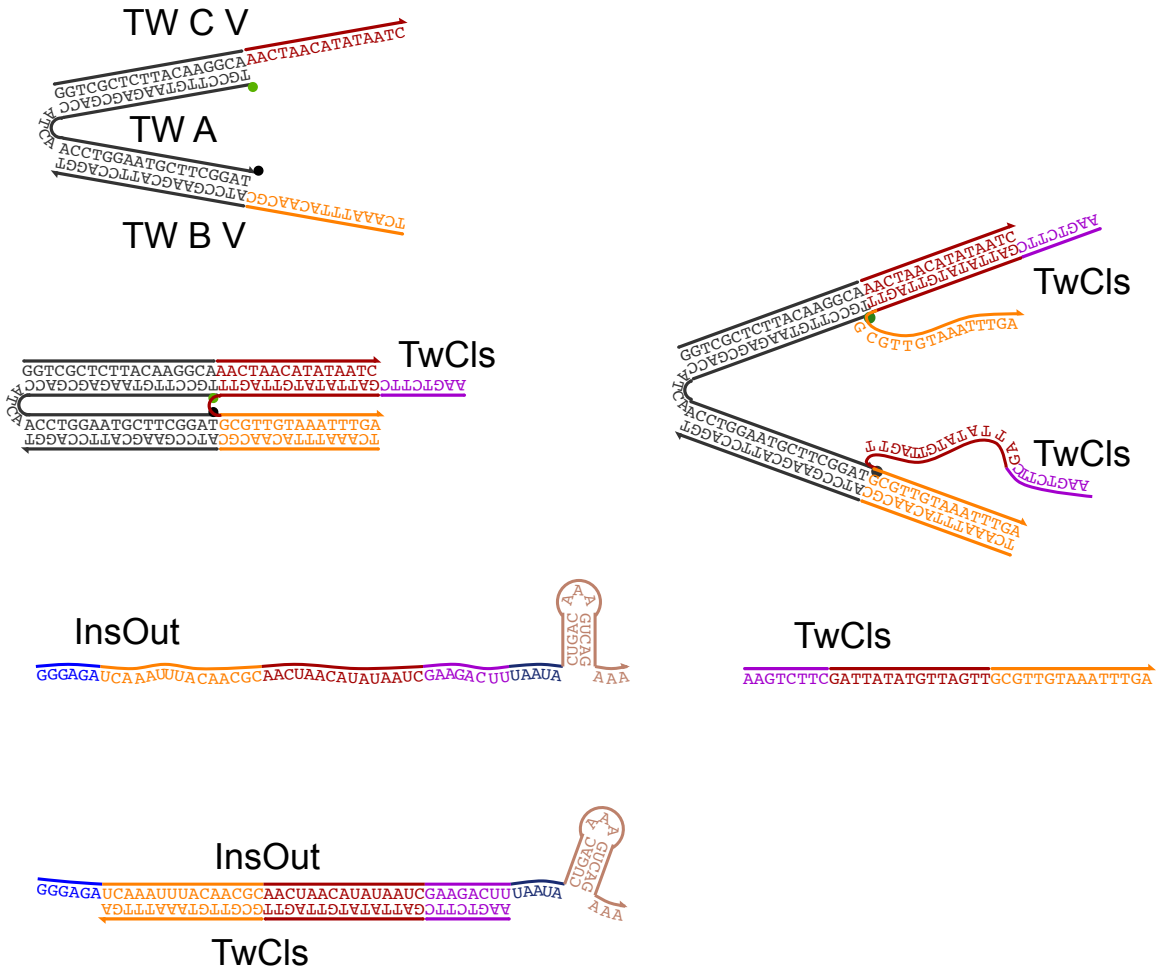


Figure 3.21: Relevant interactions for the load stage of tweezers mode V. From top to bottom, left to right: tweezers mode V, open; tweezers mode V closed by their target TwCls; tweezers mode V bound to two target molecules TwCls; RNA strand InsOut (output of the insulators at Figures 3.19 and 3.20); DNA strand TwCls; TwCls displaced by InsOut.

Reaction Pathways	
Closing:	$\mathrm{TwV} + \mathrm{TwCls} \rightharpoonup \mathrm{TwV} \cdot \mathrm{TwCls}$
Opening:	$\mathrm{TwV} \cdot \mathrm{TwCls} + \mathrm{InsOut} \rightharpoonup \mathrm{TwCls} \cdot \mathrm{InsOut} + \mathrm{TwV}$
Annihilation:	$\mathrm{TwCls} + \mathrm{InsOut} \rightharpoonup \mathrm{TwCls} \cdot \mathrm{InsOut}$
Degradation:	$\mathrm{TwCls} \cdot \mathrm{InsOut} + \mathrm{RNaseH} \rightleftharpoons \mathrm{TwCls} \cdot \mathrm{InsOut} \cdot \mathrm{RNaseH} \rightharpoonup \mathrm{TwCls} + \mathrm{RNaseH}$

Unmodeled reactions	
Double binding:	$\mathrm{TwV} + \mathrm{TwCls} + \mathrm{TwCls} \rightharpoonup \mathrm{TwV} \cdot \mathrm{TwCls} \cdot \mathrm{TwCls}$
Clearing:	$\mathrm{TwV} \cdot \mathrm{TwCls} \cdot \mathrm{TwCls} + \mathrm{InsOut} + \mathrm{InsOut} \rightharpoonup \mathrm{TwV} + \mathrm{TwCls} \cdot \mathrm{InsOut} + \mathrm{TwCls} \cdot \mathrm{InsOut}$

3.7.3 Sample notation

Several fluorescence data sets were collected for this project, at two different institutions: Caltech and Technical University in Munich (TUM). Within each set, identical DNA stock solutions and enzyme batches are used. Here, I will follow the notation originally chosen in the Supplementary Appendix of [41] to denote each data set. Sets denoted as A, B, 1, 2, and 3 were recorded at TUM; sets denoted as 4, 5, and 6 were recorded at Caltech. In this document, I will only report detailed information on the sets collected by me at Caltech. Each data set consists of several rounds of data acquisition with four samples each. One sample from each round usually is a reference sample containing the oscillator system only, unless otherwise noted. Data acquisition of one round takes one day. We will use the notation TWI-4 (for example) to indicate the mode I tweezers experiments of set 4. For more details about sample preparation please refer to Section 3.7.4.

Table 3.2: Data Sets Acquisition Overview: Caltech

SET	**Modes:**				
4	tweezers I (TWI-4)	tweezers II (TWII-4)	tweezers III (TWIII-4)	tweezers IV (TWIV-4)	tweezers V (TWV-4)
5	threshold variation I	threshold variation II	threshold variation III		
6	tweezers II (TWII-6)	tweezers II* (TWII*-6)	tweezers V* (TWV*-6)	tweezers V (TWV-6)	

3.7.4 Sample preparation

The protocols followed at Caltech are thoroughly described in this section. The transcription protocols differ mainly in the brand of the reagents used, the RNA polymerase (RNAP) and RNase H handling method, the concentration of ribonucleoside triphosphates (rNTPs), and the DNA activator and inhibitor thresholds.

Operating point

We defined our operating point for the oscillator as a trajectory providing 4–6 oscillations in 16 hours, with amplitude of 80–120 nM.

The dynamics of the core oscillator are a function of several variables: DNA concentrations, buffer composition, and concentration and activity of RNAP and RNase H. Most of our experiments were done using the Ambion T7 Megashortscript kit, for which transcription buffer and enzyme mix composition are not disclosed; each production batch is slightly tuned by the vendor to maximize transcription speed. In general, the characteristics of all off-the-shelf enzymes may vary among stocks purchased from the same supplier at different times. Therefore, all else being equal, using enzymes (and transcription buffer) from different production batches may result in significantly

different oscillation amplitude and frequency.

To achieve a consistent operating point for all our experiments, we tuned the enzyme volumes (and, in some cases, the buffer composition) when switching to a new enzyme stock. We also empirically found that the concentration of DNA thresholds (A1, A2, and dI1) influences the operating point of the system, as shown in Figure 3.40. Therefore, we adjusted the thresholds as part of our tuning procedure to reach the desired operating point.

Sets 4, 5

The final concentrations of the oscillator DNA strands were: T12 (120 nM), T21 (250 nM), dI1 (650 nM), A1 (300 nM), and A2 (550 nM). T21, T12, insulator genes, and the TW A, TW B, and TW C strands for each tweezer mode were annealed separately in 1x Ambion Megashortscript kit buffer in a digital thermal cycler (MJ Mini 48Well Personal Thermal Cycler, Bio-Rad Laboratories, Inc.) by heating for 1 minute at 95°C and cooling to room temperature in 2 h.

A transcription buffer mix was prepared before each experiment run (for four samples) to a final concentration of 1x Transcription Buffer, 7.5 mM each rNTP (Ambion Megashortscript Kit, Cat. n. AM1354), and nanopure water as appropriate. The $MgCl_2$ concentration was not adjusted. To maintain constant enzyme ratios, T7 RNAP enzyme mix (Ambion Megashortscript Kit, Cat. n. AM1354) and RNase H (Ambion, *E. coli* cloned, Cat. n. AM2292, 10 U/μl) were premixed once for each round of data acquisition (for four samples). The total enzyme pre-mix volume always exceeded by $\approx 10\%$ the volume required in the experiment.

Each experiment was carried out as follows: First, all DNA strands and all the transcription reagents were pre-mixed in two separate test tubes. The amount of each reagent is calculated to achieve the desired concentration in a final total 240 μl sample volume. The mixes were then split into the four quartz cuvettes of the data acquisition round, each to have a final volume of 60 μl. The cuvettes were pre-warmed at 37°C in the Fluorolog 3 sample chamber. The transcription buffer mix was added first, followed by the DNA mix aliquot. We found that this procedure minimized the variability of fluorescence traces across samples in the same data acquisition round.

Each sample was sealed using 35 μl of hexadecane (MP Biomedicals, LLC, Cat. n.195218) to prevent evaporation. After measuring the initial off-state fluorescence for the switch T21 in the spectrofluorimeter, the enzyme mix was added to each sample in the appropriate volume (Table 3.3), through the sealing oil layer.

Set 6

Final concentrations of the oscillator DNA strands were: T12 (120 nM), T21 (250 nM), dI1 (600 nM), A1 (250 nM), and A2 (500 nM). The oscillator switches, insulators, and the TW A, TW B, and TW C strands for each tweezers mode were annealed separately in 1x Transcription Buffer (Epicentre Biotechnologies, Cat. n. BP1001), undergoing the same thermal treatment as sets 4 and

5.

The transcription buffer mix was prepared prior to each experiment run (for four samples), mixing reagents to the following final concentrations: 1x Transcription Buffer and 10 mM dithiothreitol (DTT) (Epicentre Biotechnologies, Cat. n. BP1001), 7.5 mM each rNTP (Epicentre Biotechnologies, Cat. n. RN02825), 35 mM $MgCl_2$, and 0.015 U/μl pyrophosphatase (Sigma Aldrich, Cat. n. I1891-100UN, resuspended in Tris HCl 20 mM, pH 7.2, 50% glycerol (v/v)).

T7 RNAP was purchased from Epicentre Biotechnologies, Cat. n. TM910K (200 U/μl). *E. coli* cloned RNase H was purchased from Ambion, Cat. n. AM2292 (10 U/μl).

Each step of the experiments done for set 6 followed closely the procedure described for sets 4 and 5.

Table 3.3: Overview of the Enzyme Volumes Used in the Data Sets Collected at Caltech. Measuring the concentration of RNAP and RNase H presents several challenges. For sets 4 and 5, I used the Ambion T7 enzyme mix provided with the T7 Megashortscript Kit; the vendor does not provide accurate information regarding the RNAP concentration or weight in each batch. According to [84], the mix contains inorganic pyrophosphatase; therefore, absorbance measurements can only provide approximated estimates of the RNAP concentration. For the numerical simulations, the nominal concentration of RNAP was assumed to be 1.25 μM, according to the absorbance measurements; a nominal concentration of 1.25 μM was assumed for RNase H, as quoted by the manufacturer. For set 6, I used Epicentre Biotechnology T7 RNAP: the concentration of the lot used in the experiments was 4 μM.

SET	$\mathbf{V^{sample}}(\mu\mathbf{l})$	$\mathbf{V}_{T7}(\mu\mathbf{l})$	$\mathbf{V}_{H}(\mu\mathbf{l})$	$\mathbf{V^{tot}_{EnzMix}}(\mu\mathbf{l})$	$\mathbf{V^{sample}_{EnzMix}}(\mu\mathbf{l})$	$\mathbf{V}_{T7}/\mathbf{V}_{H}$
4	60	22	2.2	24.2	5.6	10.0
5	60	22	2.2	24.2	5.6	10.0
6	60	18	2	20	4	9.0

Table 3.4: Overview of the Oscillator Total DNA Species Concentrations Used in Each Data Set.

SETS	**[T21] (nM)**	**[A1] (nM)**	**[dI1] (nM)**	**[T12] (nM)**	**[A2] (nM)**
4	250	300	650	120	550
5 (default)	250	300	600	120	500
6	250	250	600	120	500

3.7.5 Fluorescence data processing

The raw fluorescence measurements recorded with the Horiba Fluorolog 3 were converted into molar concentrations according to the procedure described in this section. The ideal formula to convert

each fluorescence trace to a concentration trace is:

$$C(\mathrm{t_n}) = \mathrm{C_0}\frac{\mathrm{I(t_n) - I_{min}}}{\mathrm{I_{max} - I_{min}}},$$

where C_0 is the nominal total concentration of the labeled DNA strand, $\mathrm{I(t_n)}$ is the raw fluorescence intensity measured by the instrument, and $\mathrm{I_{max}}$, $\mathrm{I_{min}}$ are the maximum and minimum fluorescence signals for the strand at that specific concentration and lamp intensity. In practice, $\mathrm{I_{max}}$ is the signal that would be measured if all the tweezers (or switch) in the sample were in a fully open (or off) state; accordingly, $\mathrm{I_{min}}$ corresponds to the signal measured in a fully closed tweezers/on state switch.

Unfortunately, the values of $\mathrm{I_{max}}$ and $\mathrm{I_{min}}$ may not be both available for each trace. However, we can re–write the above formula as:

$$C(\mathrm{t_n}) = \mathrm{C_0}\frac{\mathrm{I(t_n)/I_{max} - r}}{\mathrm{1 - r}},$$

where $\mathrm{r} = \mathrm{I_{min}/I_{max}}$. We were able to successfully use this formula because first, the minimum/maximum fluorescence signal ratio r should be independent of the specific strand concentrations and lamp intensity; therefore, we measured the r values off line. Second, we could determine $\mathrm{I_{max}}$ for most of our experiments; in some cases, we could instead determine $\mathrm{I_{min}}$, from which we could estimate $\mathrm{I_{max}}$=$\mathrm{I_{min}}$/r.

To determine the r value, we first collected the average maximum fluorescence intensity $\mathrm{I_{max}}$ from a calibration sample at known concentration, containing open tweezers (TWI–TWV) or genelet in the off state (T12 or T21). Then, we measured the average minimum quenched fluorescence $\mathrm{I_{min}}$, by adding to the sample the tweezer closing strand (A1, dI1, rI2, rA1 or TwCls) or activator (quencher-labeled) strand (A1 or A2) in excess. The on/off fluorescence ratio $\mathrm{r} = \mathrm{I_{min}}/\ \mathrm{I_{max}}$ was therefore calculated. As noted before, this ratio is independent of the instrument lamp intensity and of the concentration of labeled strands, provided that the closing strand is present in excess. However, we found that r is sensitive to the overall $MgCl_2$ and rNTPs concentration in the sample, in a sequence–dependent manner. In particular, we found that the closed state fluorescence of the tweezers is the most affected by the overall ionic concentration in the sample. Specifically, the closed state fluorescence is lower in samples with higher $MgCl_2$ concentration. The buffer $MgCl_2$ concentration of the T7 Ambion Megashortscript Kit is unknown: therefore, the normalization measurements should be carried out with the same kit used for the experiments. Table 3.5 shows the r and C_0 values used to normalize data sets I collected at Caltech.

The $\mathrm{I_{max}}$ value has to be determined as "maximum off/open-state" fluorescence level of the monitored strands for each experiment. The initial T21 off-state fluorescence value prior to addition of enzymes was recorded for each trace. Therefore, $\mathrm{I_{max}}$ was calculated as the initial off-state

fluorescence value, averaged across the four samples in the same data acquisition round, decreased by 10% to account for the dilution due to the subsequent addition of the enzyme mix to each sample. Regarding the determination of $\mathrm{I_{max}}$ for the T12 genelet, the data shown in this document at Figure 3.25 were normalized by observing that the genelet is in a fully on state before adding enzymes; so we used such initial data to determine $\mathrm{I_{min}}$ as the on-state fluorescence, and estimated $\mathrm{I_{max}} = \mathrm{I_{min}}/r$.

For the molecular tweezers, the value of $\mathrm{I_{max}}$ was determined depending on the mode. For Modes I and II, across all data sets, $\mathrm{I_{max}}$ was chosen to be the first peak of each trace. For mode II*, $\mathrm{I_{max}}$ was chosen as the maximum fluorescence value over the entire time trace (choosing $\mathrm{I_{min}}$ as the minimum over the trace prior to addition of enzymes, and then estimating $\mathrm{I_{max}} = \mathrm{I_{min}}/\mathrm{r}$ was not feasible, because 400 nM of tweezers may not be fully closed by the A2 amount in solution in case of pipetting inaccuracies). For Modes III and IV, $\mathrm{I_{max}}$ was chosen as the high-fluorescence signal measured prior to addition of enzymes, when no RNA (closing strand) is present in the system. For mode V, the value of $\mathrm{I_{max}}$ was taken as the fluorescence value measured prior to adding the closing strand for set 4; for Modes V and V* in set 6, $\mathrm{I_{max}}$ was instead chosen as the fluorescence value of the first large peak of the traces at high amount of insulator (specifically, the traces shown at Figures 3.47, last row, second column, and Figure 3.46, last row, second column): such value of $\mathrm{I_{max}}$ was consistent across the data taken for mode V and V*. (Since all the data acquisition rounds for set 6 were recorded in subsequent days, without turning off the spectrofluorimeter, we assumed that lamp fluctuations could be neglected.)

The RNA–operated tweezers, Modes III and V, presented a drift, which was not caused by evaporation but presumably by the accumulation of short incomplete degradation products binding to the tweezers' hands. The experiments on mode V and V* tweezers instead presented a drift in the off-state minimum fluorescence, presumably caused by the depletion of rNTPs and by the consequently higher availability of free positive ions of Mg^{++} in solution. These two effects produced normalized concentrations exceeding the total strand amount in solution, or normalized concentrations becoming negative over time. The data affected by such fluorescence drifts were processed with a modified normalization protocol. First, we estimated the minimum fluorescence level for the trace as $\mathrm{I_{min}} = \mathrm{r} \cdot \mathrm{I_{max}}$. Then, at each data point we computed $a(\mathrm{t_n}) = (\mathrm{I}(\mathrm{t_n}) - \mathrm{I_{max}})/\mathrm{t_n}$ and $b(\mathrm{t_n}) = (\mathrm{I_{min}} - \mathrm{I}(\mathrm{t_n}))/\mathrm{t_n}$. We finally set $\bar{a}$=max$\{a(\mathrm{t_n})\}$ and $\bar{b}$=$\{\max b(\mathrm{t_n})\}$ and normalized the data as follows:

$$\mathrm{I_{max}}(\mathrm{t_n}) := \mathrm{I_{max}} + \bar{\mathrm{a}} \cdot \mathrm{t_n}, \quad \mathrm{I_{min}} := \mathrm{I_{min}} - \bar{\mathrm{b}} \cdot \mathrm{t_n}$$
$$C(\mathrm{t_n}) = \frac{\mathrm{I}(\mathrm{t_n}) - \mathrm{I_{min}}(\mathrm{t_n})}{\mathrm{I_{max}}(\mathrm{t_n}) - \mathrm{I_{min}}(\mathrm{t_n})}.$$

This is equivalent to linearly correcting $\mathrm{I_{max}}$ and $\mathrm{I_{min}}$ with the minimal slope necessary to main-

tain all the data below or above the chosen thresholds. This correction was applied to the data shown in the following figures: Figure 3.44; Figure 3.45; Figure 3.46, all data sets; Figure 3.47, all data sets.

Figure 3.22B shows the TWII-4 TYE665 channel fluorescence raw data, converted into T21 concentrations as described above. The raw data and the corresponding concentrations for the Rhodamine Green channel, TWII-4, are shown in Figure 3.23, where we also highlight the effects of a 50% error in the estimation of r. Finally in Figure 3.24 we show the raw data for TWV-6, and the effects of the I_{min} linear correction.

Table 3.5: Parameters for Conversion of Fluorescence Intensities into Concentrations

species	**r= I_{MIN}/I_{MAX}**	**C_0(nM)**
T12 (TYE563)	0.056 (sets 4, 5), 0.045 (set 6)	120
T21 (TYE665)	0.051 (sets 4, 5), 0.05 (set 6)	250
TWI	0.31 (set 4)	100, 200, 400
TWII	0.31 (set 4); 0.3 (set 6)	100, 200, 400
TWII*	0.45 (set 6)	100, 200, 400
TWIII	0.35 (set 4)	100, 200, 400
TWIV	0.47 (set 4)	100, 200, 400
TWV	0.49 (set 4), 0.5 (set 6)	100, 200, 400, 600, 800

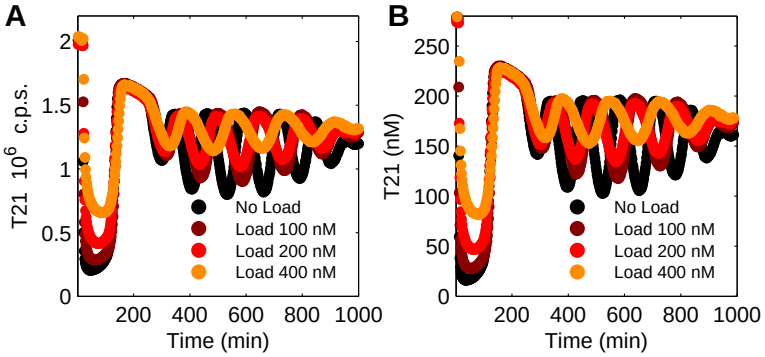


Figure 3.22: TWII-4 samples, T21 (TYE 665) channel. A. Raw fluorescence data. B. Fluorescence data converted to concentrations.

3.7.6 T12-channel data

Throughout this thesis chapter, only the T21-channel data (TYE665, or Texas Red for the data taken at TUM) are used to characterize the behavior of the oscillator. The T12 fluorescence data (TYE563, or TAMRA for the data taken at TUM) were also monitored in each experiment, but their small amplitude does not allow us to derive significant information about the system behavior.

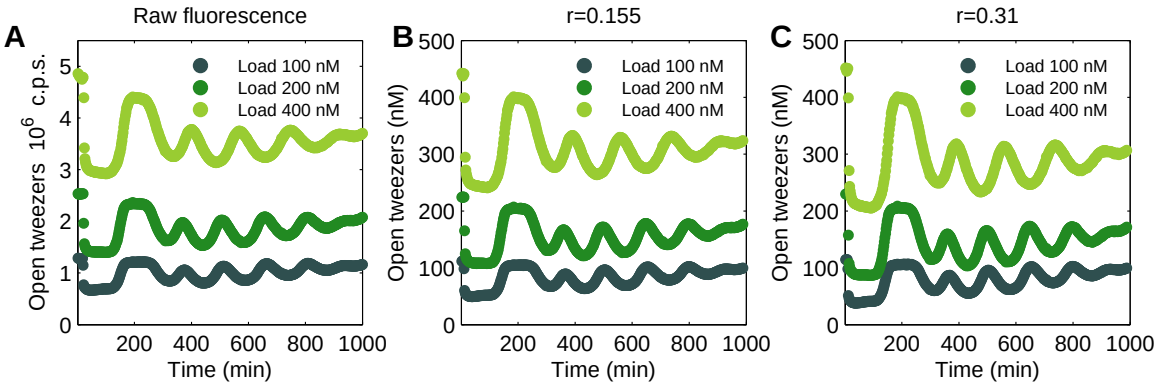


Figure 3.23: TWII-4 samples, TW II (Rhodamine Green) channel. A. Raw fluorescence data. B. Raw data converted to concentrations for TW II, with r=0.155, 50% lower than the correct value, r=0.31. C. Raw data converted to concentrations for TW II, with r=0.31.

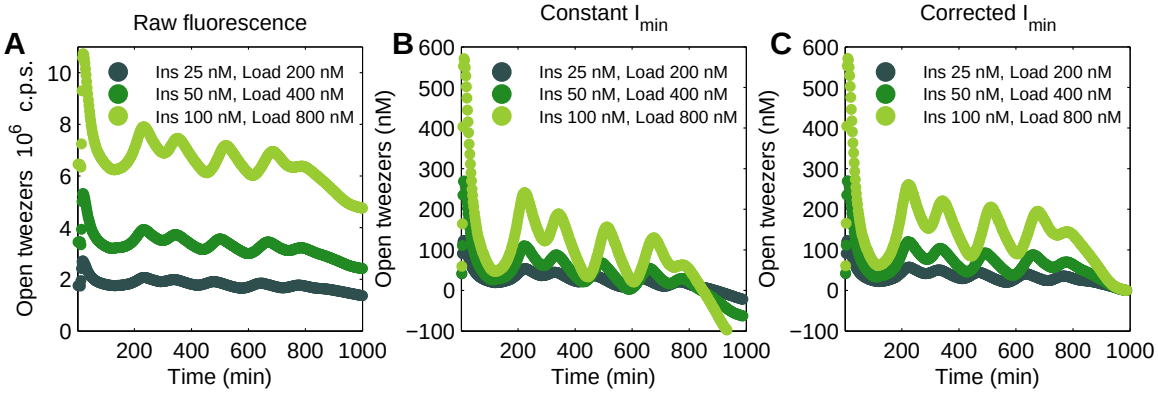


Figure 3.24: TWV-6 samples, TWV (Rhodamine Green) channel. These data corresponds to the data plotted in Figure 3.46, third row, second column. A. Raw fluorescence data for TWV-6. B. Raw data converted to concentrations for TWV-6, with constant I_{min}=r· I_{max}. C. Raw data converted to concentrations for TWV-6, with linearly adjusted I_{min}.

(This is consistent with the findings in [64].) In Figure 3.25 A, as an example, we show both the T21 and T12 normalized traces from the TWII-4 sample. (The concentrations of T12 and T21 belonging to the same sample are shown in the same color.) However, only the T21-concentration shows strong oscillations; the T12-concentration oscillates only weakly around 100 nM. For this specific trajectory, the average concentrations are <T12> = 98 nM and <T21>= 163 nM. This gives <T12A2>≈(120 - 98) nM = 22 nM and <T21A1> ≈(250 - 163) = 87 nM. Across data sets 4 and 6, we found that <T12A2> is on average 20 nM, and <T21A1> is on average 75 nM. From gel electrophoresis data (see Figure 3.35) we also know that given equal amounts of on–templates, rI2 is produced in much higher amount than rA1. While the same promoter is used for both templates, the k_{cat} values for each switch might vary and be influenced by the transcribed domain. T12 is only activated for a short time in each cycle, resulting in presumably small waves of rA1 production. Looking at the phase portrait of the sample (Figure 3.25 B), the concentration variation of T21/T21A1, Δ[T21]= Δ[T21A1] ≈ 100 nM is much larger than that of T12/T12A2 (Δ[T12]=Δ[T12A2] ≈ 10 nM). This demonstrates an asymmetry between the Switches 12 and 21: the state of SW21 is very sensitive to small changes in the state of SW12, while the state of SW12 does not respond analogously to variations in the state of SW21. Qualitatively, T12 exhibits the same behavior (small waves of modulation) in all samples and is therefore omitted in the data presented.

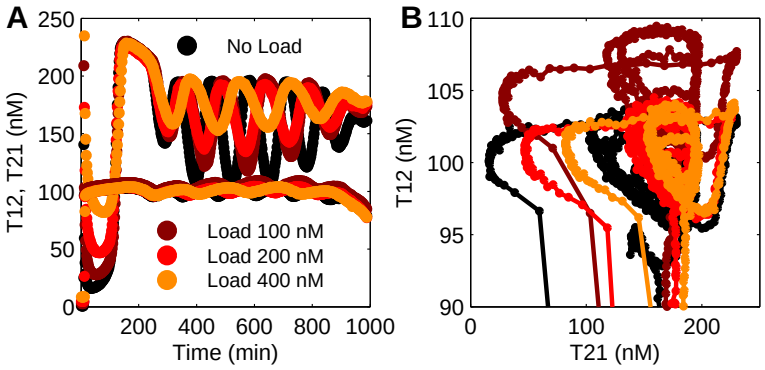


Figure 3.25: Trajectories of SW12 and SW21. A. Time traces. B. Phase portrait. The oscillation amplitude of SW12 is about one order of magnitude smaller than the amplitude of SW21. The traces of SW12 are therefore ignored in the analysis proposed in this project.

3.7.7 Analysis of the oscillations

The period and amplitude of the oscillations are time-varying in most of the experiments. To be able to compare different oscillatory traces, we processed the data in order to consider only the first three full oscillations after the first large and irregular peak.

Period: We calculate the average oscillation period from the first three full oscillations, measured between the first and the fourth minimum.

$$T(n) = t_{min}(n+1) - t_{min}(n)$$

The mean period of each sample is defined by $< T >= (T(1) + T(2) + T(3))/3$, if there are at least three full oscillations, otherwise the number is reduced accordingly.

Amplitude: We measure the oscillation amplitude of oscillation cycle n as half the difference between the peaks as defined in Figure 3.26:

$$A(n) = (c_{max}(n) - c_{min}(n))/2.$$

The mean amplitude for each time trace is defined as $< A >= (A(1) + A(2) + A(3))/3$.

As an example, in Figure 3.26 we show how the peaks and wells were selected for the control sample of Tweezers mode II in data set 3.

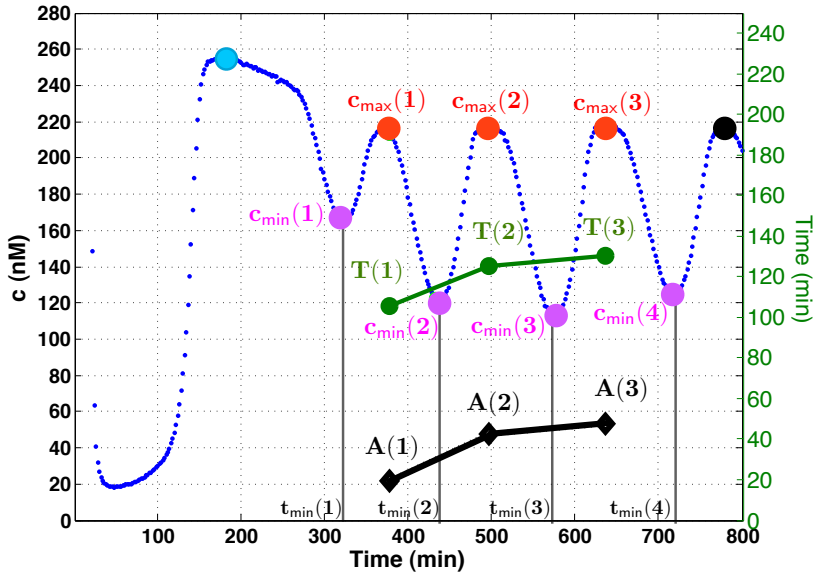


Figure 3.26: TWI-4. Red circles: selected oscillation maxima. Pink circles: selected oscillation minima. The cyan circle indicates the first large oscillation peak, not selected for period/amplitude analysis. Green trace: period per oscillation. Black trace: amplitude per oscillation. (The last peak highlighted in black is not used for the period and amplitude calculations.)

3.7.8 Day-to-day variability

In this section, I will highlight the oscillation variability introduced by pre-mixing of enzymes for use in different days.

The enzymes were instead pre-mixed right before each experiment run. Pipetting errors are particularly likely when transferring small enzyme volumes, due to their typical $\approx$ 50% glycerol storage buffer. To minimize such errors, the volumes transferred were always larger or equal to 2 μl. The total pre-mix volume always exceeded by 20% the volume required in the experiment. Such pipetting precautions allowed us to achieve a fairly stable period across different experiments, as shown in Figure 3.27, orange trace. Data set 6 (Epicentre Biotechnology reagents), brown trace, showed a more pronounced day-to-day variability in period and amplitude. This is most likely due to the fact that several transcription buffer components (and not only RNAP and RNase H as in the Ambion protocol) were mixed prior to each data acquisition round (see Section 3.7.4 for details on the protocol). For set 6, the amplitude is in particular reduced by the second small peak present in most traces.

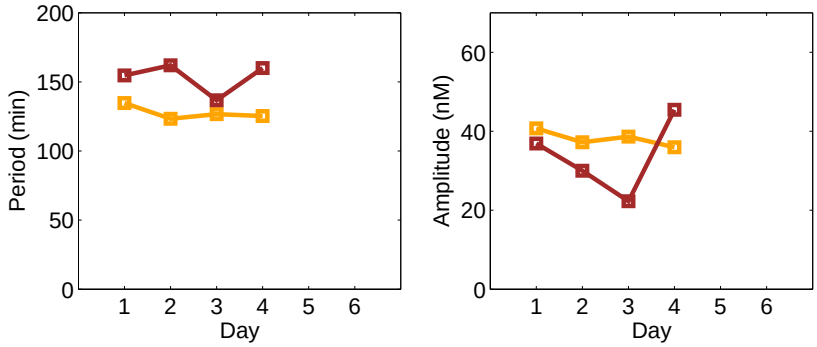


Figure 3.27: Period (left) and amplitude (right) versus acquisition round number (day of acquisition). Set 4 rounds (orange) include TWI-4, TWII-4, TWIII-4, TWIV-4. The corresponding time traces are presented in Section 3.7.18. Set 6 rounds (brown) include TWV* control from Figure 3.47, top; TWV* control from Figure 3.47, bottom; TWV control from Figure 3.46, third row; TWV control from Figure 3.46, fourth row.

3.7.9 Set-to-set variability

This section considers the overall variability of the oscillator operating point across different data sets, and includes processed data from the experiments collected at TUM. Figure 3.28 shows the mean and standard deviation of period and amplitude for all fluorescence data collected at Caltech and TUM.

Data set 6 presents high variability, particularly in the amplitude, because more reagents (Epicentre Biotechnology) are sequentially added in the transcription protocol, compared to the data sets collected using the T7 Ambion Megashortscript kit. Figure 3.29 shows a plot of period versus amplitude across different data sets.

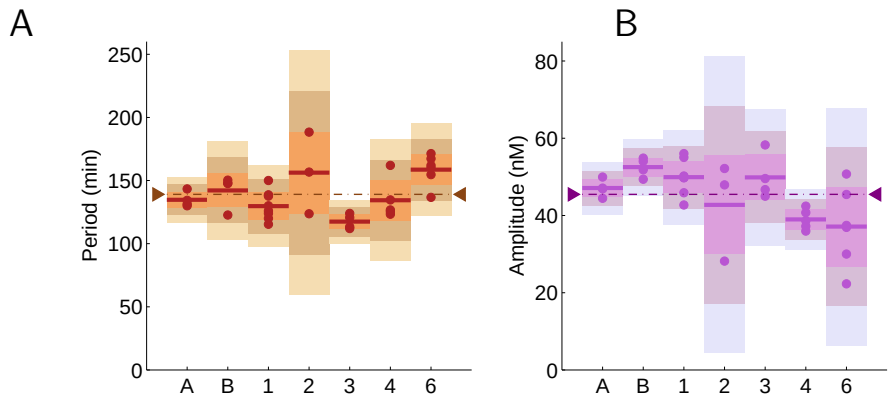


Figure 3.28: Set-to-set variability of period (A) and amplitude (B) of the oscillations. This figure shows all the processed fluorescence data collected on the oscillator in isolation, both at Caltech and TUM. For each data set, data are marked with solid dots, their mean is a solid dark line, and the shaded areas cover the one, two and three standard deviation areas. The overall mean across data sets is indicated as a dashed-dotted line with inward pointing arrows.

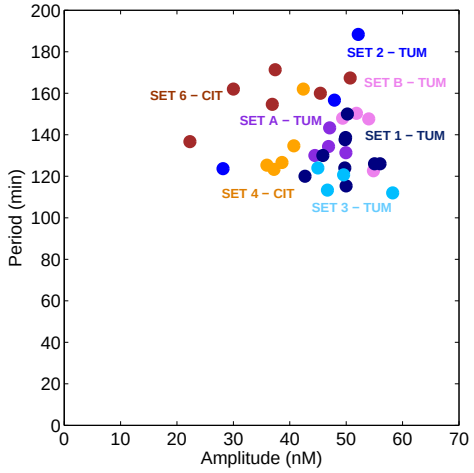


Figure 3.29: Overview of the period and amplitude correlation for all the control samples collected for this project, including all data gathered at Caltech and TUM.

3.7.10 Oscillation period

The period per each oscillation is measured between the first and the fourth minimum of each time trace. The nth minimum is defined as in Figure 3.26. Figure 3.30 gives an overview across different data sets for TWII, displaying the period from minimum n to minimum $n+1$.

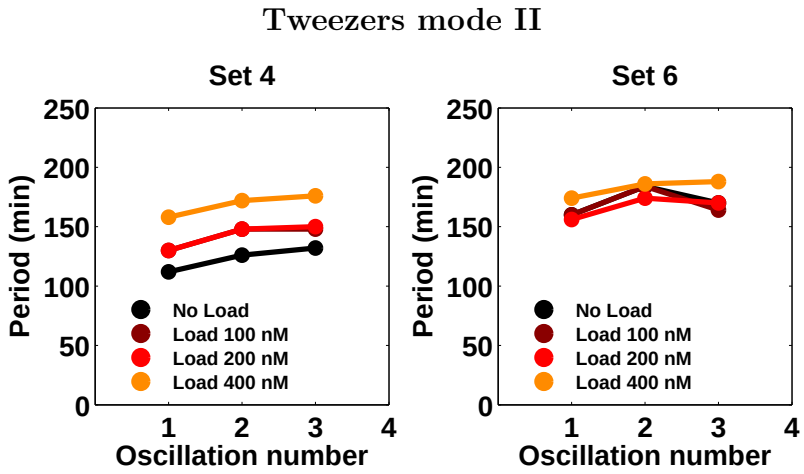


Figure 3.30: TWII, Period T(n) of each oscillation in the first three full oscillations.

3.7.11 Oscillation amplitude

Figure 3.31 gives an overview of the oscillator amplitudes as a function of the oscillation cycle number for Tweezers mode II, across different data sets. Recall that the oscillation amplitude of oscillation cycle n is calculated as half the difference between the concentration at maximum and the preceding minimum, defined as in Figure 3.26:

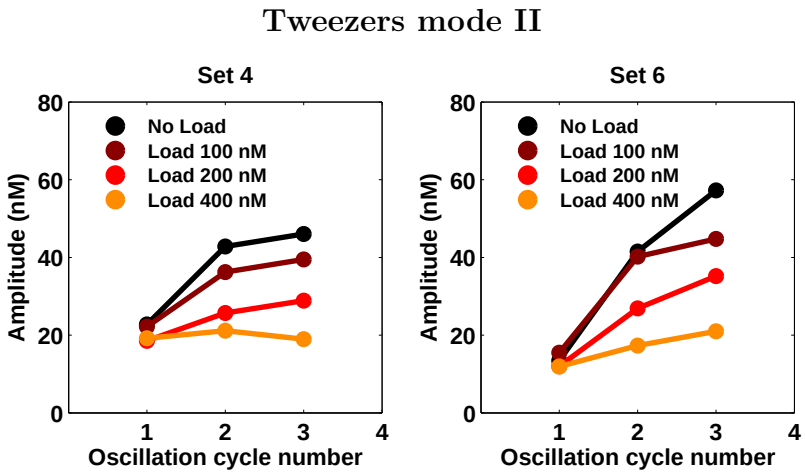


Figure 3.31: Amplitude A(n) of the oscillations in the first three full oscillations. The higher variability of sets 4 and 6 is due to the fact that in these sets the first oscillation after the plateau peak is consistently smaller than the following ones. Moreover, the protocol used for data set 6 requires more reagents to be sequentially added to the transcription mix, increasing the probability of pipetting inaccuracies.

3.7.12 Effects of the load on the oscillator performance

This section analyzes in more detail the back-action effect on the oscillator caused by each coupling mode of the tweezers. The data shown in the figures include the sets collected at TUM, stressing the consistency of our findings across different laboratories.

The core oscillator performance is affected by the presence of a load. The higher the load concentration, the stronger the depletion of one (or more) of the core oscillator components. This causes an undesired retroactivity effect, namely a distortion of the oscillatory signal. In the following, we will provide some definitions and outline the load retroactivity effects we quantified.

The concentration of load that is to be driven by the oscillator is referred to as the *nominal load concentration*. For Modes I–V this is equivalent to the concentration of the tweezers added. For the rMG aptamer production mode, the nominal load is given by the MG switch concentration. The oscillation amplitude of the concentration of downstream tweezers (Modes I–V), is here called *effective load concentration*, and is calculated as twice the maximum amplitude per oscillation of the tweezers load.

The relative period change $\Delta T/T_0$ is calculated by defining ΔT as the difference between the

loaded sample period T and the reference period T_0. We analogously calculate the relative amplitude change $\Delta A/A_0$. The relative period and amplitude changes (in all data sets) are plotted in Figure 3.32 A and C as a function of the nominal load concentration. As a guide to the eye, we calculated least-square linear fits to each of the different modes of tweezers coupling and the rMG production.

The nominal load concentration affects the oscillator period most drastically for Modes I and III, while the amplitude is affected by all modes except mode II* and mode V. Indeed, Modes II* and V show the smallest effect on the oscillator period and amplitude, when the nominal load concentrations are considered.

To evaluate the performance of the different modes, the efficiency of the coupling has to be considered. From this point of view, Modes II*, III, and IV do not qualify as successful coupling modes, because the load oscillation amplitude (defined as in Section 3.7.7) is too small (Section 3.7.18). Modes I, II, V, and V* are actuated more strongly (Figures 3.33 and 3.34) with relative effective tweezers concentrations between roughly 10% and 60%.

The relative change in oscillation period and amplitude is plotted as a function of the effective tweezers load concentration in Figure 3.32 C and D.

The period is in general increased by the presence of a load, while we find different amplitude perturbation effects. For some of the modes (I, II, III, IV, and V) a comparison with the effects of threshold variations is drawn in Section 3.7.17.

Modes I and II have similar amplitude retroactivity effects. Mode II presents a smaller period retroactivity; however, the percent effective load driven drops as a function of the nominal load, as shown in detail in Figure 3.33. It is easy to observe that the maximum concentration of tweezers mode II that can be actuated should be well below [A1] = 250 nM, whereas for mode I this boundary is given by [dI1] = 700 nM.

For a system near our default operating point, a mean concentration of genelets in the "on" state of roughly 75 nM (30%) can be deduced for SW21, while this concentration is only around 20 nM (17%) for SW12 (see Section 3.7.6). SW12 is turned on only for a short time in each cycle, resulting in a much lower concentration [rA1] as compared to [rI2]. Driving the tweezers with rA1 in mode IV therefore affects the oscillator more strongly than driving with rI2 in mode III, as the resulting reduction in rA1 concentrations yields a larger fraction effect on switch activity. Similar reasoning explains why driving with A2 in mode II* has a negligible effect on the core oscillator even with a 400 nM load (Figure 3.32 and 3.43). However, the effective concentration of tweezers driven is practically zero. The high rI2 concentration and a toehold-mediated reaction path (Figure 3.16) allows the quick removal of A2 from tweezers; presuming closing of tweezers is slower than removal, the A2 concentration still provides the same effective threshold for SW12 inhibition. Presumably, the closing of tweezers by A2 is also slower than the hybridization of A2 to T12. We can conclude

that rI2 prevents direct coupling of TWII*, obviously decreasing the retroactivity.

The insulator of mode V minimally affects the core trajectories, analogously to mode II*. On the other hand, the RNA output InsOut amplifies minimal oscillations in the state of this load switch (analogous to what happens for SW12) and this mode achieves a good signal propagation on TWV.

The insulator designed for mode V* shows very low period retroactivity and has the best performance in terms of effective load driven (Figure 3.34 A). However, the amplitude retroactivity is significant. We can try to explain the properties of mode V* as follows: First, mode V* has the same input stage of SW21. This likely means that this load genelet is in an on state for a large fraction of time as SW21, maintaining a high concentration of InsOut (similarly to what observed for rI2). This explains why, given a certain effective load, a much smaller amount of insulator V* is required, compared to insulator V (mostly off as SW12). However, the output of mode V* in turn binds to the TwCls strand forming a substrate for RNase H, which is likely to be abundant most of the time following the reasoning done for rI2 (more abundant than in mode V). This hypothesis is consistent with the large plateaus visible in Figure 3.47. Through gel electrophoresis experiments (see Figure 3.35), we also found that the insulator of mode V* has a much larger off-state transcription rate than its A1 counterpart. Leakier transcription would also result in larger amounts of InsOut in solution, and more substrate for RNase H. In fact, significantly decreasing the amount of RNase H in solution results in slower reference oscillations with larger swing amplitude (Section 3.7.16). It is thus plausible that the significant amplitude retroactivity of mode V* is caused by the presence of larger amounts of substrate for RNase H. Note that the rMG switch has the same input stage of insulator mode V* (SW21), though its retroactivity effects are different: in fact, the aptamer output does not bind to any DNA target and does not create additional substrates for RNase H.

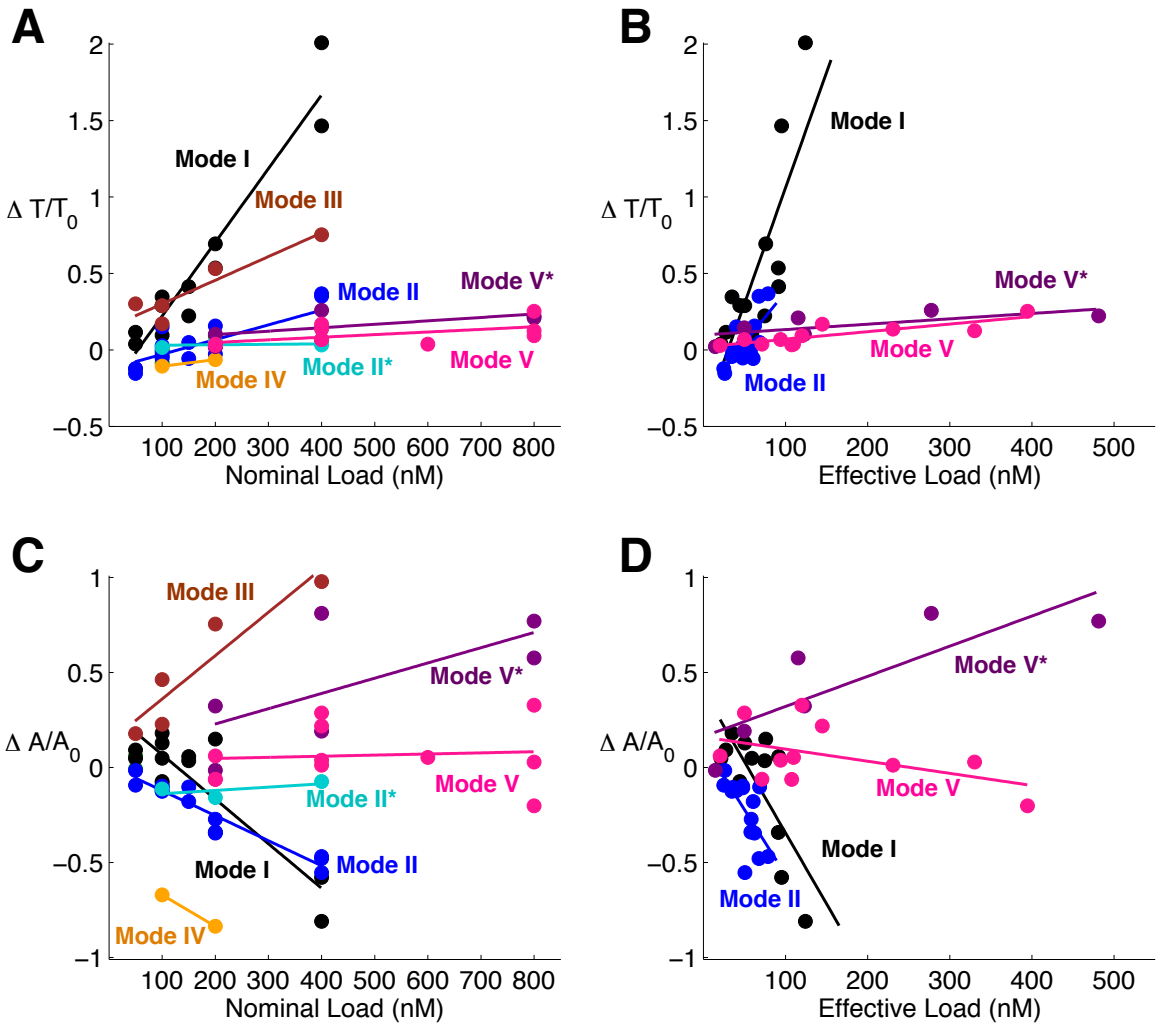


Figure 3.32: Relative period and amplitude change as a function of the load concentration for all data sets collected at Caltech and TUM. The data points are shown only when the oscillator traces exhibit a detectable amplitude and period. A. Nominal tweezer load versus core oscillator period variation. B. Effective tweezer load versus core oscillator period variation. C. Nominal tweezer load versus core oscillator amplitude variation. D. Effective tweezer load versus core oscillator amplitude variation.

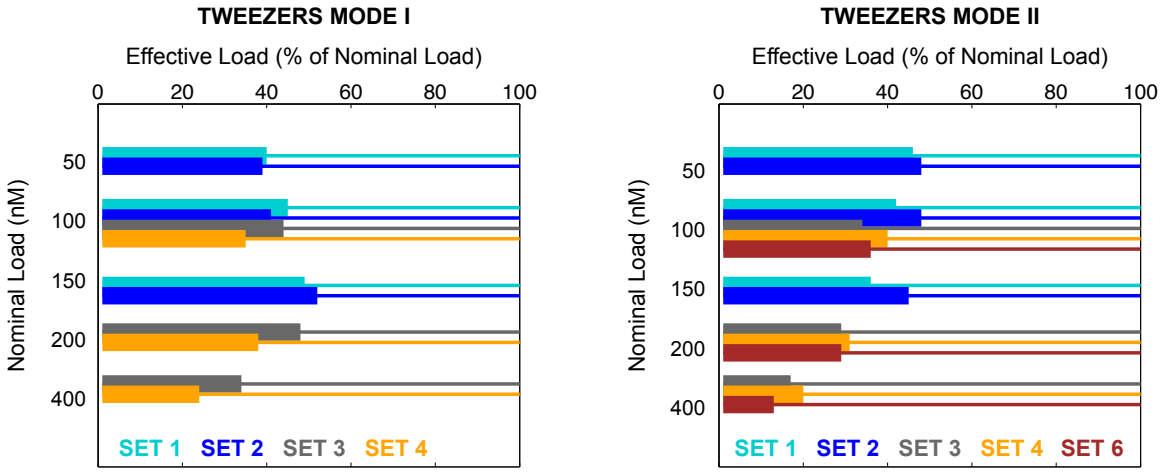


Figure 3.33: Effective load for mode I and mode II across different data sets collected at Caltech and TUM.

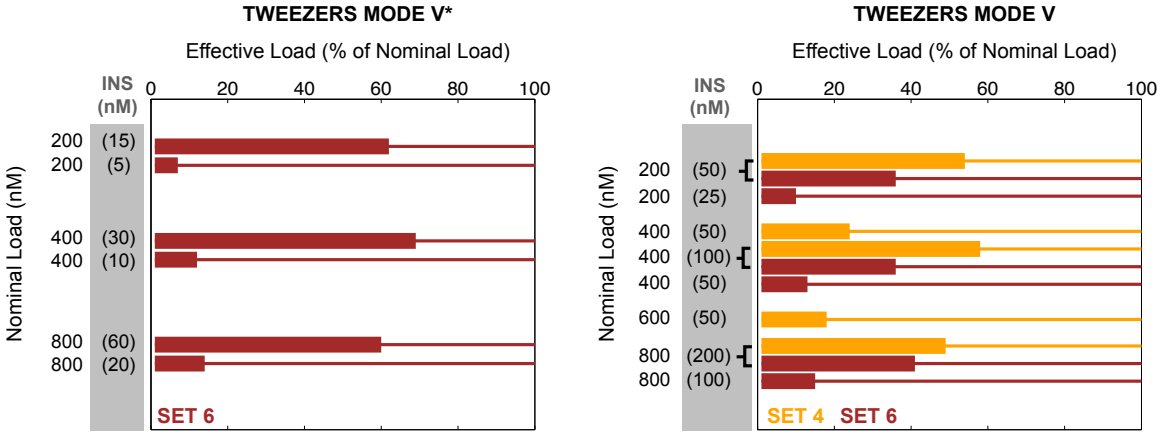


Figure 3.34: Effective load for mode V* and mode V across different data sets collected at Caltech.

3.7.13 Leak transcription from off-state switches

The synthetic genelets are switched on and off by displacing part of their nicked promoter region. The switches in the on state (T12·A2, T21·A1) have a fully double-stranded, yet nicked promoter region, whereas in the off state (T12, T21) the promoter region of the switches is partly single-stranded. We tested the off state transcription on all of our switches: T12·A2, T21·A1, Ins·A2, and Ins*·A1.

Figure 3.35 shows the gel electrophoresis results of samples from transcription reactions, for the core oscillator switches and the insulator switches. All of the switches were separately annealed and mixed with 1x Ambion Megashortscript kit reagents and 5% (v/v) Ambion T7 enzyme mix, in the presence or absence of their respective activator strand. Final concentrations of all the annealed switches were 200 nM, and the activators were added in excess, to a final concentration of 350 nM. The reaction was incubated at 37°C for 4 hours. RNA yield was quantified with 10% denaturing PAGE, run at 21°C. Template T21 has a higher off-state transcription rate than T12. The mode V* insulator, with input domain identical to T21, also exhibits higher off-state transcription than the mode V insulator.

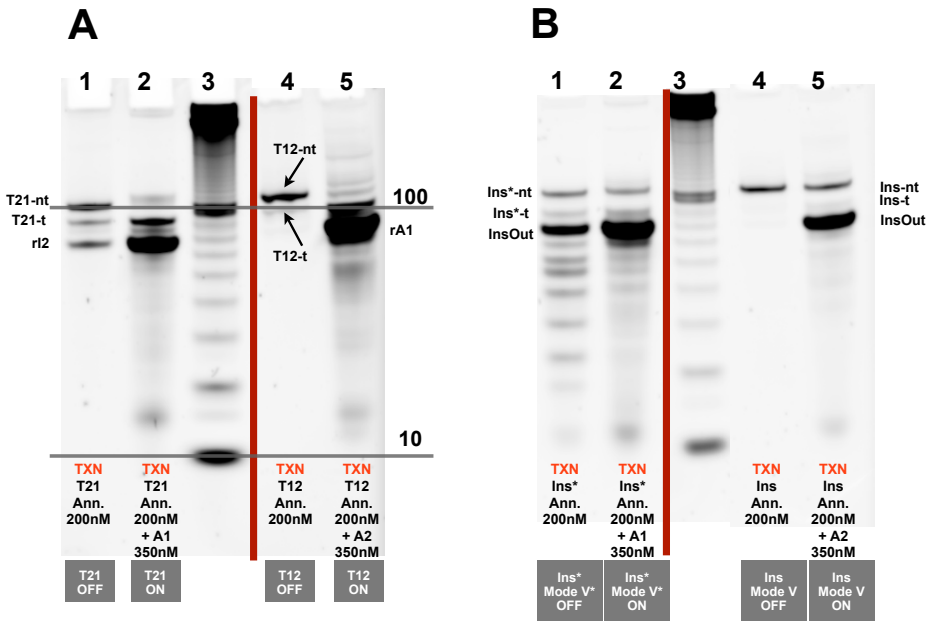


Figure 3.35: Off state versus on state transcription: A. Core oscillator switches. B. Mode V and mode V* insulator switches. Staining of the template and non-template strands is inconsistent. The red line indicates that the gel image has been processed by cropping some irrelevant bands.

3.7.14 Lack of transcription from T21·rA1 complex

We investigated the extent of transcription from the T21·rA1 complex (see Figure 3.14). Annealed T21-t and T21-nt templates at 250 nM were added with variable amounts of strands rA1. Transcription was performed in Ambion T7 Megashortscript kit 1x reagents, incubated at 37°C for 4 hours. Yield was quantified with 10% denaturing PAGE, run at 21°C. Additionally, following the same protocol, transcription of A1-activated template T21 was tested, showing strong transcriptional activity. The results show that the unwanted interaction between rA1 and SW21 produces negligible amounts of rI2 transcript, compared to the correct activation pathway (compare lanes 5 and 6 in Figure 3.36 A to lane 5 in Figure 3.36 B).

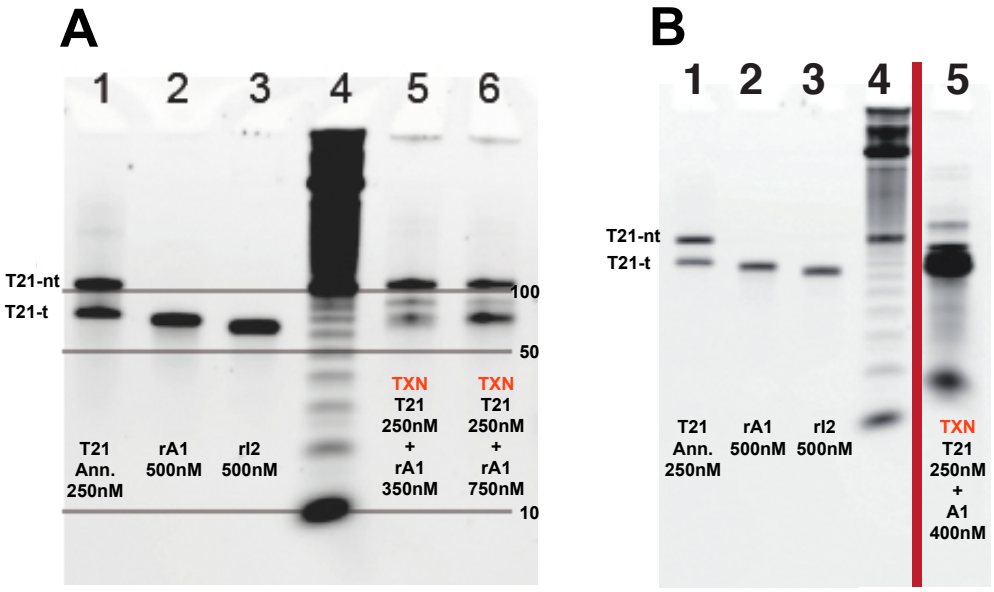


Figure 3.36:
A. Transcription from T21·rA1 complex. Lane 1: Control sample, annealed T21-t and T21-nt at 250 nM, in 1x Ambion T7 Megashortscript transcription buffer. Lane 2: Control sample rA1 at 500 nM. Lane 3: Control sample rI2 at 500 nM. Lane 4: 10 bp DNA ladder. Lane 5: Transcription of annealed T21 at 250 nM in solution with rA1 at 350 nM. Lane 6: Transcription of annealed T21 at 250 nM in solution with rA1 at 750 nM. B. Transcription from T21·A1 complex. Lane 1: Control sample, annealed T21-t and T21-nt at 250 nM, in 1x Ambion T7 Megashortscript transcription buffer. Lane 2: Control sample rA1 at 500 nM. Lane 3: Control sample rI2 at 500 nM. Lane 4: 10 bp DNA ladder. Lane 5: Transcription of annealed T21 at 250 nM in solution with A1 at 400 nM. The red line indicates that the gel image has been processed by cropping some non relevant bands between lane 4 and lane 5.

3.7.15 Interactions between enzymes and tweezers

Molecular tweezers can be a substrate for RNAP. This was observed through gel electrophoresis and fluorescence measurements on tweezer Modes I–IV. Figure 3.37 shows the results on tweezers mode II. These were tested as a transcription substrate by adding them at a concentration of 200 nM to Ambion T7 Megashortscript kit 1x reagents. A denaturing 10% PAGE run at 21°C shows unknown transcription products of different lengths. Some of these products are degraded by RNase H, as can be seen from Figure 3.37. This suggests that such products bind to the DNA tweezer strands. In fact, the fluorescence time traces in Figure 3.38 show that transcription products interact through an unknown mechanism with the tweezers, causing an increase in fluorescence. The presence of RNase H again reduces this phenomenon by degrading tweezer-bound RNA.

The extent of transcription from a tweezer substrate is sequence specific. Some of the tweezer designs we tested exhibited a drift in fluorescence over time of up to 30% even in the presence of RNase H, and had to be discarded. The presence of T7 promoter sub-sequences in such tweezer designs might have favored the interaction with RNAP.

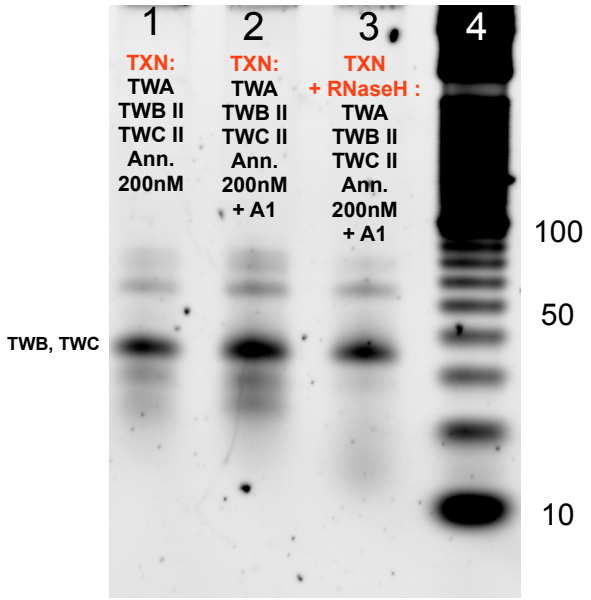


Figure 3.37: Denaturating 10% PAGE. Lane 1: Transcription run on TW II A, B, and C all at 200 nM. Lane 2: Transcription run on TW II A, B, C and A1 all at 200 nM. Note that TW II A is not visible due to the quencher present on the strand. Lane 3: Reaction products of Lane 2 after 30 min incubation with 2% (v/v) Ambion RNase H. Unknown transcription products of different lengths appear in all lanes; some of the products that most likely bind to the DNA tweezers are degraded by RNase H and disappear in Lane 3. Lane 4: 10 bp DNA ladder.

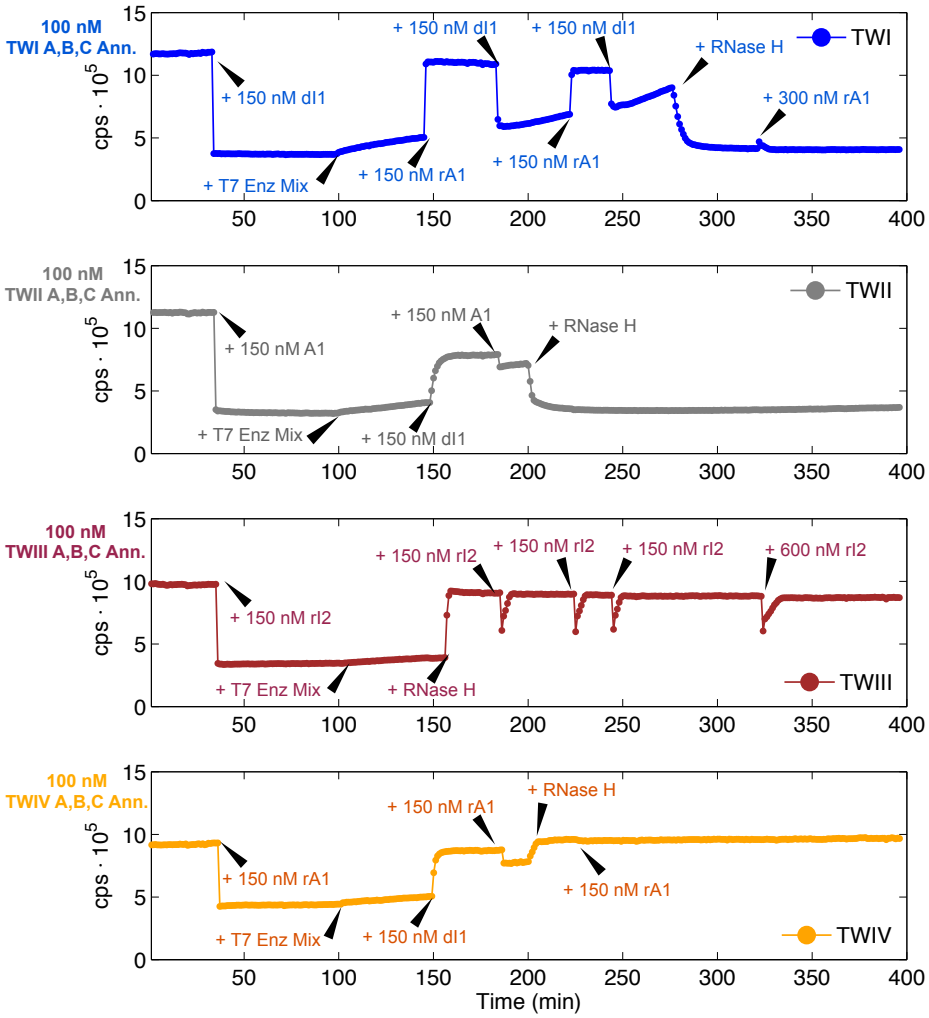


Figure 3.38: Raw fluorescence time traces showing opening and closing cycles for TW I, II, III, and IV. Initially, 100 nM annealed tweezers are present in solution with 1x Ambion T7 Megashortscript kit buffer and 7.5 mM each rNTP. Nucleic acid strands are added at different times according to the indicated labels. Enzymes are added as follows: Ambion T7 Enzyme Mix is added in 3 μL aliquots; RNase H is added in 0.2 μL aliquots (2 units from a stock of Ambion RNase H 10 U/μL). The final volume of each sample is 60 μL.

3.7.16 Effects of changing enzyme volume ratio

In this section, we briefly discuss the dependence of frequency and amplitude of the oscillator on the volume ratio of RNAP and RNase H.

As mentioned before, we tuned the enzyme amounts in the CIT and TUM protocols in order to achieve a similar operating point, defined as 4–6 full oscillations having amplitude around 100 nM. The tuning operation involved mainly an exploration of the system behavior as a function of the relative enzyme concentrations. As shown in Figure 3.39, increasing the volume of RNase H produces faster oscillations and lower amplitude.

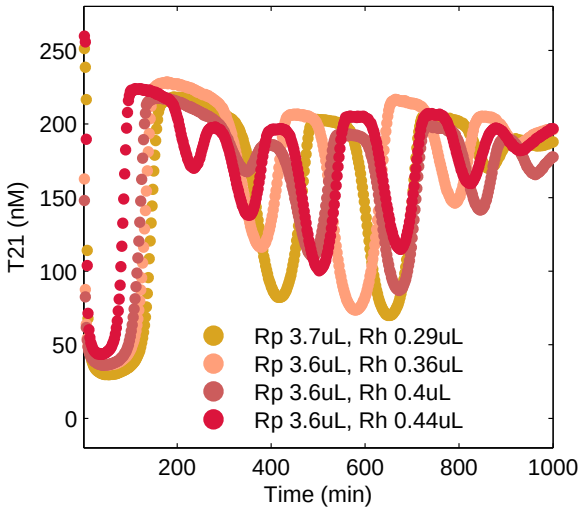


Figure 3.39: SET 6, with $[\mathrm{dI1}]_{\mathrm{tot}}$=700 nM. Data taken on different days

3.7.17 Effects of changing the DNA thresholds

The influence of different coupling schemes on the dynamics of the transcriptional oscillator can be better understood by considering the effect of the tweezers in the context of the full set of chemical reactions occurring in the system. Due to their interaction with different molecular species in the core oscillator system, the tweezers effectively change the concentrations of the threshold strands A2 and dI1. In mode I, tweezers are closed by dI1 and opened in a strand displacement reaction by rA1. In the core oscillator, strand A1 is similarly bound by dI1 and freed by a strand displacement reaction with rA1. We reasoned that an increase in the concentration of mode I tweezers therefore roughly mimics an increase in A1, which in turn corresponds to an effective reduction of the threshold set by [dI1]. In contrast, increasing the concentrations of tweezers in mode II or mode IV resembles an increase in [dI1]. In mode II, a fraction of activator strands A1 is bound to the tweezers rather than to template T21. Dynamically, the effective reduction in [A1] should be analogous to an increase in threshold by [dI1]. In mode IV, a fraction of the rA1 transcripts is sequestered by the tweezers, effectively shifting the threshold for activation of SW21 to higher values. Finally, an increase in tweezers concentration in mode III should be similar in effect to an increase in A2 concentration, as a fraction of the rI2 transcripts is bound by the tweezers, while an increase in mode II* tweezers, which are closed by A2, should correspond to a decreased A2 threshold. We experimentally challenged this interpretation of loads as effective changes in threshold values. Figure 3.40 shows fluorescence traces recorded from the unloaded oscillator, for which the concentrations of the threshold strands dI1 and A2 were systematically varied. The general trends in amplitude and period closely resemble the trends observed in Figures 3.41, 3.42, 3.43, 3.44, and 3.45 for Modes I, II, II*, III, and IV, agreeing well with the heuristic expectations detailed above. An exception is the slowing down of the oscillations with increasing tweezers concentration operated in mode II, perhaps due to the tweezers' direct effect being on A1 rather than dI1.

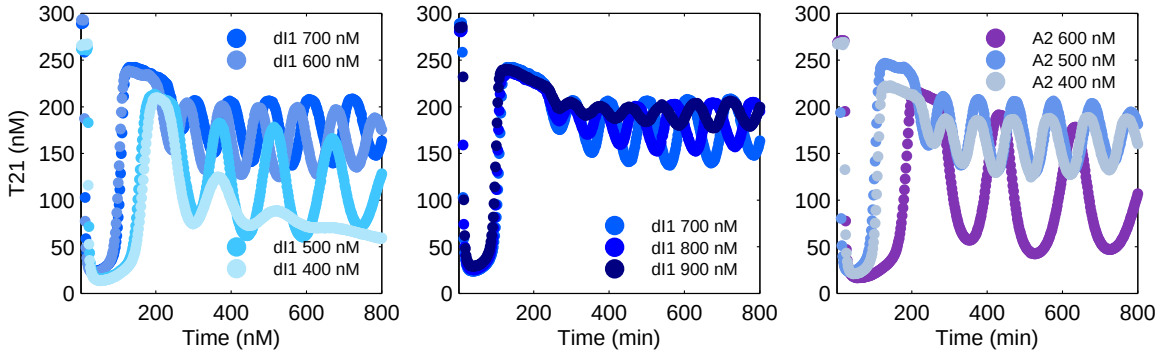


Figure 3.40: Threshold variation, set 5 samples. Oscillator traces with varying threshold concentrations. Left: the concentration of dI1 is reduced from 700 nM to 400 nM. Center: The concentration of dI1 is increased from 700 nM to 900 nM. Right: The concentration of A2 was varied from 400 nM over 500 nM (default concentration) to 600 nM.

3.7.18 Overview of all fluorescence data sets collected at Caltech

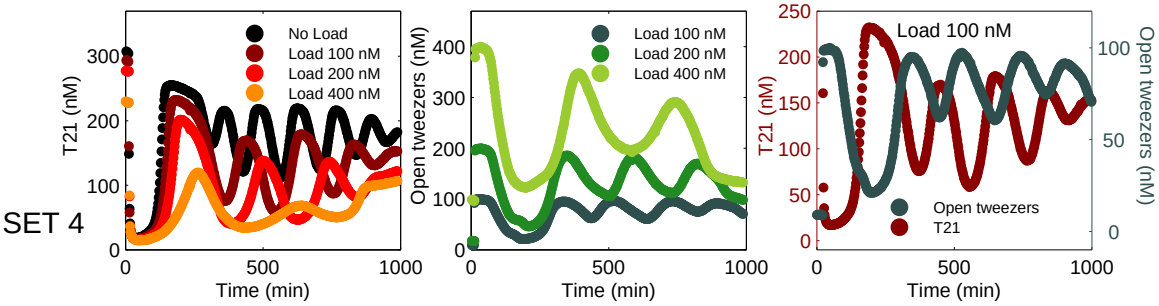


Figure 3.41: Mode I. Left: Oscillator traces. Center: Load traces. Right: Oscillator and load traces.

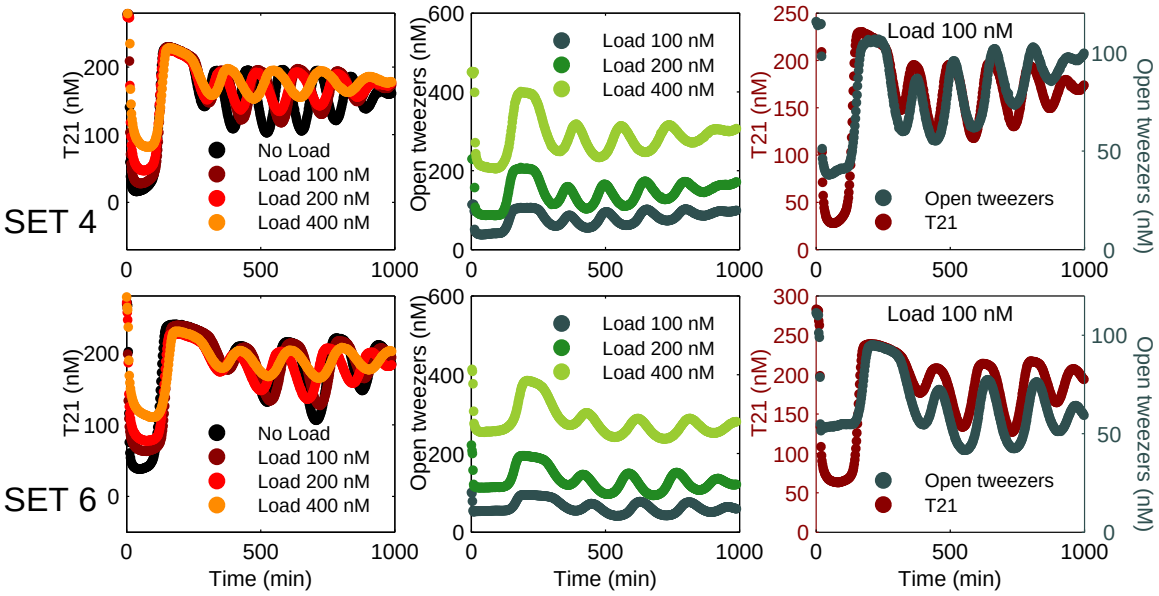


Figure 3.42: MODE II. Left: Oscillator traces. Center: Load traces. Right: Oscillator and load traces.

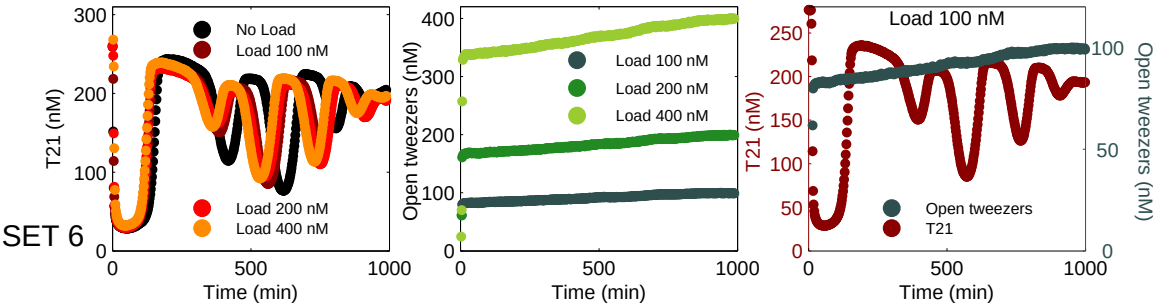


Figure 3.43: Mode II*. Left: Oscillator traces. Center: Load traces. Right: Oscillator and load traces.

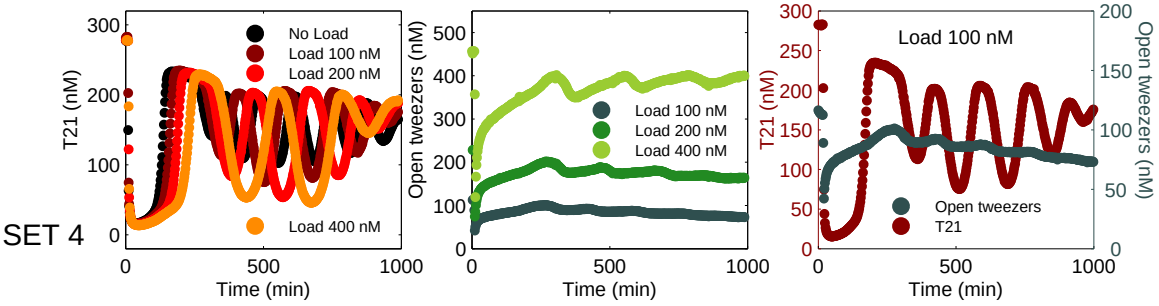


Figure 3.44: Mode III. Left: Oscillator traces. Center: Load traces. Right: Oscillator and load traces.

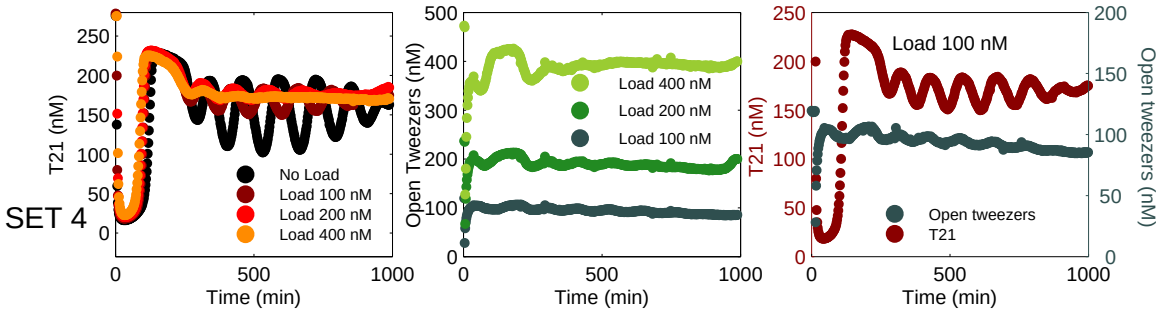


Figure 3.45: Mode IV. Left: Oscillator traces. Center: Load traces. Right: Oscillator and corresponding load traces.

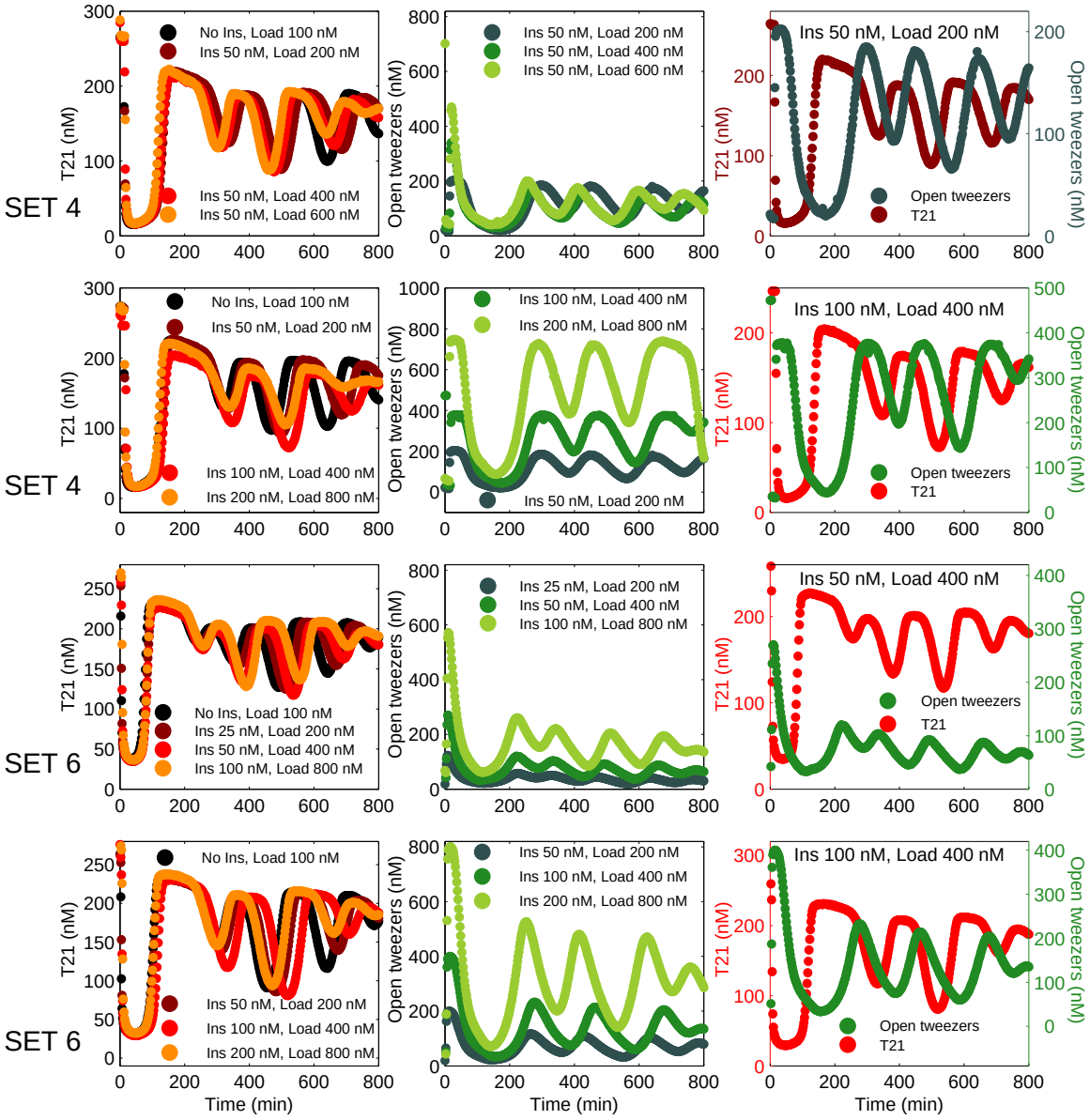


Figure 3.46: Mode V. Left: Oscillator traces. Center: Load traces. Right: Oscillator and corresponding load traces. The TwCls DNA strand was always added in 50 nM excess of TWV load.

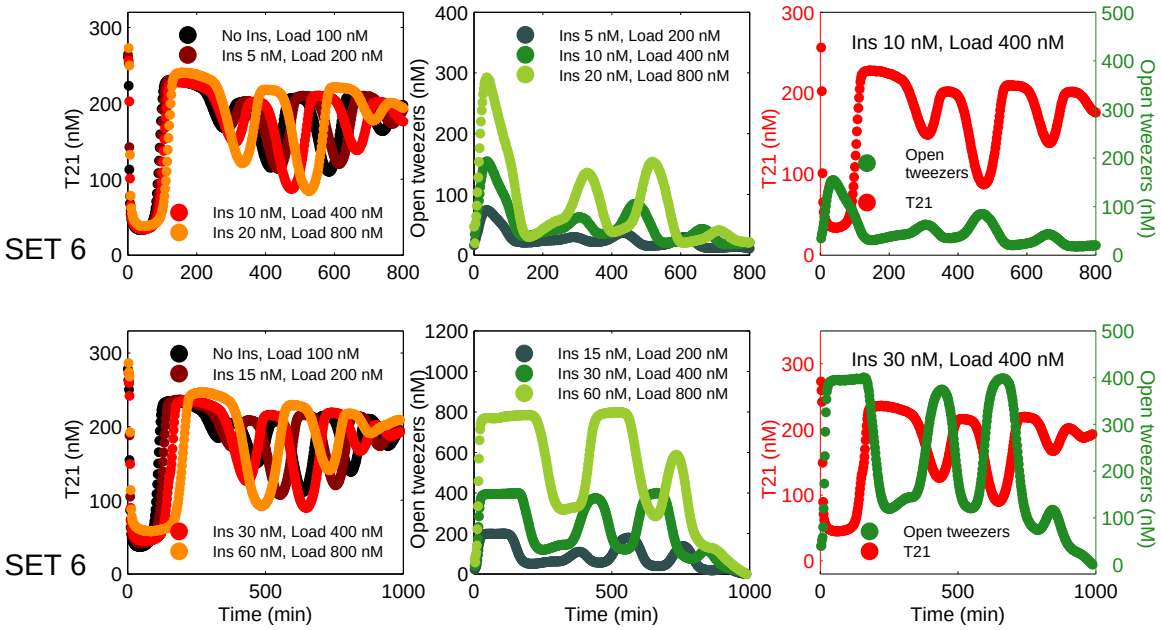


Figure 3.47: Mode V*. Left: Oscillator traces. Center: Load traces. Right: Oscillator and corresponding load traces. The DNA strand TwCls (closing the mode V tweezers) was always added in 50 nM excess of TWV load.

Chapter 4

Structural robustness in molecular networks

4.1 Introduction

In this chapter, a simple question will be asked: are there dynamical models for biological systems that have structurally stable equilibria and preserve this property robustly with respect to their specific parameters? This question has been considered before in the literature. For instance, in [94], through numerical exploration of the Jacobian eigenvalues for two-, three- and four-node networks, the authors isolated a series of interconnections which are stable, robustly with respect to the specific parameters. Such structures also turned out to be the most frequent topologies in existing biological networks databases. In another recent work [75], through extensive numerical analysis on three-node networks, the authors have shown that adaptability (defined as a significant step response followed by relaxation to the pre-stimulus equilibrium) of these systems can be investigated solely based on their structure, regardless of the chosen reaction parameters. Numerical simulation has arguably been the most popular tool to investigate robustness of biological networks [68, 45, 46, 58, 5, 127].

Analytical approaches to the study of robustness have been used in specific contexts. A series of recent papers [116, 115] focused on input/output robustness of ODE models for phosphorylation cascades. In particular, the theory of chemical reaction networks is used in [115] as a powerful tool to demonstrate the property of absolute concentration robustness. Indeed, the so-called deficiency theorems [34] are to date some of the most general results to establish robust stability of a (bio)chemical reaction network. Monotonicity is also a structural property that is useful in demonstrating robust dynamic behaviors of a class of biological models [121, 10]. Robustness has also been investigated in the context of compartmental models, which are often encountered in biology and chemistry [54].

Here, I will follow the paper I co-authored with Prof. Franco Blanchini [19] and present a simple and general theoretical tool kit for the analysis of bio-molecular systems. Such tool kit is constituted by Lyapunov and set-invariance methods. Provided that certain standard properties are verified, a

number of well-known biological networks are demonstrated to be asymptotically stable, robustly with respect to the model parameters. In some cases, robust bounds on the system performance are found. This approach does not require numerical simulation efforts.

The framework suggested here aims at formulating qualitative models without the need of exact mathematical expressions and parameters. The utilized analytical methods rely only on qualitative interactions between network components. The properties that can be derived from the models we formulate are, consequently, structurally robust because they are not inferred from specific mathematical formulas chosen to fit data. The techniques suggested are based on set-invariance and Lyapunov theory, in particular piecewise-linear functions, and show that such models are amenable for robust investigation by *engineers and mathematicians*. These techniques are effective and promising in dealing with biological robustness [3, 32]. Several models from the literature are considered, reporting the original equations, and rephrasing them in our setup as case studies. Robust certifications can be given to important properties (some of which have been established based on specific models).

4.2 Methods

Robustness

We will consider biological dynamical systems which are successfully modeled with ODEs and can be written as:

$$\dot{\mathrm{x}} = f(\mathrm{x}, \mathrm{u}), \qquad \mathrm{x}(0) = \mathrm{x}_0, \tag{4.1}$$

where x is the system state, u models external inputs, and both are vectors of appropriate dimensions. Such class of models is appropriate for biological systems where stochasticity and anisotropy can be neglected. We define robustness as follows:

Definition 1 Let $\mathcal{C}$ be a class of systems and $\mathcal{P}$ be a property pertaining to such a class. Given a family $\mathcal{F} \subset \mathcal{C}$ we say that $\mathcal{P}$ is *robustly verified* by $\mathcal{F}$, in short, robust, if it is satisfied by each element of $\mathcal{F}$.

Countless examples can be brought regarding families $\mathcal{F}$ and candidate properties. In this work, we will focus on the property of stability, which is an important feature for the equilibria of biological networks [66, 10, 94].

If we take into account a linear or linearized dynamical system, we can immediately provide an example that clarifies our definition of robustness [98]. Let $\mathcal{C}$ be the class of linear differential

systems and $\mathcal{F}$ the family of second order systems described by

$$\begin{bmatrix} \dot{\mathrm{x}}_1(\mathrm{t}) \\ \dot{\mathrm{x}}_2(\mathrm{t}) \end{bmatrix} = \begin{bmatrix} -a & b \\ -c & -d \end{bmatrix} \begin{bmatrix} \mathrm{x}_1(\mathrm{t}) \\ \mathrm{x}_2(\mathrm{t}) \end{bmatrix},$$

with positive and constant coefficients a, b, c, d. Assume $\mathcal{P}$ = *asymptotic stability of the origin as an equilibrium point.* Then we can say that $\mathcal{P}$ is robust. The situation is different if we admit that $a(\mathrm{t}), b(\mathrm{t}), c(\mathrm{t}), d(\mathrm{t})$ can vary with time, yielding a system which is possibly unstable.

If one is interested in the global system behavior, Lyapunov functions are a powerful tool providing sufficient conditions for stability. Given an equilibrium point $\bar{\mathrm{x}}$, any convex function $\mathrm{V}(\mathrm{x}-\bar{\mathrm{x}}) > 0$ for $\mathrm{x} \neq \bar{\mathrm{x}}$ and zero at the origin is a candidate Lyapunov function. If $f(\mathrm{x}, \mathrm{u})$ is continuous, and $V(\cdot)$ is smooth, then $\mathrm{V}(\cdot)$ is a Lyapunov function if

$$\dot{\mathrm{V}}(\mathrm{x} - \bar{\mathrm{x}}) = \nabla \mathrm{V}(\mathrm{x} - \bar{\mathrm{x}}) f(\mathrm{x}, \bar{\mathrm{u}}) \leq \kappa(\mathrm{x} - \bar{\mathrm{x}}),$$

where $\bar{u}$ is fixed and $\kappa(\cdot)$ is a negative definite function (i.e. $\kappa(\cdot) < 0$ on all its domain, except for $\kappa(0) = 0$).

Non-smooth Lyapunov functions

The concept of Lyapunov derivative can be generalized when the function $\mathrm{V}(\cdot)$ is non-smooth. For instance, consider the convex function:

$$\mathrm{V}(\mathrm{x} - \bar{\mathrm{x}}) = \max_{\mathrm{i}} \mathrm{V}_\mathrm{i}(\mathrm{x} - \bar{\mathrm{x}}),\ \mathrm{i} = 1, ..., \mathrm{N},$$

where each $\mathrm{V}_\mathrm{i}(\cdot)$ is smooth and convex, and assume that $\mathrm{V}(\cdot)$ is positive definite. The set of active functions is never empty and is defined as: $\mathcal{A} = \{\mathrm{i} : \mathrm{V}_\mathrm{i}(\mathrm{x} - \bar{\mathrm{x}}) = \mathrm{V}\}$. If we define the generalized Lyapunov derivative as:

$$\mathrm{D}^+\mathrm{V}(\mathrm{x} - \bar{\mathrm{x}}) \triangleq \max_{\mathrm{i} \in \mathcal{A}} \nabla \mathrm{V}_\mathrm{i}(\mathrm{x} - \bar{\mathrm{x}}) \mathrm{f}(\mathrm{x}),$$

then the condition for stability becomes:

$$\mathrm{D}^+\mathrm{V}(\mathrm{x} - \bar{\mathrm{x}}) < \kappa(\mathrm{x} - \bar{\mathrm{x}}), \quad \kappa(\cdot) \text{ negative definite.}$$

Positively invariant sets

We are interested in cases where the trajectories of system (4.1) remain trapped in bounded sets at all times, therefore behaving consistently with respect to some desired criterion.

We say that a subset $\mathcal{S}$ of the state space is positively invariant if $\mathrm{x}(0) \in \mathcal{S}$ implies that also

$\mathrm{x}(\mathrm{t}) \in \mathcal{S}$ for all $\mathrm{t} > 0$. The following theorem (which relies on the concept of Lyapunov function) provides a general necessary and sufficient condition for a set to be invariant.

Theorem 1 *(Nagumo, 1943) Assume that system* (4.1) *(for a fixed constant input* $\bar{u}$*) admits a unique solution. Consider the set:*

$$\mathcal{S} \triangleq \{\mathrm{x} \in \mathbb{R}^n : \quad s_i(\mathrm{x}) \leq \sigma_{\mathrm{i}},\, \mathrm{i} = 1, ..., \mathrm{r}\},$$

where s_i *are smooth functions, and* σ_i *are given constants. Assume that* $\nabla s_i \neq 0$, $\forall \mathrm{x} \in \partial\mathcal{S}$*. The set of active constraints is* $\mathcal{I}(\mathrm{x}) \triangleq \{i : s_i(\mathrm{x}) = \sigma_i\}$*, and is non-empty only on the boundary of* $\mathcal{S}$*. Then the set* $\mathcal{S}$ *is positively invariant if and only if*

$$\nabla s_i(\mathrm{x}) f(\mathrm{x}, \bar{\mathrm{u}}) \leq 0,\, \forall \mathrm{x} \in \partial\mathcal{S},\, \textit{and}\, \; i \in \mathcal{I}(\mathrm{x}).$$

For instance, if our constraining functions are linear, $s^\top \mathrm{x} \leq \sigma$, the Nagumo conditions are $s^\top f(\mathrm{x}, \bar{\mathrm{u}}) \leq 0$. We refer the reader to [103] for further details on positively invariant sets; more recent works on this topic are [18] and [20].

Structural robustness investigation for biological networks

Let us begin with a simple biological example. Consider a protein x_1, that represses the production of an RNA species x_2. In turn, x_2 can be the target of another RNA species u_2 (and form an inactive complex to be degraded) or it can be translated into protein x_3. A standard dynamical model [26] of this process is:

$$\begin{aligned} \dot{\mathrm{x}}_1 &= \mathrm{u}_1 - b_{11}\mathrm{x}_1, \\ \dot{\mathrm{x}}_2 &= d_{21}(\mathrm{x}_1) - b_{22}\mathrm{x}_2 - b_{2\mathrm{u}_2}\, \mathrm{x}_2\, \mathrm{u}_2, \quad d_{21}(\mathrm{x}_1) = \frac{1}{1+\mathrm{x}_1^n}, \\ \dot{\mathrm{x}}_3 &= a_{32}\mathrm{x}_2 - b_{33}\mathrm{x}_3. \end{aligned} \tag{4.2}$$

RNA species x_2 determines the production rate of protein x_3 by indexing the corresponding reaction rate as a_{32}. Following the standard notation in control theory, we assume that the production rate of protein x_1 is driven by some external signal or input u_1, and that RNA u_2 also acts as an external input on RNA x_2. We assume that all the system parameters are positive and bounded scalars. Terms a_{ij} are first-order production rates: species i is produced at a rate which is linear in species j; b_{ih} denotes, in this case, first-order degradation rates. The term $d_{21}(\mathrm{x}_1)$ is a well-known Hill function term [7].

The stability properties of this small network can be immediately assessed: x_1 will converge to

its equilibrium $\bar{\mathrm{x}}_1 = \mathrm{u}_1/b_{11}$. Similarly, $\bar{\mathrm{x}}_2 = d_{21}(\bar{\mathrm{x}}_1)/(b_{22} + b_{2\mathrm{u}_2}\mathrm{u}_2)$, $\bar{\mathrm{x}}_3 = a_{32}\bar{\mathrm{x}}_2/b_{33}$. Regardless of the specific parameter values, and therefore robustly, the system is stable. The equilibrium $\bar{\mathrm{x}}_1$ could grow unbounded with u_1, however $\bar{\mathrm{x}}_2$ is always bounded.

It should be remarked that *the knowledge of functions $a_{ij}\mathrm{x}$, $b_{ih}\mathrm{x}$, and $d(\cdot)$ is not necessary at all*: the previous conclusions can be easily derived by the qualitative information that d_{21} is strictly decreasing and asymptotically converging to 0, while $b_{11}\mathrm{x}_1$, $b_{22}\mathrm{x}_2$, $b_{2\mathrm{u}_2}\mathrm{x}_2$ u_2, $a_{32}\mathrm{x}_2$, and $b_{33}\mathrm{x}_3$ are increasing.

It is appropriate at this point to outline a series of general assumptions that will be useful in the following analysis.

We will consider a class of biological network models consisting of n first-order differential equations

$$\dot{\mathrm{x}}_i(\mathrm{t}) = \sum_{j\in\mathcal{A}_i} f_{ij}(\mathrm{x}_i, \mathrm{x}_j) - \sum_{h\in\mathcal{B}_i} g_{ih}(\mathrm{x}_i, \mathrm{x}_h) + \sum_{s\in\mathcal{C}_i} c_{is}(\mathrm{x}_s) + \sum_{l\in\mathcal{D}_i} d_{il}(\mathrm{x}_l), \tag{4.3}$$

where x_i, $i = 1, ..., n$ are the dynamic variables. For the sake of notation simplicity, external inputs are not denoted with a different symbol. Inputs can be easily included as dynamic variables $\dot{\mathrm{x}}_u = w_u(\mathrm{x}_u, t)$ which are not affected by other states and have the desired dynamics. The sets $\mathcal{A}_i$, $\mathcal{B}_i$, $\mathcal{C}_i$, $\mathcal{D}_i$ denote the subsets of variables affecting x_i. The different terms in equation (4.3) are associated with a specific biological and physical meaning. The terms $f_{ij}(\cdot,\cdot)$ represent production rates of reagents; typically, these functions are assumed polynomial in their arguments; similarly, terms $g_{ih}(\cdot,\cdot)$ model degradation or conversion rates and are also likely to be polynomial in practical cases. Finally, terms $c(\cdot)$ and $d(\cdot)$ are associated with monotonic nonlinear terms, often given by Michaelis-Menten or Hill functions [7].

We assume that system (4.3) satisfies the following assumptions:

A1 *(Smoothness) Functions $f_{ij}(\cdot,\cdot)$, $g_{ih}(\cdot,\cdot)$, $c_{is}(\cdot)$, and $d_{il}(\cdot)$ are unknown, nonnegative, continuously differentiable functions.*

A2 *$f_{ij}(\mathrm{x}_i, 0) = 0$ and $g_{ih}(\mathrm{x}_i, 0) = 0$, $\forall x$.*

A3 *Functions $f_{ij}(\mathrm{x}_i, \mathrm{x}_j)$ and $g_{ih}(\mathrm{x}_i, \mathrm{x}_h)$ are strictly increasing in x_j and x_h, respectively.*

$$\frac{\partial f_{ij}(\mathrm{x}_i, \mathrm{x}_j)}{\partial \mathrm{x}_j} > 0, \qquad \frac{\partial g_{ih}(\mathrm{x}_i, \mathrm{x}_h)}{\partial \mathrm{x}_h} > 0, \qquad \forall x$$

A4 *(Saturation) Functions $c_{is}(\mathrm{x}_s)$ and $d_{il}(\mathrm{x}_l)$ are nonnegative and, respectively, non-decreasing and non-increasing. Moreover $c_{is}(\infty) > 0$ and $d_{il}(0) > 0$.*

A5 *Functions $g_{ih}(\cdot,\cdot)$ are null at the lower saturation levels : $g_{ih}(0, \mathrm{x}_h) = 0$, $\forall \mathrm{x}_h$.*

In view of the nonnegativity assumptions and Assumption A5, the general model (4.4) is a nonlinear positive system, according to the next proposition, and its investigation will be restricted to the positive orthant.

Proposition 1 *The nonnegative orthant* $\mathrm{x}_i \geq 0$ *is positively invariant for system* (4.4).

Given the above assumptions, we can write equation (4.3) in an equivalent form. First of all, in view of A1–A3, we can write: $f_{ij}(\mathrm{x}_i, \mathrm{x}_j) = a(\mathrm{x}_i, \mathrm{x}_j)\mathrm{x}_j$, $g_{ih}(\mathrm{x}_i, \mathrm{x}_h) = b(\mathrm{x}_i, \mathrm{x}_h)\mathrm{x}_h$, with

$$a_{ij}(\mathrm{x}_i, \mathrm{x}_j) = \frac{f_{ij}(\mathrm{x}_i, \mathrm{x}_j)}{\mathrm{x}_j} \quad \text{and} \quad b_{ij}(\mathrm{x}_i, \mathrm{x}_h) = \frac{g_{ih}(\mathrm{x}_i, \mathrm{x}_h)}{\mathrm{x}_h}.$$

The above expression is always valid due to the smoothness assumption A1 (see [54], Section 2.1). Additionally, Assumption A5 requires that $b_{ih}(0, \mathrm{x}_h) = 0$, $\forall \mathrm{x}_h$, for $i \neq h$. Once we adopt this notation, we can rewrite model (4.3) as follows:

$$\dot{\mathrm{x}}_i(\mathrm{t}) = \sum_{j \in \mathcal{A}_i} a_{ij}(\mathrm{x}_i, \mathrm{x}_j)\mathrm{x}_j - \sum_{h \in \mathcal{B}_i} b_{ih}(\mathrm{x}_i, \mathrm{x}_h)\mathrm{x}_h + \sum_{s \in \mathcal{C}_i} c_{is}(\mathrm{x}_s) + \sum_{l \in \mathcal{D}_i} d_{il}(\mathrm{x}_l),\ i = 1, 2, \ldots, n. \tag{4.4}$$

To simplify the notation, we have considered functions depending on two variables at most. However, we can straightforwardly extend assumptions A1–A5 to multivariate functional terms in equation (4.3). In turn, the model structure (4.4) can be easily generalized to include terms as $a(\mathrm{x}_i, \mathrm{x}_j, \mathrm{x}_k, \ldots)$, $b(\mathrm{x}_i, \mathrm{x}_j, \mathrm{x}_k, \ldots)$, $c(\mathrm{x}_i, \mathrm{x}_j, \mathrm{x}_k, \ldots)$, $d(\mathrm{x}_i, \mathrm{x}_j, \mathrm{x}_k, \ldots)$.

If we restrict our attention to the general class of models (4.4), under assumptions A1–A5, we can proceed to successfully analyzing the robust stability properties of several biological network examples.

The structural analysis of system (4.4) can be greatly facilitated whenever it is legitimate to assume that functions a, b, c, and d have certain properties. For the reader's convenience, a list of possible properties is given below. Given a general function $f(\mathrm{x})$:

P1 $f(\mathrm{x}) = const \geq 0$ *is* nonnegative-constant.

P2 $f(\mathrm{x}) = const > 0$ *is* positive-constant.

P3 $f(\mathrm{x})$ *is* sigmoidal: *it is non-decreasing;* $f(0) = f'(0) = 0$, *if* $0 < f(\infty) < \infty$ *and its derivative has a unique maximum point,* $f'(\mathrm{x}) \leq f'(\bar{\mathrm{x}})$ *for some* $\bar{x} > 0$.

P4 $f(\mathrm{x})$ *is* complementary sigmoidal: *it is non-increasing,* $0 < f(0)$, $f'(0) = 0$, $f(\infty) = 0$, *and its derivative has a unique minimum point. In simple words,* f *is a CSM function iff* $f(0) - f(\mathrm{x})$ *is a sigmoidal function.*

P5 $f(\mathrm{x})$ *is* constant-sigmoidal*, the sum of a sigmoid and a positive constant.*

P6 $f(\mathrm{x})$ *is* constant-complementary-sigmoidal*, the sum of a complementary sigmoid and a constant.*

P7 $f(\mathrm{x})$ *is* increasing-asymptotically-constant: $f'(\mathrm{x}) > 0$, $0 < f(\infty) < \infty$, *and its derivative is decreasing.*

P8 $f(\mathrm{x})$ *is* decreasing-asymptotically-null: $f'(\mathrm{x}) < 0$, $f(\infty) = 0$, *and its derivative is increasing.*

P9 $f(\mathrm{x})$ *is* decreasing-exactly-null: $f'(\mathrm{x}) < 0$, *for* $\mathrm{x} < \bar{\mathrm{x}}$, *and* $f(\mathrm{x}) = 0$ *for* $\mathrm{x} \geq \bar{\mathrm{x}}$ *for some* $\bar{\mathrm{x}} > 0$.

P10 $f(\mathrm{x})$ *is* increasing-asymptotically-unbounded*:* $f'(\mathrm{x}) > 0$, $f(\infty) = +\infty$.

As an example, the terms $d(\cdot)$ and $c(\cdot)$ are associated with Hill functions, which are *sigmoidal* and *complementary sigmoidal* functions. A graphical sketch of their profile is in Figure 3.2 C and D.

Network graphs

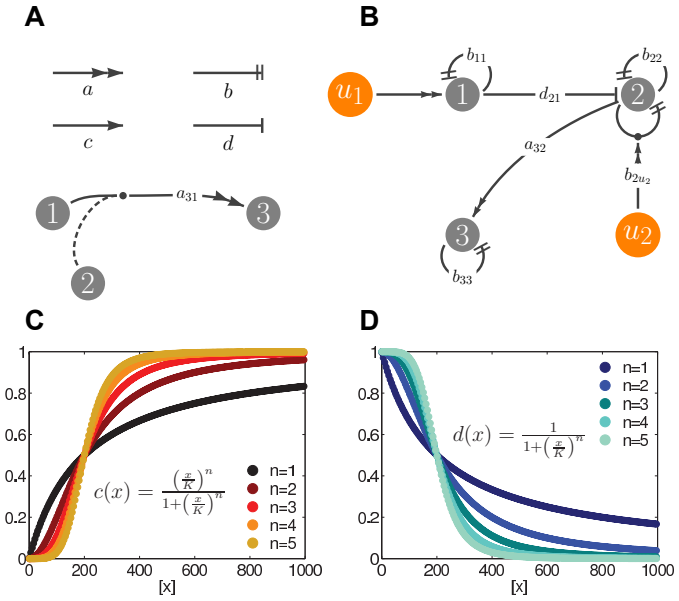


Figure 4.1: A. The arcs associated with the functions a, b, c, and d. We will use dashed arcs, connecting to arcs of the type a and b to highlight that the corresponding function is nonlinearly dependent on a species of the network: in the example above, $a_{31} = a_{31}(\mathrm{x}_2)$. B. The graph associated with equations (4.2); external inputs are represented as orange nodes. C. Examples of sigmoidal functions. D. Examples of complementary sigmoidal functions. In our general model (4.4), functions $d(\cdot)$ and $c(\cdot)$ are naturally associated with Hill function terms.

Building a dynamical model for a biological system is often a long and challenging process. For instance, to reveal dynamic interactions among a pool of genes of interest, biologists may need to selectively knockout genes, set up micro-RNA assays, or integrate fluorescent reporters in the genome. The data derived from such experiments are often noisy and uncertain, which implies that

the estimated model parameters will also be uncertain. However, in general *qualitative trends* can be reliably assessed in the dynamic or steady-state correlation of biological quantities.

Graphical representations of such qualitative trends are often used by biologists to provide intuition regarding the network main features. We believe that, indeed, such graphs may be useful even to immediately construct models analogous to (4.3). A specific method to construct such graphs can be outlined: the biochemical species of the network are associated to the nodes in the graph; the qualitative relationships between the species are instead associated with different types of arcs: in particular, the terms of a, b, c, and d are represented as arcs having different end-arrows, as shown in Figure 4.1 A.

These graphs can be immediately constructed by knowing the correlation trends among the species of the network, and aid in the construction of a dynamical model. For simple networks, this type of graph may provide intuition regarding their behavior and may facilitate their structural robustness analysis. For instance, the graph associated to equations (4.2) is shown in Figure 4.1 B. Throughout this chapter, similar case studies will be considered and their graph representation will be used as a visual support for the analysis.

Remark 1 Here, properties such as positivity, monotonicity, boundedness, and other functional characteristics are labeled as "qualitative and structural properties"[89]. Through such properties, we can draw conclusions on the dynamic behaviors of the considered systems without requiring specific knowledge of parameters and without numerical simulations. However, it is clear that this approach requires more information than other methods, such as Boolean networks and other graph-based frameworks.

Investigation method

The main objective of this work is to show that, at least for reasonably simple networks, structural robust stability can be investigated with simple analytical methods, without the need for extensive numerical analysis. A two-stage approach will be followed:

- Preliminary screening: establish essential information on the network structure, recognizing which properties (such as P1–P10) pertain to each link.
- Analytical investigation: infer robustness properties based on dynamical systems tools such as Lyapunov theory, set-invariance, and linearization.

4.3 Results and discussion

Five biological networks are considered as case studies in [19]. Three of such examples, the L-arabinose, the sRNA, and the *lac* operon networks, model the interaction and control of expression

of a set of genes. The cAMP and the MAPK pathways are instead signaling networks, namely they represent sets of chemical species interacting for transmission and processing of upstream input signals. These networks are all well-known in the literature, and have been characterized mainly through experimental and numerical methods, although the MAPK pathway, for instance, has been thoroughly analyzed using the theory of monotone systems [10].

Here, I will present two complete examples from [19], including all the technical proofs: the L-arabinose and the sRNA networks. For the remaining case studies, I will only report the a brief introduction and the main theorems without the full demonstrations. Please refer to [19] and the corresponding Supplementary Information file for the complete proofs of all the technical results on the *lac* operon, and the cAMP and MAPK pathways.

4.3.1 The L-arabinose network

The arabinose network is analyzed in [78] as an example of feedforward loop. Two genes *araBAD* and *araFGH* are regulated by two transcription factors, AraC and CRP. AraC is a repressor, but turns into an activator when bound to the sugar L-arabinose. CRP is an activator when bound to the inducer cyclic AMP (cAMP), which is produced when cells do not have access to a sufficient amount of glucose (not produced during growth on glucose). CRP also binds to the *araC* promoter and enhances transcription of AraC, which has a significant basal rate of expression (i.e., it is produced by the cell also in absence of inducer CRP). A very simple model for this network can be derived by defining the state variables x_1 and x_2, the concentrations of the transcription factor AraC and of the output protein araBAD, respectively. The concentration of the transcription factor CRP is considered an external input u:

$$\begin{aligned} \dot{\mathrm{x}}_1 &= p_1 + \beta_1 f(\mathrm{u}, \mathrm{K}_{\mathrm{ux}_1}) - \alpha_1 \mathrm{x}_1, \\ \dot{\mathrm{x}}_2 &= \beta_2 f(\mathrm{u}, \mathrm{K}_{\mathrm{ux}_2}) \cdot f(\mathrm{x}_1, K_{\mathrm{x}_1\mathrm{x}_2}) - \alpha_2 \mathrm{x}_2, \end{aligned} \tag{4.5}$$

where α_1, α_2 are the degradation and dilution rates of $\mathrm{x}_1, \mathrm{x}_2$, respectively. The basal production rate of x_1 (AraC) is p_1. The activation pathways are modeled by Hill functions $f(\mathrm{u}, \mathrm{K}) = \mathrm{u}^{\mathrm{H}}/(\mathrm{K}^{\mathrm{H}} + \mathrm{u}^{\mathrm{H}})$, where H is the Hill coefficient and K_{ij} are the activation thresholds. The model can be recast into the general structure (4.4), which includes model (4.5) as special case:

$$\begin{aligned} \dot{\mathrm{x}}_1 &= c_1 + c_{1u}(\mathrm{u}) - \mathrm{b}_{11}\mathrm{x}_1, \\ \dot{\mathrm{x}}_2 &= c_{2u1}(u, \mathrm{x}_1) - b_{22}\mathrm{x}_2, \end{aligned} \tag{4.6}$$

where u is *nonnegative-constant*; c_1, b_{11}, and b_{22} are *positive-constant*; while $c_{1u}(\mathrm{u})$ and $c_{2u1}(\mathrm{u})$ are *sigmoidal* with respect to u, the latter increasing with respect to x_1. The graph representation of

this network is in Figure 4.3 A.

For this elementary network the analysis is straightforward. Variable x_1 is not affected by x_2. Since $c_{1u}(u)$ is bounded, x_1 is also bounded and converges to an equilibrium point $\bar{x}_1(u)$ which is monotonically increasing in u. In turn, x_2 is also positive and bounded for any value of u and stably converges to a unique equilibrium point $\bar{x}_2$, which is a monotonically increasing function of u (partially activated by $\bar{x}_1(u)$). The positive term c_1 prevents $x_1(t)$ and $x_2(t)$ from staying at zero. It is worth remarking that the hierarchical structure of this network greatly facilitates the analysis; equilibria can in fact be iteratively found and their stability properties characterized.

4.3.2 The sRNA pathway

Small regulatory RNAs (sRNA) play a fundamental role in the stress response of many bacteria and eukaryotes. In short, when the organism is subject to a stimulus that threatens the cell survival, certain sRNA species are transcribed and can down-regulate the expression of several other genes. For example, when *E. coli* cells are lacking a source of iron, the sRNA RyhB (normally repressed by the ferric uptake regulator Fur) is expressed and rapidly induces the degradation of at least 18 other RNA species encoding for non-essential proteins which use up Fe molecules. This allows essential iron-dependent pathways to use the low amount of available iron. Quantitative studies of the sRNA pathways have been carried out in [70, 83, 85]. Let us define x_1 as the RNA concentration of the species which is targeted by the sRNA and x_2 as the concentration of sRNA. The model often proposed in the literature is:

$$\begin{aligned} \dot{x}_1 &= \alpha_1 - \beta_1 x_1 - k x_1 x_2, \\ \dot{x}_2 &= \alpha_2 - \beta_2 x_2 - k x_1 x_2, \end{aligned} \tag{4.7}$$

where α_1, α_2 are the transcription rates of x_1 and x_2, respectively; β_1, β_2 are their degradation rates (turnover); and k is the binding rate of the species x_1 and x_2. The formation of the inactive complex $x_1 \cdot x_2$ corresponds to a depletion of both free molecules of x_1 and x_2. If $\alpha_1 < \alpha_2$ the pathway successfully suppresses the expression of the non-essential gene encoded by x_1.

This model can be embedded in the general family:

$$\begin{aligned} \dot{x}_1 &= c_1 - b_{11} x_1 - b_{12}(x_1, x_2) x_2, \\ \dot{x}_2 &= c_2 - b_{22} x_2 - b_{21}(x_1, x_2) x_1, \end{aligned} \tag{4.8}$$

by setting $b_{12} = k x_1$ and $b_{21} = k x_2$ (note that $b_{12}(0) = b_{21}(0)$ as required). From our list of properties: c_1, c_2, b_{11}, and b_{22} are *positive-constant*; $b_{12}(x_1, x_2)$ and $b_{21}(x_1, x_2)$ are *increasing-asymptotically-unbounded* in both variables; and $b_{12}(x_1, x_2)x_2 = b_{21}(x_1, x_2)x_1$ at all times. This

network can be represented with the graph in Figure 4.2 A.

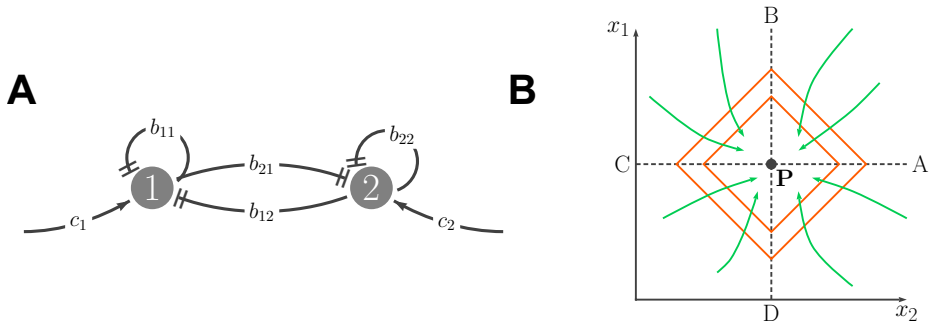


Figure 4.2: A. The graph associated with the sRNA network. B. Sectors, Lyapunov function level curves (orange), and qualitative behavior of the trajectories (green) for the sRNA system

The sRNA system is positive, because the nonnegativity Assumptions 1 and 4 are satisfied. The preliminary screening of this network tells us that each variable produces an inhibition control on the other, which increases with the variable itself. In other words x_1 is "less tolerant" to an increase of x_2 if the latter is present in a large amount. This means that the sum $\mathrm{x}_1 + \mathrm{x}_2$ is strongly kept under control. Also the mismatch between the two variables is controlled.[1] To prove stability of the (unique) equilibrium $\bar{x}$, we will use the 1–norm as Lyapunov function $\mathrm{V}(\mathrm{x} - \bar{\mathrm{x}}) = \|\mathrm{x} - \bar{\mathrm{x}}\|_1$ (see Figure 4.2 B). This choice has a remarkable interpretation: letting $\Sigma = (\mathrm{x}_1 - \bar{\mathrm{x}}_1) + (\mathrm{x}_2 - \bar{\mathrm{x}}_2)$ and $\Delta = (\mathrm{x}_1 - \bar{\mathrm{x}}_1) - (\mathrm{x}_2 - \bar{\mathrm{x}}_2)$ denote the sum and the mismatch of the two variables (referred to the equilibrium), we have

$$\begin{aligned} \mathrm{V}(\mathrm{x} - \bar{\mathrm{x}}) &= \|\mathrm{x} - \bar{\mathrm{x}}\|_1 \\ &= |\mathrm{x}_1 - \bar{\mathrm{x}}_1| + |\mathrm{x}_2 - \bar{\mathrm{x}}_2| \\ &= \max\{\Delta, \Sigma\}; \end{aligned}$$

thus the function represents the worst case between the sum and the mismatch.

The following proposition shows that the sRNA pathway is a typical system in which robustness is structurally assured. We report the full demonstration of this proposition, because its steps and the techniques used are a model for the subsequent proofs in this paper.

Proposition 2 *The variables of system (4.8) are bounded for any initial condition* $\mathrm{x}_1(0), \mathrm{x}_2(0) \geq 0$. *The system admits a unique asymptotically stable equilibrium point* $\bar{x} = (\bar{\mathrm{x}}_1, \bar{\mathrm{x}}_2)^T$ *and the convergence is exponential:*

$$\|\mathrm{x}(\mathrm{t}) - \bar{x}\|_1 \leq e^{-\beta t} \|x(0) - \bar{x}\|_1, \tag{4.9}$$

for some $\beta > 0$ *and any* $\mathrm{x}_1(0) \geq 0$, $\mathrm{x}_2(0) \geq 0$. *Moreover, no oscillations are possible around the*

[1] The concentration mismatch is more "softly" controlled, since the derivative of the difference $\dot{\mathrm{x}}_2 - \dot{\mathrm{x}}_1 = c_1 - c_2 + b_{22}\mathrm{x}_2 - b_{11}\mathrm{x}_1$ is not influenced by the nonlinear term $b_{12}(\mathrm{x}_1, \mathrm{x}_2)\mathrm{x}_2 = b_{21}(\mathrm{x}_1, \mathrm{x}_2)\mathrm{x}_1$.

equilibrium, in the sense that the condition $\mathrm{x}_1(\mathrm{t}) = \bar{\mathrm{x}}_1$ *or* $\mathrm{x}_2(\mathrm{t}) = \bar{\mathrm{x}}_2$ *occurs at most once.*

Proof To prove boundedness of the variables we need to show the existence of an invariant set

$$\mathcal{S} = \{\mathrm{x}_1 \geq 0,\ \mathrm{x}_2 \geq 0 :\ \mathrm{x}_1 + \mathrm{x}_2 \leq \kappa\}.$$

Proposition 1 guarantees that the positivity constraints are respected. Then we just need to show that the constraint $\mathrm{x}_1 + \mathrm{x}_2 \leq \kappa$ cannot be violated for sufficiently large $\kappa > 0$. The derivative of function $s(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{x}_1 + \mathrm{x}_2$ is

$$\begin{aligned}
\dot{s}(\mathrm{x}_1, \mathrm{x}_2) &= \dot{\mathrm{x}}_1 + \dot{\mathrm{x}}_2 \\
&= c_1 - b_{11}\mathrm{x}_1 - b_{12}(\mathrm{x}_1, \mathrm{x}_2)\mathrm{x}_2 + c_2 - b_{22}\mathrm{x}_2 - b_{21}(\mathrm{x}_1, \mathrm{x}_2)\mathrm{x}_1 \\
&\leq c_1 - b_{11}\mathrm{x}_1 + c_2 - b_{22}\mathrm{x}_2 \\
&\leq c_1 + c_2 - \min\{b_{11}, b_{22}\}(\mathrm{x}_1 + \mathrm{x}_2) \\
&= c_1 + c_2 - \min\{b_{11}, b_{22}\}s(\mathrm{x}_1, \mathrm{x}_2).
\end{aligned}$$

Define $\kappa = (c_1 + c_2)/\min\{b_{11}, b_{22}\}$, then for $s(\mathrm{x}_1, \mathrm{x}_2) > \kappa$ the derivative becomes negative, so $s(\mathrm{x}_1, \mathrm{x}_2)$ cannot exceed κ (See Theorem 1).

Boundedness of the solution inside a compact set assures the existence of an equilibrium point. Let $(\bar{\mathrm{x}}_1, \bar{\mathrm{x}}_2)$ be any point in which the following equilibrium conditions holds:

$$c_1 - b_{11}\bar{\mathrm{x}}_1 - b_{12}(\bar{\mathrm{x}}_1, \bar{\mathrm{x}}_2)\bar{\mathrm{x}}_2 = 0, \quad c_2 - b_{22}\bar{\mathrm{x}}_2 - b_{21}(\bar{\mathrm{x}}_1, \bar{\mathrm{x}}_2)\bar{\mathrm{x}}_1 = 0. \tag{4.10}$$

The behavior of the candidate Lyapunov function

$$\mathrm{V}(\mathrm{x}_1, \mathrm{x}_2) = |\mathrm{x}_1 - \bar{\mathrm{x}}_1| + |\mathrm{x}_2 - \bar{\mathrm{x}}_2| = \max\{\pm(\mathrm{x}_1 - \bar{\mathrm{x}}_1) \pm (\mathrm{x}_2 - \bar{\mathrm{x}}_2)\},$$

will be examined in the different sectors represented in Figure 4.2 B. Let us start by considering the sector defined by $\mathrm{x}_1 \geq \bar{\mathrm{x}}_1$ and $\mathrm{x}_2 \geq \bar{\mathrm{x}}_2$ (APB in Figure 4.2 B) for which $\mathrm{V}(\mathrm{x}_1, \mathrm{x}_2) = (\mathrm{x}_1 - \bar{\mathrm{x}}_1) +$

$(x_2 - \bar{x}_2)$. In such a sector the Lyapunov derivative is:

$$\begin{aligned} D^+V(x_1,x_2) =& \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \\ =& \quad c_1 + c_2 - b_{11}x_1 - b_{22}x_2 - b_{12}(x_1,x_2)x_1 - b_{21}(x_1,x_2)x_2 \\ =& - b_{11}(x_1 - \bar{x}_1) - b_{22}(x_2 - \bar{x}_2) - [b_{12}(x_1,x_2)x_1 - b_{12}(\bar{x}_1,\bar{x}_2)\bar{x}_1] - \\ & - [b_{21}(x_1,x_2)x_2 - b_{21}(\bar{x}_1,\bar{x}_2)\bar{x}_2] \\ <& - b_{11}(x_1 - \bar{x}_1) - b_{22}(x_2 - \bar{x}_2), \end{aligned}$$

where we have subtracted the null terms (4.10) and where we have exploited the fact that $b_{12}(x_1,x_2)x_1 = b_{21}(x_1,x_2)x_2$ is increasing in both variables. The inequality (CPD in Figure 4.2 B) $D^+V(x_1,x_2) < b_{11}(x_1 - \bar{x}_1) + b_{22}(x_2 - \bar{x}_2) < 0$ can be similarly proved to hold in the sector $x_1 \leq \bar{x}_1$ and $x_2 \leq \bar{x}_2$. Consider the sector defined by $x_1 \geq \bar{x}_1$ and $x_2 \leq \bar{x}_2$ (DPA in Figure 4.2 B) for which $V(x_1,x_2) = x_1 - \bar{x}_1$; in such a sector the Lyapunov derivative is:

$$\begin{aligned} D^+V(x_1,x_2) =& \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \\ =& c_1 - c_2 - b_{11}x_1 + b_{22}x_2 \underbrace{-b_{12}(x_1,x_2)x_1 + b_{21}(x_1,x_2)x_2}_{=\,0 \text{ by assumption}} \\ =& - b_{11}(x_1 - \bar{x}_1) + b_{22}(x_2 - \bar{x}_2) < 0. \end{aligned}$$

Note that in the last step the null terms (4.10) have been added and subtracted. In the opposite sector (BPC in Figure 4.2 B) $x_1 \leq \bar{x}_1$ and $x_2 \geq \bar{x}_2$; we can prove that $D^+V(x_1,x_2) = +b_{11}(x_1 - \bar{x}_1) - b_{22}(x_2 - \bar{x}_2) < 0$.
We just proved that

$$\begin{aligned} D^+V(x_1,x_2) &\leq -[b_{11}|x_1 - \bar{x}_1| + b_{22}|x_2 - \bar{x}_2|] \leq -\beta[|x_1 - \bar{x}_1| + |x_2 - \bar{x}_2|] \\ &\leq -\beta V(x_1,x_2), \end{aligned}$$

with $\beta = \min\{b_{11}, b_{22}\}$. This implies (4.9) and the uniqueness of the equilibrium point.

We finally need to show that there are no oscillations. To this aim, notice that the sectors DPA, $x_1 \geq \bar{x}_1$ and $x_2 \leq \bar{x}_2$, and its opposite CPB, $x_1 \leq \bar{x}_1$ and $x_2 \geq \bar{x}_2$, are both positive invariant sets.

We can apply Nagumo's theorem: consider the half-line PA originating in P, where $x_2 = \bar{x}_2$ and

$x_1 \geq \bar{x}_1$. Therefore we have that (again by adding the null term in (4.10)):

$$\begin{aligned}\dot{x}_2 = &\quad c_2 - b_{22}\bar{x}_2 - b_{21}(x_1, \bar{x}_2)x_1 - c_2 - b_{22}\bar{x}_2 - b_{21}(\bar{x}_1, \bar{x}_2)\bar{x}_1 \\ &= -\left[b_{21}(x_1, \bar{x}_2)x_1 - b_{21}(\bar{x}_1, \bar{x}_2)\bar{x}_1\right] \leq 0.\end{aligned}$$

Similarly, on half-line PD where $x_1 = \bar{x}_1$ and $x_2 \leq \bar{x}_2$, by considering (4.10) we derive

$$\dot{x}_1 = -b_{12}(\bar{x}_1, x_2)x_2 + b_{12}(\bar{x}_1, \bar{x}_2)\bar{x}_2 \geq 0;$$

hence the claimed invariance of sector DPA. The proof of the invariance of sector CPB is identical. □

Remark 2 *Note that the constructed Lyapunov function* $\|x - \bar{x}\|_1$ *does not depend on the system parameters. This fact can be used to prove that if the transcription rates* $c_1(t)$ *and* $c_2(t)$ *are time–varying, but bounded, we have convergence to a neighborhood whose amplitude, obviously, depends on the bounds of* $c_1(t)$ *and* $c_2(t)$*. It is realistic to assume that the transcription rates vary over time: for instance, if the environmental conditions change, the cell may need to down- or up-regulate entire groups of transcripts and therefore increase or decrease* $c_2(t)$*.*

The following corollary demonstrates the positive influence of c_2, which is positive over x_2 and negative over x_1.

Corollary 1 *Assume that* $x_1(0), x_2(0)$ *is at the steady-state corresponding to* $\hat{c}_1$ *and* $\hat{c}_2$*. Consider the new input* $c_2 > \hat{c}_2$ *(keeping* $c_1 = \hat{c}_1$*). Then the system converges to a new equilibrium with* $\bar{x}_1 < x_1(0)$ *and* $\bar{x}_2 > x_2(0)$*. There is no undershoot, nor overshoot.*

Proof The steady-state values $\bar{x}_1$ and $\bar{x}_2$ are, respectively, monotonically decreasing and increasing functions of c_2. Indeed, consider the steady–state condition

$$\begin{aligned}c_1 &= b_1 x_1 + b_{12}(x_1, x_2)x_2, \\ c_2 &= b_2 x_2 + b_{21}(x_1, x_2)x_1,\end{aligned}$$

and regard it as a differentiable map $(x_1, x_2) \to (c_1, c_2)$. By the uniqueness proved in Proposition 2

the map is invertible. The Jacobian of the inverse map is the inverse of the Jacobian

$$J^{-1} = \begin{bmatrix} b_1 + \frac{\partial(b_{12}x_2)}{\partial x_1} & \frac{\partial(b_{12}x_2)}{\partial x_2} \\ \frac{\partial(b_{21}x_1)}{\partial x_1} & b_2 + \frac{\partial(b_{21}x_1)}{\partial x_2} \end{bmatrix}^{-1}$$
$$= \frac{1}{det(J)} \begin{bmatrix} b_2 + \frac{\partial(b_{21}x_1)}{\partial x_2} & -\frac{\partial(b_{12}x_2)}{\partial x_2} \\ -\frac{\partial(b_{21}x_1)}{\partial x_1} & b_1 + \frac{\partial(b_{12}x_2)}{\partial x_1} \end{bmatrix},$$

where $det(J) = b_1 b_2 + b_2 \frac{\partial(b_{12}x_2)}{\partial x_1} + b_1 \frac{\partial(b_{21}x_1)}{\partial x_2} > 0$ (keep in mind that $b_{21}(x_1, x_2)x_1 = b_{12}(x_1, x_2)x_2$). The sign of the entries in the second column are negative and positive, respectively, therefore, the steady–state values $\bar{x}_1$ and $\bar{x}_2$ are decreasing and increasing functions of c_2.

The absence of overshoot and undershoot is an immediate consequence of the invariance of the sector $x_1 \geq \bar{x}_1$ and $x_2 \leq \bar{x}_2$ previously proved. □

Obviously, decreasing c_2 increases x_1 and decreases x_2. The same holds if we commute the indexes 1 and 2. It is worth noting that the same conclusions regarding the lack of multistability and oscillations for the sRNA pathway may be reached by qualitative analysis of the system's nullclines.

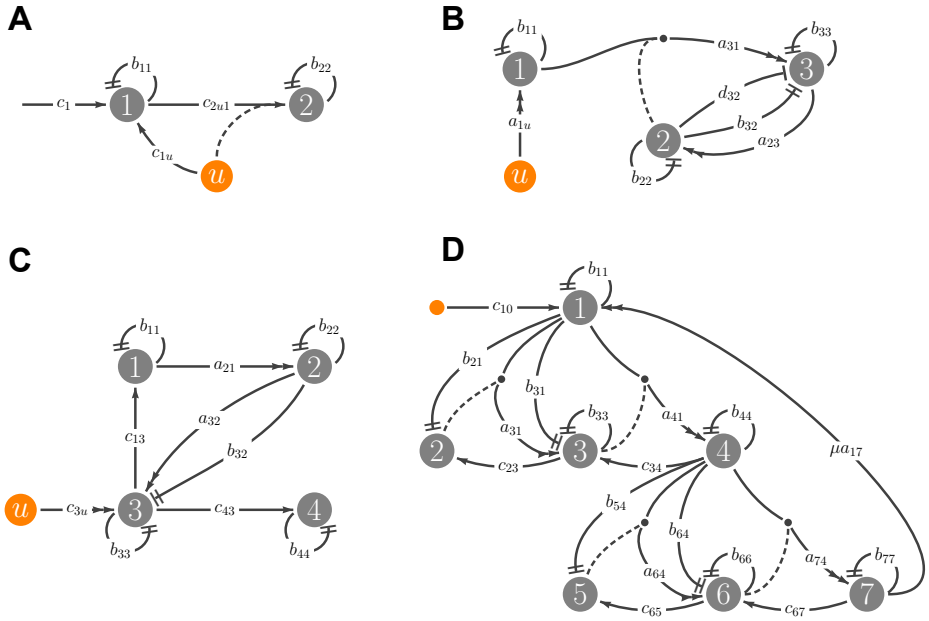


Figure 4.3: A. The graph associated with the L-arabinose network; external inputs are represented as orange nodes. B. The graph associated with the cAMP pathway. C. The graph associated with the *lac* operon network. D.The graph associated with the MAPK signaling pathway.

4.3.3 The cAMP dependent pathway

The cyclic adenosine monophosphate (cAMP) pathway can activate enzymes and regulate gene expression based on sensing of extracellular signals. Such signals are sensed by the G protein-coupled receptors on the cell membrane. When a receptor is activated by its extracellular ligand, a series of conformational changes are induced in the receptor and in the attached intracellular G protein complex; the latter activates adenylyl cyclase, which catalyzes the conversion of ATP in cAMP. In yeast, cAMP causes the activation of the protein kinase A (PKA), which in turn regulates the cell growth, metabolism, and stress response.

Stochastic models are usually proposed for numerical analysis of the cAMP pathway. However, the cAMP pathway components in yeast are present in high numbers and a deterministic modeling approach is adopted in [132]. In such paper, the pathway is decomposed in several modules; here we consider the simplified cAMP Model A, which focuses on the parts of the pathway best characterized in the literature:

$$\begin{aligned} \dot{\mathrm{x}}_1 &= k_F(\mathrm{x}_1^{tot} - \mathrm{x}_1)\mathrm{u} - k_R\mathrm{x}_1, \\ \dot{\mathrm{x}}_2 &= k_F(\mathrm{x}_2^{tot} - \mathrm{x}_2)\mathrm{x}_3 - k_R\mathrm{x}_2, \\ \dot{\mathrm{x}}_3 &= \frac{\alpha_3 + k_A\mathrm{x}_1}{1 + k_I\mathrm{x}_2} - \frac{V_{maxP_1}\mathrm{x}_2\mathrm{x}_3}{K_{MP_1} + \mathrm{x}_3} - \frac{V_{maxP_2}\mathrm{x}_3}{K_{MP_2} + \mathrm{x}_3}, \end{aligned} \tag{4.11}$$

where x_1 is the concentration of active G protein, x_2 is the concentration of active PKA protein, x_3 is the concentration of cAMP, and u is the concentration of glucose input to the network. The parameters V_{maxP_1} and V_{maxP_2} model the "feedback" effect introduced by two phosphodiesterases (Pde1p and Pde2p) on the cAMP concentration. The numerical analysis in [132] typically shows that the cAMP concentration (x_3) responds with a large overshoot to steps in the glucose (u, input) concentration. We will analytically explore the dynamic behavior of x_3, showing that under certain assumptions, a bounded overshoot is a robust characteristic in the system. The parameters k_F and k_R are forward and reverse reaction rates for the formation of active x_1 and x_2. Mass conservation allows expression of the active protein amounts as a function of the total number of molecules, $\mathrm{x}_i^{tot} = \mathrm{x}_i^{inactive} + \mathrm{x}_i$. The nonlinear expressions in equation x_3 are derived by Michaelis-Menten enzymatic steps. We can re-write the above equations according to the general model (4.4):

$$\begin{aligned} \dot{\mathrm{x}}_1 &= a_{1u}(\mathrm{x}_1)\mathrm{u} - b_{11}\mathrm{x}_1, \\ \dot{\mathrm{x}}_2 &= a_{23}(\mathrm{x}_2)\mathrm{x}_3 - b_{22}\mathrm{x}_2, \\ \dot{\mathrm{x}}_3 &= d_{32}(\mathrm{x}_2) + a_{31}(\mathrm{x}_2)\mathrm{x}_1 - b_{32}(\mathrm{x}_3)\mathrm{x}_2 - b_{33}(\mathrm{x}_3)\mathrm{x}_3, \end{aligned} \tag{4.12}$$

where u is the external signal and where $b_{23} = 0$ for $\mathrm{x}_2 = 0$ and $b_{32} = 0$ for $\mathrm{x}_3 = 0$. A qualitative graphical representation of this network is in Figure 4.3 B.

Our preliminary analysis allows us to assume: a_{1u}, a_{23}: *decreasing-exactly-null* with threshold values $\mathrm{x}_1^{\mathrm{tot}}$ and $\mathrm{x}_2^{\mathrm{tot}}$; d_{32}, a_{31}: *decreasing-asymptotically-null*; b_{32} and $g_{33} = b_{33}(\mathrm{x}_3)\mathrm{x}_3$: *increasing-asymptotically-constant*; b_{11}, b_{22}: *positive-constant*.

It can immediately be noticed that for constant u, x_1 robustly converges asymptotically to its equilibrium value such that

$$u = \frac{b_{11}\mathrm{x}_1}{a_{1u}(\mathrm{x}_1)} \doteq \xi^{-1}(\mathrm{x}_1).$$

The solution $\bar{\mathrm{x}}_1 = \xi(\mathrm{u})$ of the previous equation is uniquely defined for each u since the function $\xi^{-1}(\mathrm{x}_1)$ on the right is strictly increasing and grows to infinity, precisely $\lim_{\mathrm{x}_1 \to x_1^{tot}} \xi^{-1}(\mathrm{x}_1) = +\infty$. Biologically, this means that external glucose signals are mapped to internal active G-protein concentration with a bijection, before saturating.

Also note that the model is consistent with mass conservation: since $a_{1u}(\mathrm{x}_1)$ and $a_{23}(\mathrm{x}_2)$ are zero above the thresholds $\mathrm{x}_1^{\mathrm{tot}}$ and $\mathrm{x}_2^{\mathrm{tot}}$, we have that $\dot{\mathrm{x}}_1 < 0$ and $\dot{\mathrm{x}}_2 < 0$ for $\mathrm{x}_1 > x_1^{tot}$ and $\mathrm{x}_2 > x_2^{tot}$, respectively. Therefore we assume $\mathrm{x}_1(\mathrm{t}) \leq x_1^{tot},\ \mathrm{x}_2(\mathrm{t}) \leq x_2^{tot},$ for all $t \geq 0$.

Proposition 3 *There exists an equilibrium for system* (4.12) *if and only if*

$$d_{32}(\mathrm{x}_2^{tot}) + a_{31}(\mathrm{x}_2^{tot})\bar{\mathrm{x}}_1 < \lim_{\mathrm{x}_3 \to \infty} \left[b_{32}(\mathrm{x}_3)\mathrm{x}_2^{tot} + b_{33}(\mathrm{x}_3)\mathrm{x}_3 \right], \tag{4.13}$$

where $\bar{\mathrm{x}}_1 = \xi(\mathrm{u})$ *as previously defined. All the equilibrium values* $\bar{\mathrm{x}}_1 = \xi(\mathrm{u})$, $\bar{\mathrm{x}}_2$, *and* $\bar{\mathrm{x}}_3$ *are increasing functions of* u. *If condition (4.13) is satisfied, the equilibrium is unique and locally stable.*

The previous proposition assures only local stability, but this result can be extended to global stability. To this aim, we will assume that x_1 is at its equilibrium value $\bar{\mathrm{x}}_1$. Furthermore, under a suitable condition a performance bound on the transient values of $\mathrm{x}_3(\mathrm{t})$ can be given.

Proposition 4 *Assume that* x_1 *has reached its steady-state* $\bar{\mathrm{x}}_1$. *Then, the unique equilibrium point is globally attractive for any initial condition* $\mathrm{x}_2(0), \mathrm{x}_3(0) \geq 0$. *Moreover, assume that*

$$l_3 \doteq \lim_{\mathrm{x}_3 \to \infty} b_{33}(\mathrm{x}_3)\mathrm{x}_3 > d_{32}(0) + a_{31}(0)\xi, \tag{4.14}$$

then we can give the following bound for the transient of $\mathrm{x}_3(\mathrm{t})$

$$\mathrm{x}_3(\mathrm{t}) \leq \max\{\mathrm{x}_3(0), d_{32}(0) + a_{31}(0)\xi\}. \tag{4.15}$$

The proof can be found in Section S II of the Supplementary Information of [19].

Remark 3 *The condition (4.14) has the following interpretation. It basically states that the inhibiting term* $b_{33}(\mathrm{x}_3)\mathrm{x}_3$ *at "full force" (*x_3 *suitably large) dominates the activating term* $d_{32}(\mathrm{x}_2) + a_{31}(\mathrm{x}_2)\xi$

when x_2 *is small. Note that, indeed, the feedback terms modulated by the two phosphodiesterases act in a complementary manner, in order to maintain a bounded concentration of cAMP in the cell.*

Remark 4 *The system, even if initialized with small values* $x_2(0)$ *and* $x_3(0)$*, may exhibit a spike of cAMP* x_3 *which is bounded by (4.15), if condition (4.14) is satisfied. If* $x_3(0)$ *is small, then the bound is* $d_{32}(0) + a_{31}(0)\xi(u)$*: the amplitude of the spike is, in general, an increasing function of the glucose concentration* u*. If condition (4.14) fails, then (see Figure S2, [19] Supplementary Information file) the spike of* $x_3(t)$ *can be arbitrarily large; thus condition (4.14) can be seen as a threshold.*

4.3.4 The *lac* operon

This genetic network was originally studied by Monod and Jacob [53]. The natural nutrient for *E. coli* bacterial cells is glucose, which is metabolized by enzymes normally produced by the bacteria. When glucose is absent, but the allolactose inducer is present in their environment, *E. coli* activates a set of genes that will regulate the lactose intake and breakdown. In particular, the cells start producing a permease protein, which binds to the cell membrane and increases the inflow of lactose; and cells also start producing the β-galactosidase protein, which converts lactose in allolactose.

In this section we will consider the deterministic model proposed in [130]. This simple model does not capture the stochasticity of this genetic circuit, but it does explain the bimodal behavior of the system. Such behavior is observable experimentally: within the same population, the operon can be either induced or uninduced. Our analysis shows that for low or high intracellular inducer concentrations, the system is monostable and reaches, respectively, an uninduced or induced equilibrium; however, at intermediate inducer concentrations the system becomes multi-stable.

The state variables of the ODE model we will study are the concentration of nonfunctional permease protein x_1; the concentration of functional permease protein x_2; the concentration of inducer (allolactose) inside the cell x_3; and the concentration of β-galactosidase x_4, a quantity that can be experimentally measured. The concentration of inducer external to the cell is here denoted as an input function u.

$$\begin{aligned} \dot{x}_1 &= f_1(x_3) - \delta_1 x_1, \\ \dot{x}_2 &= \beta_1 x_1 - \delta_2 x_2, \\ \dot{x}_3 &= [f_2(u) - f_3(x_3)]x_2 + \beta_2 u - \delta_3 x_3, \\ \dot{x}_4 &= \gamma f_1(x_3) - \delta_4 x_4, \end{aligned} \tag{4.16}$$

where $\beta_1, \beta_2, \delta_1, \delta_2, \delta_3$, and γ are constants, and f_i's are functions that are experimentally measurable. In particular, at low inducer concentrations, $f_1 \approx k_1 + k_2 x_3 + k_3 x_3^2$ where k_i's are constant; at high x_3 concentrations f_1 saturates. The functions f_2 and f_3 are assumed to depend hyperbolically on

their arguments. According to the proposed setup, the previous equations can rewritten as follows:

$$\begin{aligned}
\dot{\mathrm{x}}_1 &= c_{13}(\mathrm{x}_3) - b_{11}\mathrm{x}_1, \\
\dot{\mathrm{x}}_2 &= a_{21}\mathrm{x}_1 - b_{22}\mathrm{x}_2, \\
\dot{\mathrm{x}}_3 &= a_{32}(\mathrm{u})\mathrm{x}_2 - b_{32}(\mathrm{x}_3)\mathrm{x}_2 + c_{3u}\mathrm{u} - b_{33}\mathrm{x}_3, \\
\dot{\mathrm{x}}_4 &= c_{43}(\mathrm{x}_3) - b_{44}\mathrm{x}_4,
\end{aligned} \tag{4.17}$$

where $c_{13}(\mathrm{x}_3) = f_1(\mathrm{x}_3)$, $b_{11} = \delta_1$, $a_{21} = \beta_1$, $b_{22} = \delta_2$, $a_{32}(\mathrm{u}) = f_2(\mathrm{u}) =$, $b_{32}(\mathrm{x}_3) = f_3(\mathrm{x}_3)$, $c_{3u} = \beta_2$, $b_{33} = \delta_3$, $c_{43}(\mathrm{x}_3) = \gamma f_1(\mathrm{x}_3)$, and $b_{44} = \delta_4$. This corresponds to the network in Figure 4.3 C.
From our preliminary analysis step: c_{13} is *constant-sigmoidal*, $a_{32}(\mathrm{u})$ and $b_{32}(\mathrm{x}_3)$ are *increasing-asymptotically-constant*, and the remaining functions a_{21}, b_{11}, b_{22}, and b_{33} are *positive-constant*.

We can start to study this network without any specific knowledge of the parameters in equations (4.17). First of all, as evident in Figure 4.3 C, note that the β-galactosidase concentration x_4 does not affect any other chemical species: therefore, the fourth equation can be considered separately. As long as the inducer concentration of x_3 within the cell reaches an equilibrium $\bar{\mathrm{x}}_3$, x_4 converges to $\bar{\mathrm{x}}_4 = c_{43}(\bar{\mathrm{x}}_3)/b_{44}$. Therefore, we can restrict our attention to the first three equations; this is consistent with the model proposed in [135, 136]. From now on we will consider this reduced model (see [19], Supplementary Information, Section S II), neglecting the linear term $c_{3u}u$ as in [135, 136]. We will not introduce delays in our model, as done in [136].

Our preliminary screening also shows that the evolution of this system is necessarily bounded. Indeed x_1 receives a bounded signal from x_3 and the degradation term $-b_{11}\mathrm{x}_1$ keeps x_1 bounded. In turn, x_2 remains bounded. The inducer concentration x_3 receives a bounded signal u and x_2; therefore x_3 stays bounded as well, being both $a_{32}(\mathrm{u})$ and $b_{32}(\mathrm{x}_3)$ bounded.

The following proposition evidences that fundamental results can be established starting from our general framework. These results are consistent with the findings in [136], whose analysis relies on assuming Hill-type functions in the model.

Proposition 5 *For any functional terms in Equations* (4.17)*, satisfying the general assumptions formulated above, the system admits a unique equilibrium for large* $\mathrm{u} > 0$ *or small* $\mathrm{u} > 0$.

For some chioces of such functional terms, the system may have multiple positive equilibria $\mathrm{x}^{\mathrm{A}}, \mathrm{x}^{\mathrm{B}}, \mathrm{x}^{\mathrm{C}}, \cdots \in \mathbb{R}^3$ *(typically three) for intermediate values of* u. *If multiple equilibria exist, then they are ordered in the sense that* $\mathrm{x}^{\mathrm{A}} \leq \mathrm{x}^{\mathrm{B}} \leq \mathrm{x}^{\mathrm{C}} \ldots$ *where the inequality has to be considered component-wise. If the equilibria are all distinct, then they are alternatively stable and unstable. In the case of three equilibria,* $\mathrm{x}^{\mathrm{A}}, \mathrm{x}^{\mathrm{B}}, \mathrm{x}^{\mathrm{C}}$ *they are stable, unstable, and stable, respectively. Finally, given any equilibrium point, the positive and negative cones* $\mathrm{x} \leq \mathrm{x}^*$ *and* $\mathrm{x} \geq \mathrm{x}^*$ *are positively invariant.*

The proof is given in [19], Supplementary Information, Section S III. The cone–invariance property

implies that the state variables cannot exhibit oscillations around their equilibria. For instance, if x^A is the first (hence stable) equilibrium, given any initial condition upper bounded by x^A ($x(0) \leq x^A$) in the domain of attraction, the convergence to x^A has no overshoot (and if $x(0) \geq x^A$ there is no undershoot).

Remark 5 It is interesting to notice that, due to the competition between terms a_{32} and b_{32}, the considered model for the *lac* network is not a monotone system according to the definition in [121], where a different model was considered.

4.3.5 The MAPK signaling pathway

Mitogen-activated protein (MAP) kinases are proteins that respond to the binding of growth factors to cell surface receptors. The pathway consists of three enzymes, MAP kinase, MAP kinase kinase (MAP2K), and MAP kinase kinase kinase (MAP3K) that are activated in series. Activation or phosphorylation means the addition of a phosphate group to the target protein. Extracellular signals can activate MAP3K, which in turn phosphorylates MAP2K at two different sites; in the last round, MAP2K phosphorylates MAPK at two different sites. The MAP kinase signaling cascade can transduce a variety of growth factor signals, and has been evolutionary conserved from yeast to mammals.

Several experimental studies have highlighted the presence of feedback loops in this pathway, which result in different dynamic properties. This work will focus on a specific positive-feedback topology, where doubly-phosphorylated MAPK has an activation effect on MAP3K. Such positive feedback has been extensively studied in the literature since the biochemical analysis of Huang and Ferrell [48, 37] on the MAPK cascade found in *Xenopus* oocytes. In this type of cell, Mos (MAP3K) can activate MEK (MAP2K) through phosphorylation of two residues (converting unphosphorylated MEK to monophosphorylated MEK-P and then bisphosphorylated MEK-PP). Active MEK then phosphorylates p42 (MAPK) at two residues. Active p42 can then promote Mos synthesis, completing the closed positive-feedback loop.

The presence of such positive feedback in the MAPK cascade has been linked to a bi-stable behavior: the switch between two stable equilibria in *Xenopus* oocytes denotes the transition between the immature and mature state. A standard ODE model for the cascade is proposed in [10], where the authors demonstrate bi-stability of the system by applying the general theory of monotone

systems. We adopt such a model, which is reported below:

$$\begin{aligned}
\dot{x} &= f(\mathrm{x})\mathrm{u} + V_1 - \frac{V_2\mathrm{x}}{k_2 + \mathrm{x}},\\
\dot{\mathrm{y}}_1 &= \frac{V_6\mathrm{y}_2}{K_6 + \mathrm{y}_2} - \frac{V_3\mathrm{x}\mathrm{y}_1}{K_3 + \mathrm{y}_1},\\
\dot{\mathrm{y}}_2 &= \frac{V_3\mathrm{x}\mathrm{y}_1}{K_3 + \mathrm{y}_1} + \frac{V_5\mathrm{y}_3}{K_5 + \mathrm{y}_3} - \frac{V_4\mathrm{x}\mathrm{y}_2}{K_4 + \mathrm{y}_2} - \frac{V_6\mathrm{y}_2}{K_6 + \mathrm{y}_2},\\
\dot{\mathrm{y}}_3 &= \frac{V_4\mathrm{x}\mathrm{y}_2}{K_4 + \mathrm{y}_2} - \frac{V_5\mathrm{y}_3}{K_5 + \mathrm{y}_3},\\
\dot{\mathrm{z}}_1 &= \frac{V_{10}\mathrm{z}_2}{K_{10} + \mathrm{z}_2} - \frac{V_7\mathrm{y}_3\mathrm{z}_1}{K_7 + \mathrm{z}_1},\\
\dot{\mathrm{z}}_2 &= \frac{V_7\mathrm{y}_3\mathrm{z}_1}{K_7 + \mathrm{z}_1} + \frac{V_9\mathrm{z}_3}{K_9 + \mathrm{z}_3} - \frac{V_8\mathrm{y}_3\mathrm{z}_2}{K_8 + \mathrm{z}_2} - \frac{V_{10}\mathrm{z}_2}{K_{10} + \mathrm{z}_2},\\
\dot{\mathrm{z}}_3 &= \frac{V_8\mathrm{y}_3\mathrm{z}_2}{K_8 + \mathrm{z}_2} - \frac{V_9\mathrm{z}_3}{K_9 + \mathrm{z}_3},
\end{aligned} \tag{4.18}$$

where x is the concentration of Mos (MAP3K); y_1 is the concentration of unphosphorylated MEK (MAP2K); y_2 is the concentration of phosphorilated MEK-P; y_3 is the concentration of MEK-PP; z_1, z_2, and z_3 are, respectively, the concentrations of unphosphorylated, phosphorylated, and doubly-phosphorylated p42 (MAPK). Finally, u is the input to the system.

While bi-stability may occur due to other phenomena, such as multisite phosphorylation [79], rather than due to feedback loops, a large body of literature focuses on bi-stability induced by the positive-feedback in the Huang-Ferrel model in *Xenopus* [95, 36] reported above. In [48] the feedback $f(\mathrm{u})$ was characterized, through *in vitro* studies, as an activating Hill-function with high cooperativity. In [10] instead, $f(\mathrm{u})$ was assumed to be a first-order linear term in the concentration of active MAP3K, x_7. In Proposition 6, the effects of different qualitative functional assumptions on the feedback loop dynamics $f(\mathrm{u})$ will be explored. The system loses its well-known bi-stability not only in the absence of feedback, but also when the feedback becomes unbounded. An unbounded positive feedback would be caused, for instance, by an autocatalytic process of MAP3K activation, mediated by active MAPK. We choose to rewrite the above model as follows:

$$
\begin{aligned}
\dot{\mathrm{x}}_1 &= a_{17}(\mathrm{x}_1)\ \mu \mathrm{x}_7 + c_{10} - b_{11}(\mathrm{x}_1)\mathrm{x}_1,\\
\dot{\mathrm{x}}_2 &= c_{23}(\mathrm{x}_3) - b_{21}(\mathrm{x}_2)\mathrm{x}_1,\\
\dot{\mathrm{x}}_3 &= a_{31}(\mathrm{x}_2)\mathrm{x}_1 + c_{34}(\mathrm{x}_4) - b_{31}(\mathrm{x}_3)\mathrm{x}_1 - b_{33}(\mathrm{x}_3)\mathrm{x}_3,\\
\dot{\mathrm{x}}_4 &= a_{41}(\mathrm{x}_3)\mathrm{x}_1 - b_{44}(\mathrm{x}_4)\mathrm{x}_4,\\
\dot{\mathrm{x}}_5 &= c_{56}(\mathrm{x}_6) - b_{54}(\mathrm{x}_5)\mathrm{x}_4,\\
\dot{\mathrm{x}}_6 &= a_{64}(\mathrm{x}_5)\mathrm{x}_4 + c_{67}(\mathrm{x}_7) - b_{64}(\mathrm{x}_6)\mathrm{x}_4 - b_{66}(\mathrm{x}_6)\mathrm{x}_6,\\
\dot{\mathrm{x}}_7 &= a_{74}(\mathrm{x}_6)\mathrm{x}_4 - b_{77}(\mathrm{x}_7)\mathrm{x}_7.
\end{aligned}
\tag{4.19}
$$

The term $\mu \mathrm{x}_7$ introduces the positive feedback loop and represents a key parameter for the analysis to follow. A preliminary screening of the system immediately highlights the following properties: function $b_{11}(\mathrm{x}_1)\mathrm{x}_1$; functions $c_{23}(\mathrm{x}_3)$, $b_{21}(\mathrm{x}_2)$, $a_{41}(\mathrm{x}_3)$, and $b_{44}(\mathrm{x}_4)\mathrm{x}_4$; functions $c_{56}(\mathrm{x}_6)$, $b_{54}(\mathrm{x}_5)$, $a_{74}(\mathrm{x}_6)$, and $b_{77}(\mathrm{x}_7)\mathrm{x}_7$ are *increasing-asymptotically-constant*. Moreover, $a_{31}(\mathrm{x}_2) = b_{21}(\mathrm{x}_2)$, $c_{34}(\mathrm{x}_4) = b_{44}(\mathrm{x}_4)\mathrm{x}_4$, $b_{31}(\mathrm{x}_3) = a_{41}(\mathrm{x}_3)$, $b_{33}(\mathrm{x}_3)\mathrm{x}_3 = c_{23}(\mathrm{x}_3)$ and $a_{64}(\mathrm{x}_5) = b_{54}(\mathrm{x}_5)$, $c_{67}(\mathrm{x}_7) = b_{77}(\mathrm{x}_7)\mathrm{x}_7$, $b_{64}(\mathrm{x}_6) = a_{74}(\mathrm{x}_6)$, $b_{66}(\mathrm{x}_6)\mathrm{x}_6 = c_{56}(\mathrm{x}_6)$. We assume c_{10} to be a *positive-constant*.

The graph in Figure 4.3 D can be partitioned considering three aggregates of variables, precisely $\{\mathrm{x}_1\}$, $\Sigma_{234} = \{\mathrm{x}_2, \mathrm{x}_3, \mathrm{x}_4\}$ and $\Sigma_{567} = \{\mathrm{x}_5, \mathrm{x}_6, \mathrm{x}_7\}$. Signal x_1 is the only input for Σ_{234}; signal x_4 is the only input for Σ_{567}. x_7 is then fed back to the first subsystems by arc a_{17}. Without the positive feedback loop, we will demonstrate that the system is a pure stable cascade. Note also that Σ_{234}, and Σ_{567} can be reduced, since $\dot{\mathrm{x}}_2 + \dot{\mathrm{x}}_3 + \dot{\mathrm{x}}_4 = 0$, and $\dot{\mathrm{x}}_5 + \dot{\mathrm{x}}_6 + \dot{\mathrm{x}}_7 = 0$ and therefore the following sums are constant

$$
\begin{aligned}
\mathrm{x}_2(\mathrm{t}) + \mathrm{x}_3(\mathrm{t}) + \mathrm{x}_4(\mathrm{t}) &= k,\\
\mathrm{x}_5(\mathrm{t}) + \mathrm{x}_6(\mathrm{t}) + \mathrm{x}_7(\mathrm{t}) &= h,
\end{aligned}
\tag{4.20}
$$

with $k \doteq \mathrm{x}_2(0) + \mathrm{x}_3(0) + \mathrm{x}_4(0)$ and $h \doteq \mathrm{x}_5(0) + \mathrm{x}_6(0) + \mathrm{x}_7(0)$. Since $\mathrm{x}_i \geq 0$, all the variables but x_1 are bounded. The system can be studied by removing variables $\mathrm{x}_3 = k - \mathrm{x}_2 - \mathrm{x}_4$ and $\mathrm{x}_6 = h - \mathrm{x}_5 - \mathrm{x}_7$. We must assume that $c_{10} < \lim_{\mathrm{x}_1 \to \infty} b_{11}(\mathrm{x}_1)\mathrm{x}_1$, otherwise no equilibrium is possible. The following result is proved in Section S IV of the Supplementary Information of [19].

Proposition 6 *For $\mu = 0$ the system admits a unique globally asymptotically stable equilibrium.*

For $\mu > 0$, the system may have multiple equilibria, for specific choices of the functions a, b, c.

For $\mu > 0$ suitably large and $a_{17}(\mathrm{x}_1)$ lower bounded by a positive number, then the system has no equilibria.

For $\mu > 0$ bounded and $a_{17}(\mathrm{x}_1)$ increasing, or non-decreasing, and bounded, if multiple simple[2] *equilibria exist, then such equilibria are alternatively stable and unstable. In the special case of three equilibria, then the system is bi-stable.*

For $\mu > 0$ bounded and $a_{17}(\mathrm{x}_1)$ increasing asymptotically unbounded, then the number of equilibria is necessarily even (typically 0 or 2). Moreover, if we assume that there exists $\mu^* > 0$ such that the system admits two distinct equilibria for any $0 < \mu \leq \mu^*$, then one is stable, while the other is unstable.

The proof of this last proposition also shows that multiple equilibria $x^A, x^B, ...$ have a partial order: $\bar{\mathrm{x}}_1^A \leq \bar{\mathrm{x}}_1^B \leq \bar{\mathrm{x}}_1^C \ldots$, $\bar{\mathrm{x}}_4^A \leq \bar{\mathrm{x}}_4^B \leq \bar{\mathrm{x}}_4^C \ldots$, $\bar{\mathrm{x}}_7^A \leq \bar{\mathrm{x}}_7^B \leq \bar{\mathrm{x}}_7^C \ldots$, while $\bar{\mathrm{x}}_2$ and $\bar{\mathrm{x}}_5$ have the reverse order $\bar{\mathrm{x}}_2^A \geq \bar{\mathrm{x}}_2^B \geq \bar{\mathrm{x}}_2^C \ldots$ and $\bar{\mathrm{x}}_5^A \geq \bar{\mathrm{x}}_5^B \geq \bar{\mathrm{x}}_5^C \ldots$.

Remark 6 *The simplest case of constant a_{17} has been fully developed in [10]* [3] *and [121], and it turns out that the system may exhibit bi-stability for suitable values of the feedback strength μ. Here it was shown that, for constant a_{17}, bi-stability is actually a robust property. These results are consistent with the fact that the MAPK cascade is a monotone system and some of them could be demonstrated with the same tools used in [10, 121]. With respect to such literature, the contribution of this work is that of inferring properties such as number of equilibria and mono- or bi-stability starting from qualitative assumptions on the dynamics of the model, without invoking monotonicity.*

Remark 7 *Finally, it is necessary to remark that our results on the MAPK pathway robust behaviors hold true given the model (4.19) and its structure. Other work in the literature shows that feedback loops are not required to achieve a bi-stable behavior in the MAPK cascade [79] when the dual phosphorylation and de-phosphorylation cycles are non-processive (i.e., sites can be phosphorylated/de-phosphorylated independently) and distributed (i.e., the enzyme responsible for phosphorylation/de-phosphorylation is competitively used in the two steps).*

4.4 Conclusions

A property is structurally robust if it is satisfied by a class of systems of a given structure, regardless of the choice of specific expressions adopted and of the parameter values in the model. We have considered five relevant biological examples and proposed capturing their dynamics with parameter-free, qualitative models. It was shown that specific robust properties of such models can be assessed by means of solid theoretical tools based on Lyapunov methods, set–invariance theory, and matrix theory.

[2] I.e., the nullclines have no common tangent lines.

[3] See the erratum: `http://www.math.rutgers.edu/~sontag/FTPDIR/angeli-ferrell-sontag-pnas04-errata.txt` and [104].

Robustness is often tested through simulations, at the price of exhaustive campaigns of numerical trials and, more importantly, with no theoretical guarantee of robustness. We are far from claiming that numerical simulation is useless. It it important, for instance, to falsify "robustness conjectures" by finding suitable numerical counter-examples. Furthermore, for very complex systems in which analytic tools can fail, simulation appears be the last resort. Indeed a limit of the considered theoretical investigation is that its systematic application to more complex cases is challenging. However, the set of techniques we employed can be successfully used to study a large class of simple systems, and are in general suitable for the analytical investigation of structural robustness of biological networks, complementary to simulations and experiments.

Chapter 5

Summary and future work

In this last chapter, I will briefly summarize the contributions of each chapter and outline future research plans.

5.1 Flux regulation

In this chapter I proposed two network architectures based on negative and positive feedback, to regulate and match the output flow rate of two interconnected systems. Feedback is implemented through mass action chemical reactions, which down- or up-regulate the activity of the molecules generating the network output. To my knowledge, this design has not been considered elsewhere in the literature. Numerical simulations and data suggest that feedback confers robustness to the system with respect to certain parametric variations and to initial conditions.

The analytical and experimental results presented in Chapter 2 need substantial refinement. First, the simple model problems 2.3 and 2.7 will be non-dimensionalized in search of key parameter aggregates and nullcline characteristic behaviors. Systematic numerical analysis of the systems will be a useful aid, starting from the results in [42]. Additionally, parameter-free models will be considered, along the lines of those presented in Chapter 4, to explore the structural properties of these feedback schemes. Additional experiments and analysis need to be carried out.

• **Negative auto-regulation scheme:** The experiments shown for this case will be repeated, focusing on gel-based quantitation of the RNA concentration in solution. The gel electrophoresis data currently available were processed using the DNA ladder as a control for concentration, and they may lack accuracy. It will also be interesting to explore the robustness of the system to larger variations of the template concentrations over time and to external disturbances/load processes. The data fits will be improved, extending the fitted parameter set and including gel electrophoresis data.

• **Cross-activation scheme** Further experiments will be run to characterize the unsatisfactory aspects of the current design, with the purpose of understanding which design details should be

improved. In particular, I will focus on the transcription leak and undesired inhibitory pathways. In the future, I plan to fully re-design this system. I will consider the use of translator gates, or of decoupling genelets, to avoid the self inhibitory reactions structurally present in the current design.

It will be interesting to consider a circuit design incorporating both self-inhibition and cross-activation and compare it to the two described schemes. An additional interesting series of experiments will consist of systematically varying the toehold lengths in order to speed up or slow down the feedback loops, and assessing the robustness of the schemes with respect to such rates.

I also plan on exploring the theoretical interconnections between flux matching and consensus problems [109]. This might be useful in finding general feedback schemes to match flows of n interconnected systems.

5.2 Oscillatory systems

Oscillators are a fundamental component in all silicon devices: modern digital clocks synchronize the operations of millions of transistors. This chapter was dedicated to the experimental study of a molecular oscillator to be used as a clock for a downstream molecular device. Mathematical modeling and experiments show that interconnecting the oscillator to its load in a direct manner, i.e. by stoichiometric binding and release, results in undesired back-action effects and loss of the original signal. Loosely speaking, the back-action is primarily caused by mass conservation constraints. This issue is mitigated by the introduction of a molecular insulator, a node draining a small amount of molecules from the oscillator and using them to amplify its signal [27]. Experiments are carried out using the tool kit of transcriptional circuits. Given the results presented in this thesis, it is conceivable to use molecular clocks to orchestrate large-scale molecular processes in parallel. For instance, DNA-based molecular logic gates [110, 97] that operate reaching a single steady state could be dynamically driven with our molecular oscillator.

Another important research milestone is the design of a molecular oscillator whose amplitude and frequency are separately tunable. The *in vitro* oscillator described in this thesis either exhibits fast oscillations with low amplitude, or very slow oscillations with a large swing. Unlike silicon-based oscilloscopes, most synthetic oscillators suffer from some form of correlation between their amplitude and frequency.

Further analysis on the dynamics exhibited by our oscillator is also currently under way using the mathematical framework of monotone systems [11]. Preliminary results have been obtained using the simplified oscillator model in [123]: the system can be interpreted as the interconnection of two monotone systems, where loop gain can be changed through the enzyme production and degradation rates. Additionally, the DNA activator and inhibitor thresholds can be interpreted as elements introducing a delay in the system. The effects of the threshold variations on the oscillator

dynamics shown in Section 3.7.17 may be explained in terms of internal delay variations.

5.3 Robustness in molecular networks

The general analysis framework presented in Chapter 4 relies on two main features: 1) The formulation of simple, parameter-free biological models capturing only essential functional features of the considered system. 2) The use of Lyapunov functions and invariant sets theory to isolate specific properties of the model; because models are parameter–free, such properties are naturally robust with respect to parametric variabilities the real system might present.

Currently, this framework is being applied to the context-dependent pathways exhibited by the MAPK network in neural rat cells [107]. First, two dynamic models (parameter independent) have been formulated starting from the MAPK graph topologies in [107]. (Such graphs were experimentally derived in [107] using the Modular Response Analysis technique [60].) Then, it is possible to demonstrate that depending on the input present, the two topologies lead to either a bi-stable or a mono-stable response. These results, soon to be submitted, were obtained in collaboration with F. Blanchini.

Another application of these techniques is in the area of oscillatory molecular systems. This research is motivated by [129], where the features of several simple models for biological oscillators are explored numerically. It is likely that some of the properties considered in [129] can be proved analytically.

Bibliography

[1] Integrated DNA technologies: http://www.idtdna.com.

[2] Nucleic acid package: http://www.nupack.org.

[3] A. Abate, A. Tiwari, and S. Sastry. Box invariance for biologically-inspired dynamical systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 5162–5167, 2007.

[4] J. Ackermann, B. Wlotzka, and J. S. McCaskill. In vitro DNA-based predator-prey system with oscillatory kinetics. *Bulletin of Mathematical Biology*, 60(2):329–354, 1998.

[5] M. Aldana and P. Cluzel. A natural class of robust networks. *Proceedings of the National Academy of Sciences of the USA*, 100(15):8710–8714, 2003.

[6] U. Alon. Biological networks: The tinkerer as an engineer. *Science*, 301(5641):1866–1867, 2003.

[7] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, 2006.

[8] U. Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, 2007.

[9] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397(6715):168–171, 1999.

[10] D. Angeli, J. E. Ferrell, and E. D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proceedings of the National Academy of Sciences of the USA*, 101(7):1822–1827, 2004.

[11] D. Angeli and E. Sontag. Oscillations in I/O monotone systems. *IEEE Transactions on Circuits and Systems: Special Issue on Systems Biology*, 55:166–176, 2008.

[12] K. J. Åstrom and R. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2009.

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

[14] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913–917, 1997.

[15] C. J. Bashor, N. C. Helman, S. Yan, and W. A. Lim. Using engineered scaffold interactions to reshape map kinase pathway signaling dynamics. *Science*, 319(5869):1539–1543, 2008.

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

[17] R. P. Bhattacharyya, A. Remenyi, M. C. Good, C. J. Bashor, A. M. Falick, and W. A. Lim. The ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science*, 311(5762):822–826, 2006.

[18] F. Blanchini. Set invariance in control - a survey. *Automatica*, 35(11):1747–1767, 1999.

[19] F. Blanchini and E. Franco. Structurally robust biological networks. *Bio Med Central Systems Biology*, 5(1):74, 2011.

[20] F. Blanchini and S. Miani. *Set-theoretic methods in control*, volume 22 of *Systems & Control: Foundations & Applications*. Birkhäuser, Boston, 2008.

[21] N. E. Buchler and F. R. Cross. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Molecular Systems Biologyl*, 5:272, 2009.

[22] T. R. Cech. Crawling out of the RNA world. *Cell*, 136(4):599–602, 2009.

[23] M. Chamberlin and J. Ring. Characterization of T7-specific ribonucleic-acid polymerase .2. inhibitors of enzyme and their application to study of enzymatic-reaction. *Journal of Biological Chemistry*, 248:2245–2250, 1973.

[24] H.-L. Chen, A. De, and A. Goel. Towards programmable molecular machines. In *Foundations of Nanoscience*, 2008.

[25] T. Danino, O. Mondragon-Palomino, L. Tsimring, and J. Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326–330, 2010.

[26] H. De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology*, 9:67–103, 2002.

[27] D. Del Vecchio, A. Ninfa, and E. Sontag. Modular cell biology: Retroactivity and insulation. *Molecular Systems Biologyl*, 4:161, 2008.

[28] R. M. Dirks, J. S. Bois, J. M. Schaeffer, E. Winfree, and N. Pierce. Thermodynamic analysis of interacting nucleic acid strands. *Society for Industrial and Applied Mathematics Review*, 49:65–88, 2007.

[29] W. U. Dittmer, S. Kempter, J. O. Radler, and F. C. Simmel. Using gene regulation to program DNA-based molecular devices. *Small*, 1:709–712, 2005.

[30] M. J. Doktycz and M. L. Simpson. Nano-enabled synthetic biology. *Molecular Systems Biologyl*, 3(125), 2007.

[31] H. El-Samad, H. Kurata, J. Doyle, C. Gross, and M. Khammash. Surviving heat shock: Control strategies for robustness and performance. *Proceedings of the National Academy of Sciences of the USA*, 102(8):2736–2741, 2005.

[32] H. El-Samad, S. Prajna, A. Papachristodoulou, J. Doyle, and M. Khammash. Advanced methods and algorithms for biological networks analysis. *Proceedings of the IEEE*, 94(4):832–853, 2006.

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

[34] M. Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors I. the deficiency zero and deficiency one theorems. *Chemical Engineering Science*, 42:2229–2268, 1987.

[35] R. Ferrari, C. Rivetti, and G. Dieci. Transcription reinitiation properties of bacteriophage T7 RNA polymerase. *Biochemical and Biophysical Research Communications*, 315:376–380, 2004.

[36] J. E. Ferrell, J. R. Pomerening, S. Y. Kim, N. B. Trunnell, W. Xiong, C.-Y. F. Huang, and E. M. Machleder. Simple, realistic models of complex biological processes: Positive feedback and bistability in a cell fate switch and a cell cycle oscillator. *Federation of European Biochemical Societies letters*, 583(24):3999–4005, 2009.

[37] J. E. J. Ferrell and E. M. Machleder. The Biochemical Basis of an All-or-None Cell Fate Switch in Xenopus Oocytes. *Science*, 280(5365):895–898, 1998.

[38] A. C. Forster and G. M. Church. Synthetic biology projects in vitro. *Genome Research*, 17:1–6, 2007.

[39] E. Franco, D. Del Vecchio, and R. M. Murray. Design of insulating devices for *in vitro* synthetic circuits. In *Proceedings of the IEEE Conference on Decision and Control*, 2009.

[40] E. Franco, P.-O. Forsberg, and R. M. Murray. Design, modeling and synthesis of an *in vitro* transcription rate regulatory circuit. In *Proceedings of the American Control Conference*, 2008.

[41] E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree, and F. C. Simmel. Timing molecular motion and production with a synthetic transcriptional clock. *Proceedings of the National Academy of Sciences*, 108(40):E784–E793, 2011.

[42] E. Franco and R. M. Murray. Design and performance of *in vitro* transcription rate regulatory circuits. In *Proceedings of the IEEE Conference on Decision and Control*, 2008.

[43] A. Friedland, T. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins. Synthetic gene networks that count. *Science*, 324(5931):1199–1202, 2009.

[44] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S.-G. Lee, and J. C. Liao. A synthetic gene-metabolic oscillator. *Nature*, 435:118–122, 2005.

[45] J. Gómez-Gardenes, M. Y., and L. M. Floría. On the robustness of complex heterogeneous gene expression networks. *Biophysical Chemistry*, 115:225–229, 2005.

[46] A. Gorban and O. Radulescu. Dynamical robustness of biological networks with hierarchical distribution of time scales. *IET Systems Biology*, 1(4):238–246, 2007.

[47] H. Gu, J. Chao, S.-J. Xiao, and N. C. Seeman. Dynamic patterning programmed by DNA tiles captured on a DNA origami substrate. *Nature Nanotechnology*, 4(4):245–248, 2009.

[48] C.-Y. F. Huang and J. J. Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proceedings of the National Academy of Sciences of the USA*, 93:10078–10083, 1996.

[49] D. Imburgio, M. Rong, K. Ma, and W. T. McAllister. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry*, 39:10419–10430, 2000.

[50] F. Isaacs, D. Dwyer, C. Ding, D. Pervouchine, C. Cantor, and J. J. Collins. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7):841–847, 2004.

[51] F. J. Isaacs and J. J. Collins. Plug-and-play with RNA. *Nature Biotechnology*, 23(3):306–307, 2005.

[52] M. Isalan, C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao, E. Raineri, M. Garriga-Canut, and L. Serrano. Evolvability and hierarchy in rewired bacterial gene networks. *Nature*, 452(7189):840–845, 2008.

[53] F. Jacob, D. Perrin, C. Sanchéz, and J. Monod. L'opéron: groupe de gènes à expression coordonnée par un opérateur. *Comptes Rendus de l'Académie des Sciences, Paris*, 250:1727–1729, 1960.

[54] J. Jacquez and C. Simon. Qualitative theory of compartmental systems. *Society for Industrial and Applied Mathematics Review*, 35(1):43–79, 1993.

[55] M. Jewett, K. Calhoun, A. Voloshin, J. Wuu, and J. Swartz. An integrated cell-free metabolic platform for protein production and synthetic biology. *Molecular Systems Biologyl*, 4(220), 2008.

[56] Y. Jia and S. S. Patel. Kinetic mechanism of transcription initiation by bacteriophage T7 RNA polymerase. *Biochemistry*, 36(14):4223–4232, 1997.

[57] M. Jiang, M. Rong, C. Martin, and W. T. McAllister. Interrupting the template strand of the T7 promoter facilitates translocation of the DNA during initiation reducing transcript slippage and the release of abortive products. *Journal of Molecular Biology*, 310:509–522, 2001.

[58] O. Kartal and O. Ebenhöh. Ground state robustness as an evolutionary design principle in signaling networks. *Public Library of Science One*, 4(12):e8001, 2009.

[59] R. J. Kershner, L. D. Bozano, C. M. Micheel, A. M. Hung, A. R. Fornof, J. N. Cha, C. T. Rettner, M. Bersani, J. Frommer, P. W. K. Rothemund, and G. M. Wallraff. Placement and orientation of individual DNA shapes on lithographically patterned surfaces. *Nature Nanotechnology*, 4(9):557–561, 2009.

[60] B. N. Kholodenko, A. Kiyatkin, F. J. Bruggeman, E. Sontag, H. V. Westerhoff, and J. B. Hoek. Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proceedings of the National Academy of Sciences of the USA*, 99(20):12841–12846, 2002.

[61] J. Kim. In vitro *synthetic transcriptional networks*. PhD thesis, California Institute of Technology, 2007.

[62] J. Kim, J. J. Hopfield, and E. Winfree. Neural network computation by in vitro transcriptional circuits. In *Advances in Neural Information Processing Systems*, volume 17, pages 681–688, 2004.

[63] J. Kim, K. S. White, and E. Winfree. Construction of an *in vitro* bistable circuit from synthetic transcriptional switches. *Molecular Systems Biologyl*, 1:68, 2006.

[64] J. Kim and E. Winfree. Synthetic *in vitro* transcriptional oscillators. *Molecular Systems Biologyl*, 7:465, 2011.

[65] G. C. King, C. Martin, T. T. Pham, and J. E. Coleman. Transcription by T7 RNA-polymerase is not zinc-dependent and is abolished on amidomethylation of cystein-347. *Biochemistry*, 25:36–40, 1986.

[66] H. Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, 2002.

[67] K. Kurin-Csörgei, I. R. Epstein, and M. Orbán. Systematic design of chemical oscillators using complexation and precipitation equilibria. *Nature*, 433(7022):139–142, 2005.

[68] Y.-K. Kwon and K.-H. Cho. Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics. *Bioinformatics*, 24(7):987–994, 2008.

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

[70] E. Levine, Z. Zhang, T. Kuhlman, and T. Hwa. Quantitative characteristics of gene regulation by small RNA. *Public Library of Science Biology*, 5(9):e229, 2007.

[71] T. Liedl and F. C. Simmel. Switching the conformation of a DNA molecule with a chemical oscillator. *Nano Letters*, 5(10):1894–1898, 2005.

[72] P. L. Luisi, F. Ferri, and P. Stano. Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften*, 93(1):1–13, 2006.

[73] P. L. Luisi and P. Stano. *The Minimal Cell*. Springer, 1^{st} edition, 2011.

[74] K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. N. Stojanovic, N. G. Walter, E. Winfree, and H. Yan. Molecular robots guided by prescriptive landscapes. *Nature*, 465(7295):206–210, 2010.

[75] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang. Defining network topologies that can achieve biochemical adaptation. *Cell*, 138(4):760–773, 2009.

[76] J. Mallet-Paret and H. L. Smith. The poincare-bendixson theorem for monotone cyclic feedback systems. *Journal of Dynamics and Differential Equations*, 2:367–421, 1990.

[77] D. Mandell and T. Kortemme. Computer-aided design of functional protein interactions. *Nature Chemical Biology*, 5(11):797–807, 2009.

[78] S. Mangan, A. Zaslaver, and U. Alon. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of Molecular Biology*, 334:197–204, 2003.

[79] N. I. Markevich, J. B. Hoek, and B. N. Kholodenko. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J. Cell Biol.*, 164(3):353–359, 2004.

[80] M. Markham, N. R. & Zuker. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research*, 33:577–581, 2005.

[81] J. R. McCarrey and A. D. Riggs. Determinator-inhibitor pairs as a mechanism for threshold setting in development: a possible function for pseudogenes. *Proceedings of the National Academy of Sciences of the USA*, 83:679–683, 1986.

[82] K. E. McGinness and G. F. Joyce. Substitution of ribonucleotides in the T7 RNA polymerase promoter element. *Journal of Biological Chemistry*, 277(4):2987–2991, 2002.

[83] P. Mehta, S. Goyal, and N. S. Wingreen. A quantitative comparison of sRNA-based and protein-based gene regulation. *Molecular Systems Biologyl*, 4, 2008.

[84] S. Milburn, M. Goldrick, and M. Winkler. Compositions and methods for increasing the yields of in vitro RNA transcription and other polynucleotide synthetic reactions, U. S. Patent 5256555, 1993.

[85] N. Mitarai, J. Benjamin, S. Krishna, S. Semsey, Z. Csiszovszki, E. Masse, and K. Sneppen. Dynamic features of gene expression control by small regulatory RNAs. *Proceedings of the National Academy of Sciences of the USA*, 106(26):10655–10659, 2009.

[86] K. Montagne, R. Plasson, Y. Sakai, T. Fujii, and Y. Rondelez. Programming an in vitro DNA oscillator using a molecular networking strategy. *Molecular Systems Biologyl*, 7, 2011.

[87] S. Mukherji and A. van Oudenaarden. Synthetic biology: understanding biological design from synthetic circuits. *Nature reviews. Genetics*, 10:859–871, 2009.

[88] M. Nakajima, K. Imai, H. Ito, T. Nishiwaki, Y. Murayama, H. Iwasaki, T. Oyama, and T. Kondo. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science*, 308(5720):414–415, 2005.

[89] S. Nikolov, E. Yankulova, O. Wolkenhauer, and V. Petrov. Principal difference between stability and structural stability (robustness) as used in systems biology. *Nonlinear Dynamics, Psychology, and Life Sciences*, 11(4):413–33, 2007.

[90] V. Noireaux, R. Bar-Ziv, and A. Libchaber. Principles of cell-free genetic circuit assembly. *Proceedings of the National Academy of Sciences of the USA*, 100:12672–12677, 2003.

[91] V. Noireaux and A. Libchaber. A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences of the USA*, 101(51):17669–17674, 2004.

[92] J. A. Papin, N. D. Price, S. J. Wiback, D. A. Fell, and B. O. Palsson. Metabolic pathways in the post-genome era. *Trends in Biochemical Sciences*, 28(5):250–258, 2003.

[93] A. Phillips and L. Cardelli. A programming language for composable DNA circuits. *Journal of The Royal Society Interface*, 2009.

[94] R. J. Prill, P. A. Iglesias, and A. Levchenko. Dynamic properties of network motifs contribute to biological network organization. *Public Library of Science Biology*, 3(11):e343, 2005.

[95] L. Qiao, R. B. Nachbar, I. G. Kevrekidis, and S. Y. Shvartsman. Bistability and Oscillations in the Huang-Ferrell Model of MAPK Signaling. *Public Library of Science Computational Biology*, 3(9):e184, 2007.

[96] X. Qin, M. Byrne, T. Mori, P. Zou, D. R. Williams, H. Mchaourab, and C. H. Johnson. Intermolecular associations determine the dynamics of the circadian KaiABC oscillator. *Proceedings of the National Academy of Sciences of the USA*, 107(33):14805–14810, 2010.

[97] L. Quian and E. Winfree. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 3:1196–1201, 2011.

[98] N. Radde, N. Bar, and M. Banaji. Graphical methods for analysing feedback in biological networks - a survey. *International Journal of Systems Science*, 41(1):35–46, 2010.

[99] P. Rapp. Analysis of biochemical phase-shift oscillators by a harmonic balancing technique. *Journal Of Mathematical Biology*, 3(3 - 4):203–224, 1976.

[100] D.-K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong, and J. D. Keasling. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–943, 2006.

[101] N. Rosenfeld, M. B. Elowitz, and U. Alon. Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology*, 323:785–793, 2002.

[102] P. W. K. Rothemund. Folding DNA to create nanoscale shapes and paterns. *Nature*, 440:297–302, 2006.

[103] N. Rouche, P. Habets, and M. Laloy. *Stability theory by Liapunov's direct method*, volume 22 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1977.

[104] C. Russo, C. Giuraniuc, R. Blossey, and J.-F. Bodart. On the equilibria of the MAPK cascade: Cooperativity, modularity and bistability. *Physica A: Statistical Mechanics and Its Applications*, 388(24):5070 – 5080, 2009.

[105] G. Russo, M. di Bernardo, and E. D. Sontag. Global entrainment of transcriptional systems to periodic inputs. *Public Library of Science Computational Biology*, 6(4):e1000739, 2010.

[106] J. Saez-Rodriguez, A. Kremling, and E. Gilles. Dissecting the puzzle of life: modularization of signal transduction networks. *Computers & Chemical Engineering*, 29(3):619–629, 2005.

[107] S. D. M. Santos, P. J. Verveer, and P. I. H. Bastiaens. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nature Cell Biology*, 9(3):324–330, 2007.

[108] M. A. Savageau. Design of molecular control mechanisms and the demand for gene expression. *Proceedings of the National Academy of Sciences of the USA*, 74:5647–5651, 1977.

[109] L. Scardovi, M. Arcak, and E. Sontag. Synchronization of interconnected systems with applications to biochemical networks: An input-output approach. *IEEE Transactions on Automatic Control*, 55:1367–1379, 2010.

[110] G. Seelig, D. Soloveichik, D. Y. Zhang, and E. Winfree. Enzyme-free nucleic acid logic circuits. *Science*, 314(5805):1585–1588, 2006.

[111] G. Seelig, B. Yurke, and E. Winfree. DNA hybridization catalysts and catalyst circuits. *DNA Computing*, 3384:329–343, 2005.

[112] Y. Shimizu. Cell-free translation reconstituted with purified components. *Nature biotechnology*, 19(8):751–755, 2001.

[113] Y. Shimizu. Cell-free translation systems for protein engineering. *The Federation of European Biochemical Societies journal*, 273(18):4133–4140, 2006.

[114] Y. Shimizu and U. Takuya. PURE technology. In J. M. Walker, editor, *Methods in Molecular Biology*, volume 607, pages 11–21. Springer-Verlag, NY, 2010.

[115] G. Shinar and M. Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327(5971):1389–1391, 2010.

[116] G. Shinar, R. Milo, M. Rodrìguez Martìnez, and U. Alon. Input-output robustness in simple bacterial signaling systems. *Proceedings of the National Academy of Sciences of the USA*, 104:19931–199935, 2007.

[117] O. Shoval, L. Goentoro, Y. Hart, A. Mayo, E. Sontag, and U. Alon. Fold-change detection and scalar symmetry of sensory input fields. *Proceedings of the National Academy of Sciences of the USA*, 107(36):15995–16000, 2010.

[118] F. Simmel and W. U. Dittmer. DNA nanodevices. *Small*, 1(3):284–299, 2005.

[119] M. L. Simpson. Cell-free synthetic biology: a bottom-up approach to discovery by design. *Molecular Systems Biologyl*, 2:69, 2006.

[120] D. Soloveichik, G. Seelig, and E. Winfree. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences of the USA*, 2010.

[121] E. Sontag. Monotone and near-monotone biochemical networks. *Systems and Synthetic Biology*, 1:59–87, 2007.

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

[123] C. Sturk, E. Franco, and R. M. Murray. Tuning a synthetic in vitro oscillator using control-theoretic tools. In *Proceedings of the IEEE Conference on Decision and Control*, 2010.

[124] P. Subsoontorn, J. Kim, and E. Winfree. Bistability of an in vitro synthetic autoregulatory switch. http://arxiv.org/abs/1101.0723, 2011.

[125] J. W. Szostak, D. P. Bartel, and P. L. Luisi. Synthesizing life. *Nature*, 409:387–390, 2001.

[126] M. Takinoue, D. Kiga, K. Shohda, and A. Suyama. Experiments and simulation models of a basic computation element of an autonomous molecular computing system. *Physical Review E*, 78(041921), 2008.

[127] T. Tian. Robustness of mathematical models for biological systems. *ANZIAM J.*, 45:C565–C577, 2004.

[128] M. Tigges, T. T. Marquez-Lago, J. Stelling, and M. Fussenegger. A tunable synthetic mammalian oscillator. *Nature*, 457(7227):309–312, 2009.

[129] T. Y.-C. Tsai, Y. S. Choi, W. Ma, J. R. Pomerening, C. Tang, and J. E. J. Ferrell. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science*, 321(5885):126–129, 2008.

[130] J. M. G. Vilar, C. Guet, and S. Leibler. Modeling network dynamics: the lac operon, a case study. *Journal of Cell Biology*, 161(3):471–476, 2003.

[131] D. M. Widmaier, D. Tullman-Ercek, E. A. Mirsky, R. Hill, S. Govindarajan, J. Minshull, and C. A. Voigt. Engineering the salmonella type III secretion system to export spider silk monomers. *Molecular Systems Biologyl*, 5:–, 2009.

[132] T. Williamson, J.-M. Schwartz, D. B. Kell, and L. Stateva. Deterministic mathematical models of the cAMP pathway in saccharomyces cerevisiae. *Bio Med Central Systems Biology*, 3, 2009.

[133] A. T. Winfree. *The Geometry of Biological Time*. Springer-Verlag, New York, NY, 1980.

[134] T. M. Yi, Y. Huang, M. I. Simon, and J. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences of the USA*, 97(9):4649–4653, 2000.

[135] N. Yildirim and M. Mackey. Feedback regulation in the lactose operon: A mathematical modeling study and comparison with experimental data. *Biophysical Journal*, 84(5):2841–2851, 2003.

[136] N. Yildirim, M. Santillan, D. Horike, and M. Mackey. Dynamics and bistability in a reduced model of the *lac* operon. *Chaos*, 14(2):279–292, 2004.

[137] P. Yin, H. M. T. Choi, C. R. Calvert, and N. A. Pierce. Programming biomolecular self-assembly pathways. *Nature*, 451(7176):318–322, 2008.

[138] B. Yurke and A. P. Mills. Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines*, 4:111–122, 2003.

[139] B. Yurke, A. J. Turberfield, A. P. Mills Jr, F. C. Simmel, and J. L. Neumann. A DNA-fuelled molecular machine made of DNA. *Nature*, 406:605–608, 2000.

[140] D. Y. Zhang, A. J. Turberfield, B. Yurke, and E. Winfree. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 318:1121–1125, 2007.

[141] M. Zuker and P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1):133–148, 1981.