

Modular design and analysis of synthetic biochemical networks

Citation for published version (APA):

Roekel, van, H. W. H. (2015). *Modular design and analysis of synthetic biochemical networks*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Biomedical Engineering]. Technische Universiteit Eindhoven.

Document status and date:

Published: 10/12/2015

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Modular Design and Analysis of Synthetic Biochemical Networks

A catalogue record is available from the Eindhoven University of Technology Library

ISBN: 978-90-386-3974-1

Copyright © 2015 by Rik van Roekel

All rights are reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the author.

Cover Design: ICMS Animation Studio, TU/e

Printed by: Ipskamp Drukkers, Enschede

This work was financially supported by The Netherlands Organisation for Scientific Research (NWO): Graduate Programme 2010

Modular Design and Analysis of Synthetic Biochemical Networks

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Eindhoven,
op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens,
voor een commissie aangewezen door het College voor Promoties,
in het openbaar te verdedigen
op donderdag 10 december 2015 om 16:00 uur

door

Hendrik Wouter Hubertus van Roekel

geboren te Venlo

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

| | |
|--------------------------|------------------------------|
| voorzitter: | prof.dr. J.P.W. Pluim |
| 1 ^e promotor: | prof.dr. P.A.J. Hilbers |
| copromotor(en): | dr.ir. T.F.A. de Greef |
| leden: | prof.dr. W.T.S. Huck (RUN) |
| | prof.dr. P.R. ten Wolde (VU) |
| | prof.dr.ir. M.W.J. Prins |
| | prof.dr. M.E.J. Karttunen |
| | prof.dr. E.W. Meijer |

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

TABLE OF CONTENTS

Chapter 1 – An Introduction to Programmable Chemical Reaction Networks

| | |
|--|----|
| ABSTRACT | 2 |
| 1.1 INTRODUCTION | 3 |
| 1.2 HOMOGENEOUS INORGANIC REACTION NETWORKS | 4 |
| 1.3 TEMPORAL BIOCHEMICAL REACTION NETWORKS | 8 |
| 1.4 SPATIALLY EXTENDED BIOCHEMICAL REACTION NETWORKS | 15 |
| 1.5 EMERGING APPLICATIONS | 22 |
| 1.6 OUTLINE OF THIS THESIS | 24 |
| 1.7 REFERENCES | 25 |

Chapter 2 – Automated Design of Programmable Enzyme-Driven DNA Circuits

| | |
|---|----|
| ABSTRACT | 30 |
| 2.1 INTRODUCTION | 31 |
| 2.2 METHODS | 33 |
| 2.2.1 General | 33 |
| 2.2.2 ODEs, sampling procedure and response classification | 35 |
| 2.2.3 Robustness analysis | 40 |
| 2.2.4 Automated DNA sequence design | 43 |
| 2.3 RESULTS AND DISCUSSION | 46 |
| 2.3.1 Automated Design of Programmable PEN Toolbox Networks | 46 |
| 2.3.2 Robustness Analysis and Automated DNA Sequence Design | 49 |
| 2.4 CONCLUSION | 52 |
| 2.5 REFERENCES | 53 |

Chapter 3 – Adaptation Dynamics in a PEN-Based Incoherent Feedforward Loop

| | |
|---|----|
| ABSTRACT | 58 |
| 3.1 INTRODUCTION | 59 |
| 3.2 MATERIALS AND METHODS | 60 |
| 3.2.1 Experimental characterization of the PEN toolbox-based IFFL circuit | 60 |

| | |
|---|----|
| 3.2.2 Incorporation of unproductive sequestration of exonuclease by protected ssDNA strands into the IFFL model | 62 |
| 3.2.3 Toy models to analyze the effect of unproductive sequestration of exonuclease | 63 |
| 3.3 RESULTS AND DISCUSSION | 66 |
| 3.3.1 Enzyme kinetics | 66 |
| 3.3.2 The extended model | 70 |
| 3.3.3 Analysis of adaptation dynamics in the toy models | 73 |
| 3.4 CONCLUSION | 75 |
| 3.5 REFERENCES | 77 |

Chapter 4 – Design and Analysis of a Tunable Oscillating Enzymatic Network

| | |
|---|----|
| ABSTRACT | 80 |
| 4.1 INTRODUCTION | 81 |
| 4.2 MATERIALS AND METHODS | 82 |
| 4.2.1 The synthetic network, derivation of the model and parameter sampling | 82 |
| 4.2.2 Experimental data and model optimization | 87 |
| 4.3 RESULTS AND DISCUSSION | 90 |
| 4.3.1 Sampling of the network | 90 |
| 4.3.2 Optimization of the model with time-course data | 91 |
| 4.3.3 Profile likelihood | 95 |
| 4.3.4 Tunability | 97 |
| 4.4 CONCLUSION | 98 |
| 4.5 REFERENCES | 98 |

Chapter 5 – Evaporative Self-Assembly of Single-Chain, Polymeric Nanoparticles

| | |
|----------------------------------|-----|
| ABSTRACT | 102 |
| 5.1 INTRODUCTION | 103 |
| 5.2 MATERIALS AND METHODS | 104 |
| 5.2.1 Experimental | 104 |
| 5.2.2 Computational | 106 |
| 5.2.3 Image analysis | 110 |
| 5.3 RESULTS AND DISCUSSION | 112 |

| | |
|----------------------|-----|
| 5.4 REFERENCES | 117 |
|----------------------|-----|

Epilogue – General Discussion, Ethics and Outlook

| | |
|-------------------------------|-----|
| 6.1 GENERAL DISCUSSION | 120 |
| 6.2 ETHICAL ASPECTS | 121 |
| 6.3 FUTURE PERSPECTIVES | 123 |
| 6.4 REFERENCES | 123 |

Appendix – Supplement to the Chapters

| | |
|--|-----|
| A.1 MODEL EQUATIONS, PARAMETER VALUES, SAMPLING RANGES AND NUPACK SETTINGS CORRESPONDING TO CHAPTER 2 | 126 |
| A.2 MODEL EQUATIONS AND DERIVATIONS CORRESPONDING TO CHAPTER 3 | 138 |
| A.3 DERIVATIONS CORRESPONDING TO CHAPTER 4 | 139 |

| | |
|--------------------------------|-----|
| SUMMARY | 142 |
| NEDERLANDSE SAMENVATTING | 144 |
| CURRICULUM VITAE | 146 |
| LIST OF PUBLICATIONS | 147 |
| DANKWOORD | 148 |

GLOSSARY OF ABBREVIATIONS

| | |
|-------------|--|
| AFM | Atomic force microscopy |
| BRN | Biochemical reaction network |
| BZ reaction | Belousov-Zhabotinsky reaction |
| CRN | Chemical reaction network |
| CSTR | Continuously stirred tank reactor |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxynucleotide triphosphate |
| dsDNA | Double-stranded DNA |
| IFFL | Incoherent feedforward loop |
| IFFLP | Incoherent feedforward loop with a proportioner node |
| MC | Monte Carlo |
| NFBLB | Negative feedback loop with a buffer node |
| ODE | Ordinary differential equation |
| PEN toolbox | Polymerase-exonuclease-nickase toolbox |
| RNA | Ribonucleic acid |
| SCPN | Single-chain polymeric nanoparticle |
| ssDNA | Single-stranded DNA |
| TMI | Toy model I |
| TMII | Toy model II |
| TxTl | Transcription-translation |

Chapter 1

An Introduction to Programmable Chemical Reaction Networks

This work has been published:

Hendrik W. H. van Roekel*, Bas J. H. M. Rosier*, Lenny H. H. Meijer*, Peter A. J. Hilbers, Albert J. Markvoort, Wilhelm T. S. Huck and Tom F. A. de Greef (2015), Programmable chemical reaction networks: emulating regulatory functions in living cells using a bottom-up approach. *Chemical Society Reviews* **44**, 7465–7483.

* Authors contributed equally

ABSTRACT

Living cells are able to produce a wide variety of biological responses when subjected to biochemical stimuli. It has become apparent that these biological responses are regulated by complex chemical reaction networks (CRNs). Unraveling the function of these circuits is a key topic of both systems biology and synthetic biology. Recent progress at the interface of chemistry and biology together with the realization that current experimental tools are insufficient to quantitatively understand the molecular logic of pathways inside living cells has triggered renewed interest in the bottom-up development of CRNs. This builds upon earlier work of physical chemists who extensively studied inorganic CRNs and showed how a system of chemical reactions can give rise to complex spatiotemporal responses such as oscillations and pattern formation. Using purified biochemical components, *in vitro* synthetic biologists have started to engineer simplified model systems with the goal of mimicking biological responses of intracellular circuits. Emulation and reconstruction of system-level properties of intracellular networks using simplified circuits are able to reveal key design principles and molecular programs that underlie the biological function of interest. In this introductory Chapter, we present an accessible overview of this emerging field starting with key studies on inorganic CRNs followed by a discussion of recent work involving purified biochemical components. Finally, we review recent work showing the versatility of programmable biochemical reaction networks (BRNs) in analytical and diagnostic applications.

1.1 INTRODUCTION

Cells are complex biochemical systems sustained by an underlying network of dynamic regulatory circuits and signal transduction pathways which are ultimately responsible for the basic functionality and robustness of life¹⁻³. These biochemical networks are composed of thousands of chemical species that interact *via* nonlinear kinetic processes and dissipate chemical energy to achieve their function. In the last decade, impressive work in systems biology has revealed some of the design principles correlating the topology of biochemical circuits to the range of biological responses that can be produced by these networks^{4,5}. These studies have shown that chemical self-organization—the process by which a dissipative (bio)chemical system of several components becomes ordered in space and time—is a key factor determining the output of intracellular circuits⁶⁻⁸. Unfortunately, delineating the exact relation between biochemical network topology and the range of accessible spatiotemporal behaviors in living cells is not possible in many cases due to a limitation of current experimental tools.

One viable strategy to disentangle the complexity of cellular signaling networks is by applying a bottom-up approach, consisting of the construction of cell-free circuits that are able to display self-organizing behaviors such as signal transduction, oscillations, self-replication, entrainment and programmed pattern formation (Figure 1.1). Chemical self-organization has been studied since the discovery of inorganic oscillators, such as the Belousov-Zhabotinsky (BZ) reaction in the 1960s. In the previous century, impressive work in this field led to a detailed understanding of the connection between feedback, diffusion and the emergence of complex self-organized behavior such as chemical patterns and the ability of chemical systems to display higher-order dynamics such as oscillations and bistability⁹. These early studies on chemical self-organization in relatively simple physicochemical systems have provided a new generation of physicists, chemists, and cell biologists with a comprehensive framework to tackle one of the key questions in science: to understand how living cells process information and control their global organization using hierarchical layers of biochemical networks^{4,6}.

These studies on self-organization resulted in the rise of two related and partially overlapping fields, *i.e.* systems chemistry¹⁰ and *in vitro* synthetic biology^{11,12}. While systems chemistry focuses on the emulation of complex emergent phenomena encountered in living systems utilizing a framework that is based on mostly abiotic homogeneous chemistry, *in vitro* synthetic biology is concerned with the bottom-up engineering of new and existing life-like functionalities using biological components. Both fields apply a bottom-up methodology to recapitulate the dynamic organization of living cells by replacing biology's complex and idiosyncratic regulatory networks by simpler and more predictable molecular circuitry. One of the key goals of *in vitro* synthetic biology is to reduce specific spatiotemporal functions of regulatory circuits in living cells to simple biochemical model systems. Using this approach, the relevant characteristics and interactions that determine the specific system-level behavior of interest can be isolated and, importantly, mathematically described in a framework of simple theories. In addition, novel self-organization principles employed by cells to receive, transmit and process stimuli into complex biochemical and physiological outputs can be discovered by stepwise increasing the complexity of these reductionist biochemical model systems.

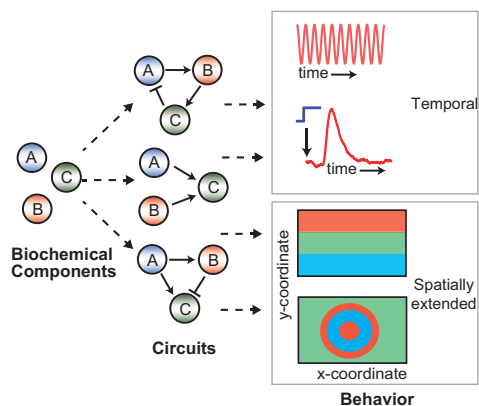


Figure 1.1 Programmable cell-free biomolecular circuits built from modular biochemical components present a unique and uncharted platform to systematically explore the molecular logic and physical design principles of regulatory networks in the living cell. Complex (spatio)temporal behavior exhibited by such circuits are based on elementary regulations such as activation and inhibition.

enzymatic systems that display a wide range of self-organized temporal behaviors under well-stirred conditions such as bistability, oscillations and adaptation. However, in the intracellular environment, diffusion and crowding are fundamental drivers of spatiotemporal ordering. Therefore, we discuss several studies where programmed pattern formation and spatial information processing has been achieved using *in vitro* biomolecular systems. Finally, we review emerging applications of *in vitro* synthetic biology in the area of cell-free biosensing and diagnostics and materials science.

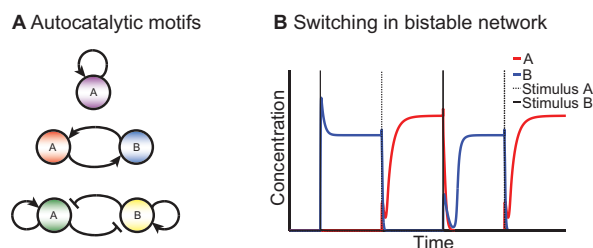


Figure 1.2 Autocatalysis and bistability. (A) Common autocatalytic motifs, with (top) direct autocatalysis, (middle) indirect autocatalysis and (bottom) mutual inhibition with autocatalysis. (B) An autocatalytic network that can exist in two steady states, *i.e.* one where A is dominant (red) and one where B is dominant (blue). Switching between these states is inducible by external stimuli.

1.2 HOMOGENEOUS INORGANIC REACTION NETWORKS

Homogeneous chemical reaction networks (CRNs) that display dynamic self-organization by coupling of reaction and diffusion have been studied since the second half of the 20th century. In CRNs coupling of feedback and diffusion can give rise to new behaviors. For example, autocatalysis—the process in which a reactant directly or indirectly drives the formation of itself—gives rise to global bistability in a well-stirred environment, while propagating fronts develop when the reaction is carried out in a quasi-two-dimensional unstirred setup. A system exhibits bistability when it is able to exist in two stable steady-states. Figure 1.2A displays several topologies of autocatalytic networks while Figure 1.2B displays bistability which can be triggered into one of the two possible steady-states using a stimulus. When appropriate negative feedback is introduced, the well-stirred setup becomes either excitable (*i.e.* an input above a certain threshold leads to a spike in the output before returning to the initial steady state) or oscillatory. In a spatially extended system, this leads to a single propagating pulse or a wave train of constant velocity, respectively. The BZ reaction was the first example of a class of inorganic chemical reactions that shows such out-of-equilibrium phenomena, *e.g.* bistability, and also oscillations and synchronization (*vide infra*) that are visible to the naked eye⁹. Oscillations are signals in which temporal variations about a central

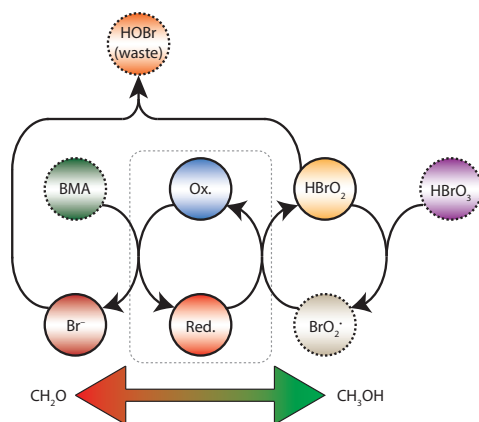


Figure 1.3 Key steps in the BZ reaction. The oxidation of the metal-ion catalyst (middle, enclosed with dashed box) is accompanied by conversion of BrO_2^\bullet to HBrO_2 resulting in an autocatalytic formation of HBrO_2 (right). The reduction of the metal-ion catalyst (accompanied by the oxidation of BMA, left), produces Br^- that reacts with HBrO_2 , serving as inhibition of the autocatalytic species under formation of waste product HOBr . The BZ reaction can be excited in a so-called activator-controlled mode by methanol, or in an inhibitor-controlled mode by formaldehyde, as these species slightly shift the balance to autocatalysis of HBrO_2 or Br^- production, respectively.

value occur with well-defined frequencies. The basic mechanism responsible for oscillations in the BZ reaction is the autocatalytic production of bromous acid (HBrO_2) that is inhibited by bromide (Br^-) via a delayed negative feedback mechanism (Figure 1.3). This network motif is commonly encountered in other chemical and biochemical oscillators as well (*vide infra*). Moreover, the BZ reaction and equivalent inorganic CRNs are excitable by external stimuli. In the following, we discuss the BZ reaction and similar excitable inorganic reaction networks that can be triggered to display higher-order spatiotemporal behavior, both in batch reactors and open chemostats. Furthermore, we highlight efforts to synchronise inorganic oscillators in space and time functionalities.

Excitability of the BZ reaction was first noted by Zaikin and Zhabotinsky in 1970¹³, when they found that a BZ reaction mixture dispersed as a thin layer in a Petri dish can be triggered into generating spatiotemporal waves originating from a single point called a pacemaker by perturbation of the medium. In an attempt to unravel the physicochemical principles that determine wave initiation, Bishop and Grzybowski¹⁴ used wet stamping to deliver chemical stimuli from complex geometries to a gel containing the BZ reagents. This setup resulted in excitation of the system in a so-called activator-controlled mode, or inhibitor-controlled mode. Depending on the triggering agent present in the star-shaped stamp, *i.e.* formaldehyde for inhibitor-controlled mode (Figure 1.4A) and methanol for activator-controlled mode (Figure 1.4B), the autocatalytic step is initiated either in the star's tips or between the star's arms, resulting in focused wave emission. This phenomenon is explained by a slight shift in the balance between the activating and inhibiting perturbation, which is established by localized control of the autocatalytic process mainly by delivery or removal of the inhibitor species into or from the oscillatory region by diffusion. These diffusion processes are affected by the curvature of the star-shaped stamp and as a result, changing the triggering agent in the star directly affects the mode of wave emission from the star by inducing a spatiotemporal interplay between reaction kinetics and diffusion.

Rather than studying CRNs in batch reactors, the investigation of CRNs in an open-system setup prevents the ultimate decay of the reaction mixture towards thermodynamic equilibrium. Noszticzius *et al.*¹⁵ studied the BZ reaction in an open-system with an experimental set-up that allows two solutions with BZ reactants to come into contact in an annular gel, resulting in initiation of the oscillatory reaction in the ring. Embedded with a pacemaker that triggers the reaction, and a wave barrier that attenuates counter clockwise waves, the system is instructed to generate waves

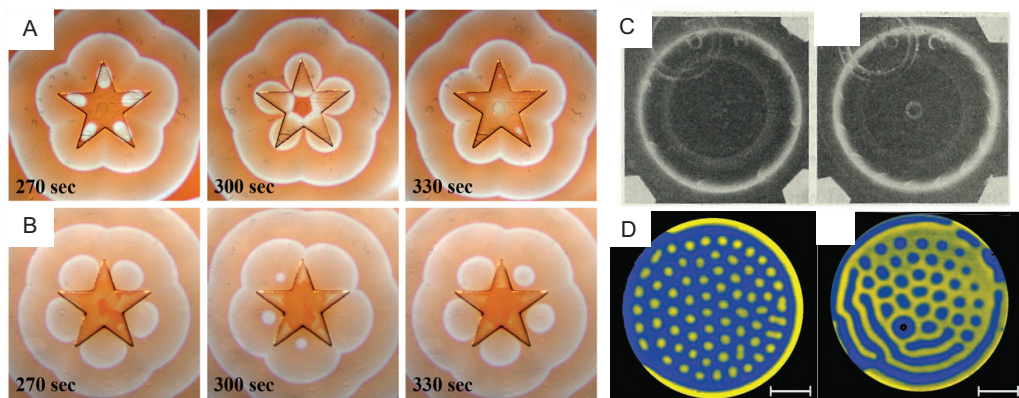


Figure 1.4 Excitability and spatiotemporal pattern formation in inorganic CRNs. (A) Inhibitor-controlled pattern formation and (B) activator-controlled pattern formation in the BZ reaction by a star-shaped stamp. Image taken from ref. 14. (C) Equidistant spatiotemporal waves resembling chemical pinwheels on an annular gel in which the BZ reaction takes place. Waves with a larger gap between them travel faster (left) than waves that are closer together (right). Image taken from ref. 15. (D) Turing pattern formation in uncomplexed protons (yellow: low pH, blue: high pH) for high inhibitor concentration (left) and low inhibitor concentration (right). Image taken from ref. 16.

with a defined frequency and direction. The resulting pattern resembles a pinwheel, Figure 1.4C, as the distance between subsequent waves becomes constant. This study presents one of the first realizations of an open reactor in which the BZ reaction is excited into generating stable chemical patterns in time.

Open-system studies were taken to the next pioneering level in 2009, when De Kepper and co-workers¹⁶ employed the iodate-sulfite reaction, which autocatalytically produces protons, to generate Turing-like patterns. The authors accomplished this by coupling the autocatalytic production of protons to a negative feedback on this autocatalytic loop by oxidation of an inhibitor species. Short-range spatiotemporal activation was realized by diffusive control of protons *via* reverse complexation to high-molecular weight ions, impairing their movement with respect to diffusion of the inhibitor. The resulting differential diffusivity of the inhibitor and the autoactivating species ensures long-range inhibition. A disc shaped spatial reactor takes up chemicals *via* diffusion on one face of the disc, resulting in pattern formation in the bulk of the disc. If the inhibitor exceeds a preset critical concentration, the inorganic CRN displays spatiotemporal self-organization that is characteristic of a Turing bifurcation (the hexagonal patterns in Figure 1.4D) in terms of localized low pH values. This study shows the possibility of higher-order spatiotemporal organization by chemical reactions as a result of the judicious interplay of mass-transfer, reaction rates and differential diffusivity of the components paving the way for future efforts aimed at engineering life-like biomolecular systems. Diffusion also plays an important organizational role on the μm scale in systems that operate with biological components (Section 1.4).

Model systems to investigate synchronization by diffusive coupling were recently employed by the groups of Epstein¹⁷ and Showalter¹⁸. Synchronization is the process by which independent oscillators interact and assume the same temporal frequency. Epstein *et al.* investigated the communication between BZ droplets in oil *via* diffusive coupling by bromine employing a microfluidic set-up. The droplets showed out-of-phase synchronization (Figure 1.5A) and stationary

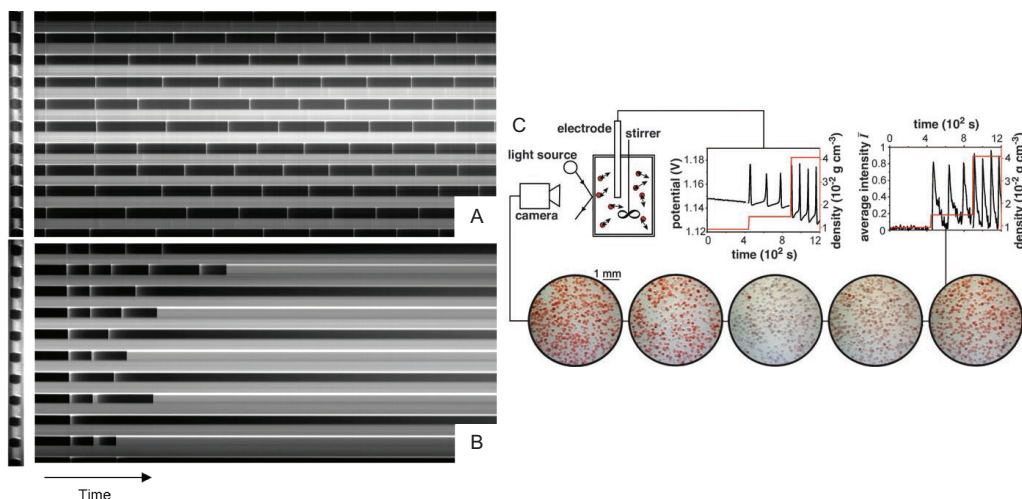


Figure 1.5 Chemical synchronization in the BZ reaction. (A) Space-time plot showing out-of-phase oscillations in a microfluidic set-up containing BZ droplets. (B) Space-time plot displaying stationary Turing patterns where the oxidized and reduced alternately appear from an initial oscillatory state. Image taken from ref. 17. (C) Experimental set-up for detecting quorum-sensing in a population of microparticles at high exchange rate: oscillations appear after a critical particle density. The bottom shows images of a single oscillation, visible as a color change of the BZ catalyst immobilized on the beads. Image taken from ref. 18.

Turing structures (Figure 1.5B), depending on the droplet size and the initial concentration of malonic acid. The authors identified by experiment and by simulation that the inhibitor bromine is mainly responsible for communication between the droplets, as its hydrophobic character allows it to move between compartments. Therefore, bromine is responsible for the resulting synchronization between droplets. Using a different approach, Taylor and Showalter *et al.* embedded microparticles with immobilized BZ catalyst and suspended the particles in a BZ reaction mixture lacking the catalyst. The experimental setup is displayed in Figure 1.5C, along with a series of images showing a visibly changing synchronized population of microparticles during a single oscillation. The exchange rate of activator and inhibitor among the microparticles was controlled by the stirring rate of the suspension. At high exchange rates, the microparticles remain quiescent at low particle densities, and above a certain critical particle density, synchronized oscillatory behavior is observed in a near step-wise fashion (graphs in Figure 1.5C), reminiscent of the dynamics of quorum sensing. At low exchange rates, there is a linear increase of synchronization with particle density, indicating that weak coupling is the main mechanism of information exchange. The exchange of chemical information between BZ droplets or microparticles concerning their oscillatory state serves as a model system to rationalize global coupling of intracellular oscillators by intercellular cell-to-cell communication.

The studies discussed thus far provide key insights into the establishment of complex spatiotemporal dynamics *via* a bottom-up approach. We now stress the notion that inorganic CRNs can serve as useful model systems or even emulators for biological systems in several aspects. Firstly, all types of inorganic reactions discussed here are homogeneous reactions that maintain their far-from-equilibrium thermodynamic state for a significant amount of time, making them excellent functional model systems for nonequilibrium biological behavior. Analogously, intracellular circuits

use high-energy molecules to maintain steady-state transcription and translation rates. Secondly, biochemical synchronization between living systems is observed in cell populations, *i.e.* quorum-sensing, which can be emulated by a suspension of microparticles. Finally, inorganic CRNs can be employed for information processing¹⁹ and pattern storage²⁰ by excitable chemical systems, which could provide renewed insights on memory storage and decision making in intracellular networks where critical events such as cell-cycle progression and cellular differentiation are regulated by bistable network modules.

1.3 TEMPORAL BIOCHEMICAL REACTION NETWORKS

Recently, complex out-of-equilibrium behavior has been established using biological macromolecules rather than small inorganic chemicals. This shift in approach allows the use of ambient conditions rather than the harsh acidic environments characteristic of the BZ reaction. In addition, CRNs based on biological molecules more closely resemble regulatory networks *in vivo*. Many pioneering studies conducted are concerned with the *in vitro* reconstitution of existing intracellular networks, such as oscillatory phosphorylation of KaiC by synergistic action of KaiA and KaiB²¹. Indeed, reconstruction of system-level properties of regulatory circuits using purified biochemical components can reveal key design principles and molecular programs that underlie the biological function of interest²². However, the focus of this Thesis is on bottom-up rational design strategies to build cell-free model systems *de novo*. Therefore, we will focus on biochemical reaction networks (BRNs) that display various higher-order temporal functionalities including switching and oscillations. We limit the discussion to enzymatically driven systems, *i.e.* those that rely on enzymes to maintain their thermodynamic nonequilibrium state. While non-enzymatic biomolecular networks are highly programmable and useful synthetic biological tools, intracellular networks use enzymatic reactions to achieve their function.

In the previous section, bistable CRNs were briefly mentioned as core memory units. Indeed, living systems harness bi- and multistable molecular programs to store information concerning the state of a cell, and subsequently process this information to induce decision-making events. Kim, Winfree and co-workers²³ succeeded in the bottom-up design of an *in vitro* bistable enzymatically driven circuit based on synthetic DNA templates known as genelets that are regulated by RNA transcripts. The researchers developed the so-called genelet toolbox in which RNA polymerase (RNAP) transcribes RNA regulators from genelets that in turn serve as excitatory or inhibitory regulators for the transcription of RNA regulators from other genelets. In this way, the elimination of genes and their associated production and degradation of proteins that are not involved in the circuit allows for systematic analysis by means of a fairly accurate quantitative description that is impossible to achieve in *in vivo* systems or crude cell extracts. In Figure 1.6A, the implementation of a genelet-based bistable switch is shown that consists of transcriptionally active and inactive double-stranded DNA (dsDNA) species and interacting RNA regulators. RNA regulators serve as inhibitors I_1 and I_2 of transcription from genelets Sw_{21} and Sw_{12} , respectively, by binding to the respective single-stranded DNA (ssDNA) activators A_1 and A_2 . Sequestration of activator DNA prevents formation of fully dsDNA resulting in a significant reduction of transcriptional activity. These reactions are designed to be very specific by appropriate choice of the base sequence. Activation is accomplished by degradation of RNA transcripts bound to the activator ($A_i I_i$) resulting in the release of the sequestered DNA activator. Figure 1.6B shows the corresponding topology of

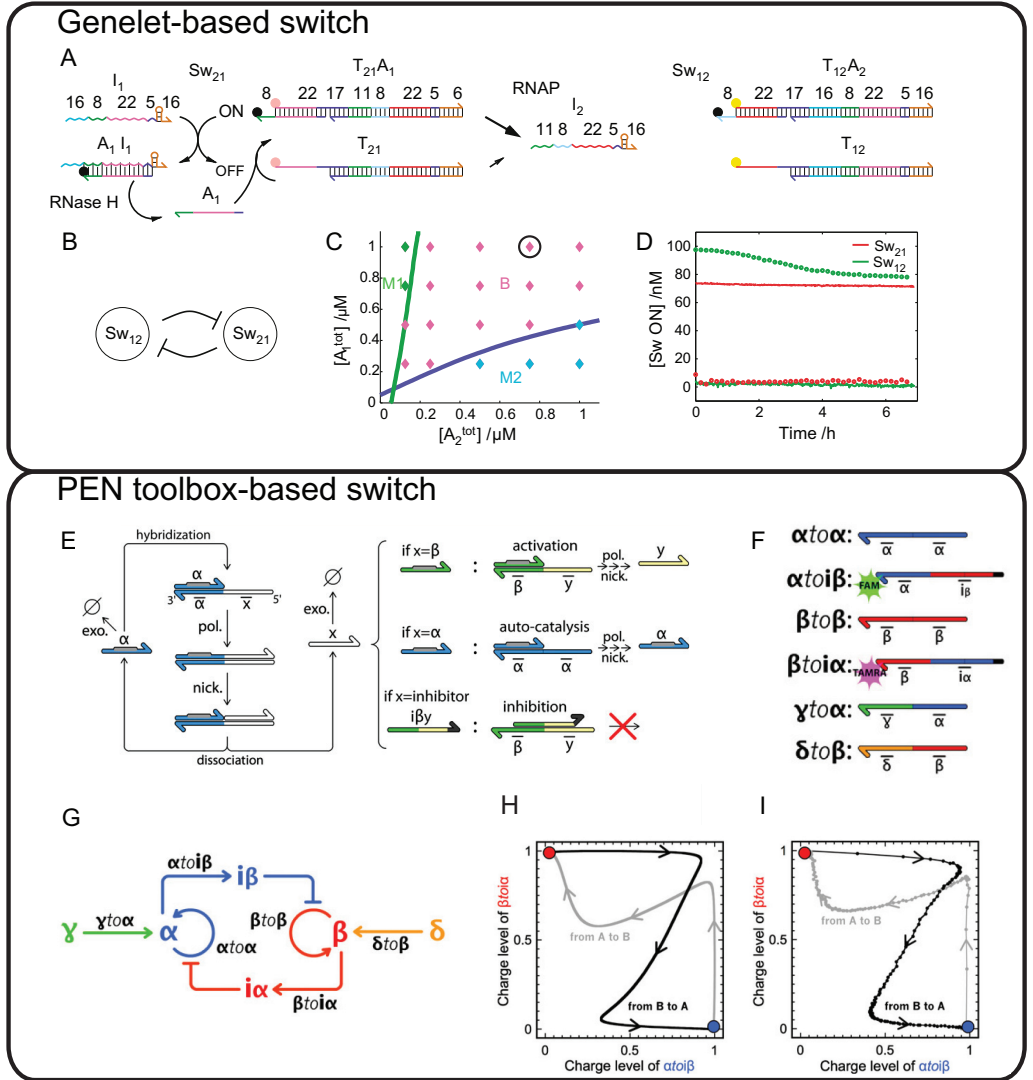


Figure 1.6 *In vitro* synthetic bistable switches. The top panel shows a genelet-based switch, and the bottom panel a polymerase-exonuclease-nickase (PEN) toolbox-based switch. (A) Implementation of the genelet-based switch based on synthetic template DNA from which RNA polymerase (RNAP) produces RNA regulators that in turn serve as excitatory or inhibitory regulators for the transcription of RNA regulators from other genelets, while RNase H degrades regulators from $A_i I_i$ complexes. (B) Topology of the reaction scheme in A. (C) Experimentally constructed bifurcation diagram, with pink data points corresponding to bistable behavior. (D) Switch activity time courses corresponding to the encircled data point in C, with high initial concentration of RNA inhibitor I_1 (circles) or high initial concentration of RNA inhibitor I_2 (lines). Image taken from ref. 23. (E) Reactions in the PEN toolbox: activator ssDNA α primes the synthetic template, after which it is extended by DNA polymerase resulting in dsDNA. Nickase cleaves this dsDNA in the backbone of the newly formed strand, resulting in dissociation of signal strand X . Depending on the sequence of X , three types of reactions can occur, *i.e.* activation, autocatalysis and inhibition. (F) The six synthetic DNA templates of the bistable switch shown in G. (G) Topological representation of the PEN toolbox-based switch. The single-stranded input signals γ and δ induce switching to the α and β state, respectively. (H) Simulated predictions of switching to the β state by δ (grey line) and of switching to the α state by γ (black line).

(I) A single switching experiment starting in the α state, which is first switched to the β state by δ (grey line), and subsequently switched back to the α state by γ (black line). Image taken from ref. 25.

the bistable BRN, while Figure 1.6C displays the experimentally constructed bifurcation diagram with the two activator ssDNA concentrations (A_2^{tot} for Sw_{12} and A_1^{tot} for Sw_{21}) as bifurcation parameters. Figure 1.6D displays the switch activity time courses corresponding to high initial concentration of RNA inhibitor I_1 (circles) or high initial concentration of RNA inhibitor I_2 (lines). The system remains in the steady state that the investigators initiated, showing that two possible steady states exist. Further experiments to test the bistable circuit on hysteresis suggest that switching between steady states in a single experiment is possible at least once. The characterization of individual genelets proved sufficient for the authors to assemble a bistable circuit, based on modularity and programmability of the components. The modularity of the genelet-toolbox have also allowed Kim, Murray and colleagues to assemble a synthetic transcriptional network that shows adaptation and fold change detection based on the incoherent feedforward loop (IFFL)²⁴.

Rondelez and co-workers²⁵ used a similar strategy to construct a bistable toggle switch, by introducing a modular toolbox by which biochemical networks of arbitrary complexity can be engineered *de novo* under well-controlled *in vitro* conditions. The toolbox is based on DNA and the enzymes polymerase, exonuclease and nickase (PEN toolbox). Synthetic DNA templates are triggered by short ssDNA strands acting as regulating signals on the production of an output signal. Activation of templates by these ssDNA signals enables DNA polymerase to extend the oligomer-template pair resulting in fully hybridized dsDNA. Subsequently, these dsDNAs are nicked resulting in double-stranded complexes (Figure 1.6E). Because these reactions are performed at a temperature close to the melting temperature of the partial duplexes, the input and output ssDNAs dissociate from their templates. Templates are inhibited by ssDNA strands that are complementary to part of the template's sequence, lack a nickase recognition site and possess a two-base mismatch at their 3' ends which prevents extension of the partial duplexes, rendering the template strands inactive. Signal and inhibition strands are degraded over time by exonuclease from which the template strands are protected by phosphorothioate modifications at their 5' ends. Figure 1.6F shows the six templates that comprise the toggle switch. Two templates, $\alpha\text{to}\alpha$ and $\beta\text{to}\beta$ autocatalytically produce ssDNA species α and β while templates $\alpha\text{to}\beta$ and $\beta\text{to}\alpha$ produce inhibitors upon activation by α and β respectively. These four templates comprise the bistable core (Figure 1.6G). Positive feedback is necessary in this system as reciprocal inhibition alone does not provide the ultrasensitive response necessary for bistability. To independently switch the system between the α and β state two additional templates, *i.e.* $\gamma\text{to}\alpha$ and $\delta\text{to}\beta$ were employed (Figure 1.6F-G). The authors computationally predicted switching to the β state by addition of single-stranded δ and switching to the α state by addition of single-stranded γ (Figure 1.6H) and verified this bistable behavior of the circuit experimentally (Figure 1.6I). The circuit could be switched two consecutive times in a single experiment after which resources were depleted and the concentrations of templates became too low, leading to loss of functionality of the switch. Finally, the authors presented the construction of a push-push memory circuit that responds to a single input that switches the state of the bistable core. Due to the size of the circuit (eight template strands), the reaction times of the push-push circuit are large and the circuit could only be switched once. Nonetheless, the size of the circuit is comparable to *in vivo* complexity of eight genes, which is amongst the largest realizations of synthetic genetic regulatory networks performed in cells. Moreover, the systematic engineering of

the push-push circuit further demonstrates the modularity of the PEN toolbox. To even further predict *in silico* the dynamical behavior of *in vitro* PEN-based reaction networks, we developed an automated approach in which PEN-based networks of arbitrary complexity were screened for target temporal behavior (Chapter 2)²⁶.

As explicitly pointed out in the previous section, oscillations are a canonical example of out-of-equilibrium behavior. Recently, several *in vitro* biochemical circuits exhibiting oscillations have been reported. Here, we highlight several pioneering studies in which oscillatory dynamics have been engineered in enzymatically driven networks, DNA circuits and *in vitro* transcription-translation-based systems, both under batch conditions and in open reactors. Examples of oscillating BRNs that function under batch conditions include cell-free *in vitro* biochemical oscillators based on the genelet toolbox (Figure 1.7A) and the PEN toolbox (Figure 1.7B).

Kim and Winfree constructed an oscillator using the genelet toolbox, which consists of two synthetic DNA templates that comprise a delayed negative feedback loop²⁷. Two genelets, regulated by RNA activator rA_1 and RNA inhibitor rI_2 transcribed from genelet templates Sw_{12} and Sw_{21} respectively, govern the oscillatory behavior. The circuit could show up to five complete cycles. The researchers then added a positive feedback loop with the use of a third genelet that modulated and extended the oscillatory regime and, importantly, demonstrated the modularity of this system. To further explore the modularity and resilience of the system, Murray, Simmel and co-workers coupled the transcriptional oscillator to several downstream loads including a DNA-based nanomechanical device and the production of a functional RNA molecule (Figure 1.7A)²⁸. The coupling of dynamic networks to downstream loads affects the dynamics of the original system, a process which is known as retroactivity and should be minimized. Different mechanisms were employed to couple the oscillator to several downstream processes, and for each mechanism the impact of the load on the performance of the upstream ‘core oscillator’ was assessed on key characteristics such as the frequency and amplitude of the oscillatory response. Reduction of retroactivity and efficient activation of the downstream reaction circuit was achieved when the coupled process was driven by appropriately fast kinetics in combination with a high-amplitude oscillatory signal. The authors achieved near zero retroactivity with the introduction of a genelet module that functions as an insulator. The insulator is activated by DNA strand A_2 and inhibited by RNA strand rI_2 , which are utilized and produced by the core oscillator, respectively (Figure 1.7A). The insulator acts as an amplifier because small increases in the active genelet (only mildly affecting the core oscillator) drive transcription of large amounts of RNA transcript *InsOut* that drives the opening of a DNA tweezer. Degradation of *InsOut* occurs fast enough for the tweezer to follow the oscillations of the core BRN by periodically opening and closing. Therefore, the insulator module not only demonstrates the modularity of the genelet toolbox but also greatly improves the utility of *in vitro* BRNs as insulation of the core circuit allows these networks to be utilized to drive functional downstream processes, for example in DNA-based diagnostics.

In an additional study, Simmel and co-workers compartmentalized the *in vitro* transcriptional oscillator into microdroplets to study the effect of micro-scale encapsulation on *in vitro* biochemical reaction networks with complex, far-from-equilibrium dynamical behavior²⁹. Contrary to the study on compartmentalized BZ droplets¹⁷, biochemical reactions occur at low concentrations and, hence, show fluctuations and variability reflected by stochastic kinetics. To study these effects, the

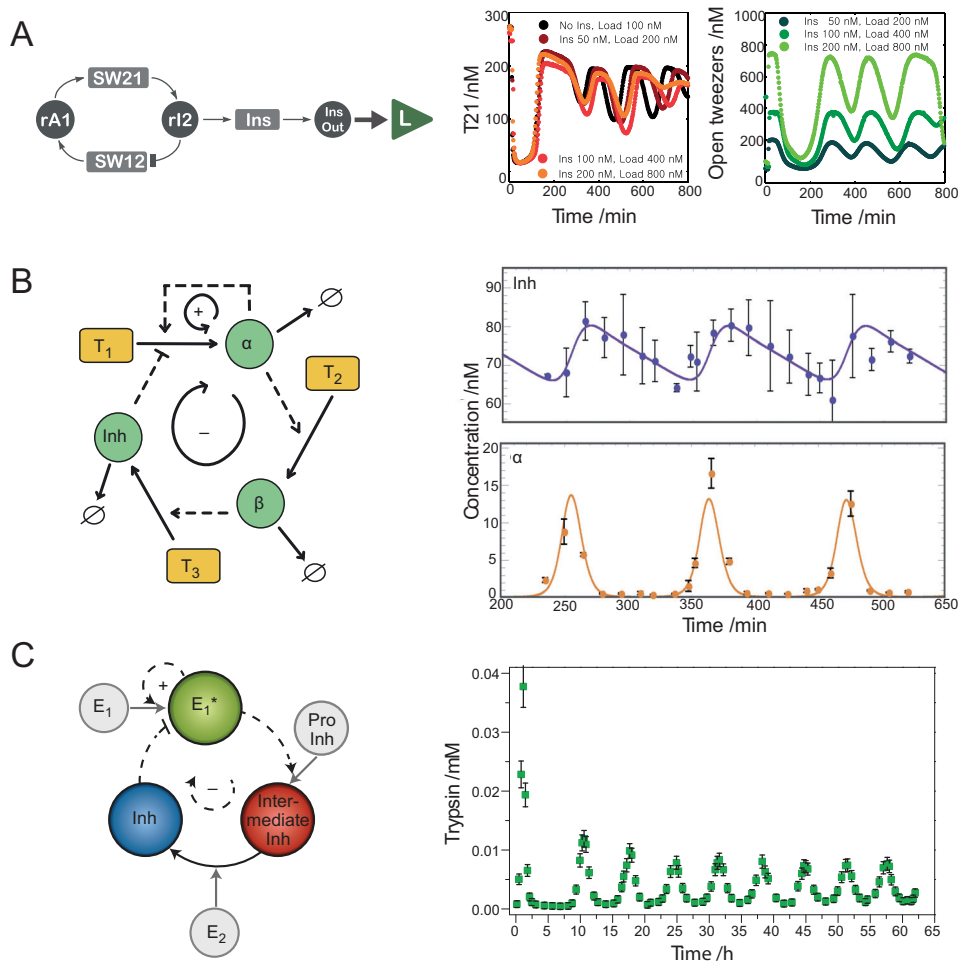


Figure 1.7 Oscillatory networks designed *de novo*. (A) Schematic representation of a genelet oscillator coupled downstream to a load via an insulator module (left). Experimental traces of the transcriptional oscillator coupled to the downstream process shows the insensitivity of the system towards the load for various insulator and DNA tweezer concentrations. Image taken from ref. 28. (B) Schematic representation of the PEN-based oscillator, including autocatalysis, activation, inhibition and degradation of primers and single-stranded DNA strands (left). The oscillatory traces of single-stranded species from experiments (dots) are shown on the right and can be fully described by a computational model (line). Image taken from ref. 30. (C) Schematic representation of the trypsin-based enzymatic oscillator (left). Experimental traces show sustained oscillations in the concentration of trypsin for approximately 65 hours. Image taken from ref. 32.

robustness of the *in vitro* transcriptional oscillator was assessed at low concentrations by measuring large populations of microdroplets simultaneously. In contrast to the work of Epstein¹⁷, synchronization among compartmentalized oscillators did not occur due to a lack of controlled communication between droplets, resulting in dynamical diversity of oscillatory behavior among droplets in terms of amplitude, frequency and damping. Based on results of deterministic and stochastic models the researchers concluded that the dynamical diversity in oscillatory behavior

could be attributed to the statistical variation of reactant concentrations as a result of partitioning the system into droplets, rather than reflecting the intrinsic stochastic kinetics of the chemical reaction network itself. This extrinsic noise as a result of statistical variations during encapsulation of biochemical components is an important parameter when engineering artificial cells with preprogrammed temporal dynamics.

A PEN toolbox-based oscillator was constructed by Rondelez and co-workers (Figure 1.7B)³⁰, in an analogous manner as described for their bistable switch. The network topology consists of autocatalytic synthesis of primer α which activates synthesis of primer β which in turn serves as a primer for the production of inhibitor *inh*. Delayed inhibition of the autocatalytic synthesis of primer α results from the reversible sequestration of oligomer *inh* to T_1 . Highlighting the versatility of the methodology, the same group designed an *in vitro* enzymatic DNA-based network that emulated the complex dynamics of ecological systems³¹, including predator-prey oscillations, competition-induced chaos, and symbiotic synchronization.

Several groups have successfully engineered oscillating BRNs in open chemostats. Huck *et al.* constructed a trypsin-based enzymatic oscillator under flow conditions and provided a modular approach to rationally design dissipative, tunable and robust chemical networks³². The methodology is based on translation of a basic network topology known to generate oscillations, *i.e.* a short positive feedback loop in combination with a delayed negative feedback mechanism. Positive feedback is implemented by autocatalytic production of trypsin (E_1^* , Figure 1.7C) from trypsinogen (E_1), while delayed inhibition is realized by a masked inhibitor that is converted into a fully active inhibitor by an enzymatic cascade based on trypsin and aminopeptidase. As the two enzymes that make up the negative feedback loop show a high selectivity towards their substrates, the rate of each step could be independently tuned. Using mathematical modeling, the correct balance between production and inactivation of trypsin necessary for sustained oscillatory behavior was established resulting in sustained oscillations in the experimental concentration of trypsin for a wide range of flow rates (Chapter 4). Furthermore, the authors show signal modulation by coupling of two biochemical networks each in a separate fluidic reactor, and by coupling of the out-of-equilibrium enzymatic network to complex coacervates.

In vitro transcription and translation chemistry (TxTl) allows the engineering of biomolecular circuits based on transfer of information between genes. Because the transcription and translation rates of *in vitro* TxTl systems decrease in time under batch conditions, experiments need to be conducted in a continuous flow setup allowing constant steady-state transcription and translation rate by exchange of reagents. Maerkl *et al.* constructed and implemented a genetic oscillator using pneumatically controlled microfluidic reactors with volumes in the nanoliter regime³³. Importantly, the genetic oscillator was subjected to continuous flow allowing exchange of reagents and forcing the system out-of-equilibrium resulting in sustained oscillations for extended periods of time (Figure 1.8A). The circuit design combines a short range positive feedback loop with delayed repression. The positive feedback loop is based on an autogene consisting of a gene encoding for T3 RNA polymerase and its cognate T3 promoter. Inhibition of the autogene is achieved by production of TetR repressor which binds to TetR operator sites on the T3 promoter of the autogene. Because production of TetR repressor has to be delayed to achieve sustained oscillations, the authors introduced an Amber stop codon in the *tetR* gene resulting in premature termination of *tetR* mRNA

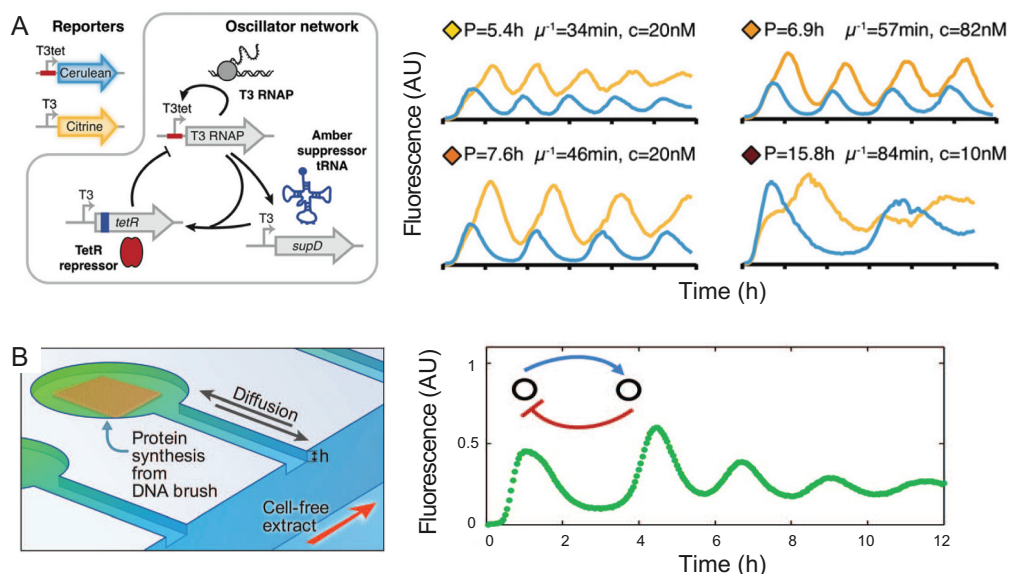


Figure 1.8 *In vitro* oscillators based on transcription and translation chemistry. (A) Operation scheme of the genetic oscillator consisting of an autocatalytic loop and delayed inhibition (left). The dynamic state of the system is measured using Cerulean and Citrine which are produced in parallel to T3RNAP and TetR repressor respectively. Oscillatory traces of Cerulean and Citrine show varying profiles and periods (P) at different *supD* gene concentrations (c) and different residence times (μ^{-1}) (right). Image taken from ref. 33. (B) Left: Design of the microfluidic chip in which DNA brushes are assembled in circular compartments and are connected through a capillary to a feed channel. Activator-repressor genetic networks are assembled with sigma factor for activation and the lambda phage cI for repression. Right: GFP profile of oscillatory gene dynamics in an activator-repressor network measured in the compartment. Image taken from ref. 34.

translation. The translation rate can be increased by production of Amber suppressor tRNA which is in turn under control of the T3 autogene. Inspection of the phase diagram of the *in vitro* TxTI oscillator with respect to the *supD* DNA template concentrations and the dilution rate reveals that oscillations are observed for almost all concentrations of *supD* template. Furthermore, it reveals a decrease in the oscillation period with increasing dilution rate with oscillatory behavior generally being established at high dilution rates.

Showing the generality of *in vitro* TxTI systems as a tool to engineer programmable molecular circuits, Bar-Ziv, Noireaux and colleagues assembled artificial cells based on two-dimensional compartments containing DNA-brushes. Programmable protein synthesis is achieved by diffusion of *Escherichia coli* cell extract through thin capillaries that connect the compartments to a central feeding channel (Figure 1.8B)³⁴. Oscillatory dynamics of the genetic network is accomplished by a mixed DNA-brush encoding for an activator-repressor gene network in each compartment. The temporal period in the activator concentration scaled linearly with the length of the capillary channel, demonstrating the effect of diffusion and compartment geometry on oscillatory dynamics under non-stirred conditions. Moreover, by separating DNA-brushes encoding activator and repressor proteins into two connected compartments, the researchers demonstrated diffusion-based information transfer resulting in a spatiotemporal pulse with a period that scales linearly with the distance between the respective compartments. The above studies illustrate enzymatic and nucleic acid-based reaction networks that display complex high-order temporal functionalities, including

switching and oscillations, using a minimal set of components. Clearly, *in vitro* synthetic biologists are capable of emulating temporal life-like out-of-equilibrium dynamics using a relatively small number of biomolecular components.

1.4 SPATIALLY EXTENDED BIOCHEMICAL REACTION NETWORKS

In the previous section, several enzymatic BRNs that display emergent temporal functionalities, such as oscillations and bistability were discussed. While the displayed out-of-equilibrium behavior under well-stirred conditions has provided key insights into the fundamental design principles behind their function, the coupling of reactions with diffusion in spatially extended systems is essential to understand the complex spatial ordering and dynamic behavior of the intra- and intercellular environment. Indeed, reaction-diffusion mechanisms are responsible for a wide range of important regulatory functions *in vivo*, such as neuron polarization, spatial organization during mitosis and morphogenesis^{6,7}.

In this section, we discuss *in vitro* biochemical model systems that display higher-order spatiotemporal behavior when subjected to spatially extended or spatially modulated conditions. Herein, we mainly consider enzymatic systems, where positive or negative feedback is introduced *via* autocatalysis or inhibition. We also highlight the use of molecular programs based on nucleic acids to generate spatial patterns.

Many of these phenomena can be modeled and understood with simple sets of differential equations, often using a Michaelis-Menten description of the reaction kinetics. For example, the well-known FitzHugh-Nagumo model was developed as a model for neuron excitability, and is now used to study a wide range of biochemical reaction-diffusion systems. In fact, by combining these systems with accurate kinetic models, they can also be utilized to estimate physicochemical parameters of biochemical reactions, as demonstrated elegantly by Grzybowski and co-workers³⁵. Using micropatterned agarose stamps to introduce freely-diffusing enzymes into a polyacrylamide hydrogel containing immobilized ligands, the authors were able to estimate protein-ligand dissociation constants down to the picomolar range by fitting the experimental reaction-diffusion patterns to computational models. To engineer higher-order spatial behavior in enzymatic reaction networks, feedback is required and, in the simplest case, can be generated by an intrinsic property of the enzyme itself. Taylor and co-workers³⁶ investigated the urea-urease reaction, in which urease catalyses the conversion of urea to CO₂ under production of base, NH₃. Because the enzyme-substrate complex exists in an active protonated form and an inactive biprotonated form, a bell-shaped rate-pH dependence arises with a maximal reaction rate at pH 7. Therefore, if the initial pH of the base-producing reaction is adjusted to a pH below 7, the reaction accelerates as it progresses. This autocatalytic behavior was first investigated in a continuously stirred tank reactor (CSTR), in which a solution of urea and urease produced a sharp switch from pH 3 to pH 9 after an induction period of approximately 1000 s. Next, a quasi-two-dimensional setup was constructed and a propagating pH front with constant velocity could be observed after initiation of the system (Figure 1.9A). Both the front velocity and the number of spontaneous initiations increased at higher urease concentrations. Simulations, assuming Michaelis-Menten kinetics and including substrate and product inhibition, could reproduce the experimental observations and showed that the diffusion of basic products drives front propagation, while the sharpness of the front is maintained throughout the reaction (Figure 1.9B).

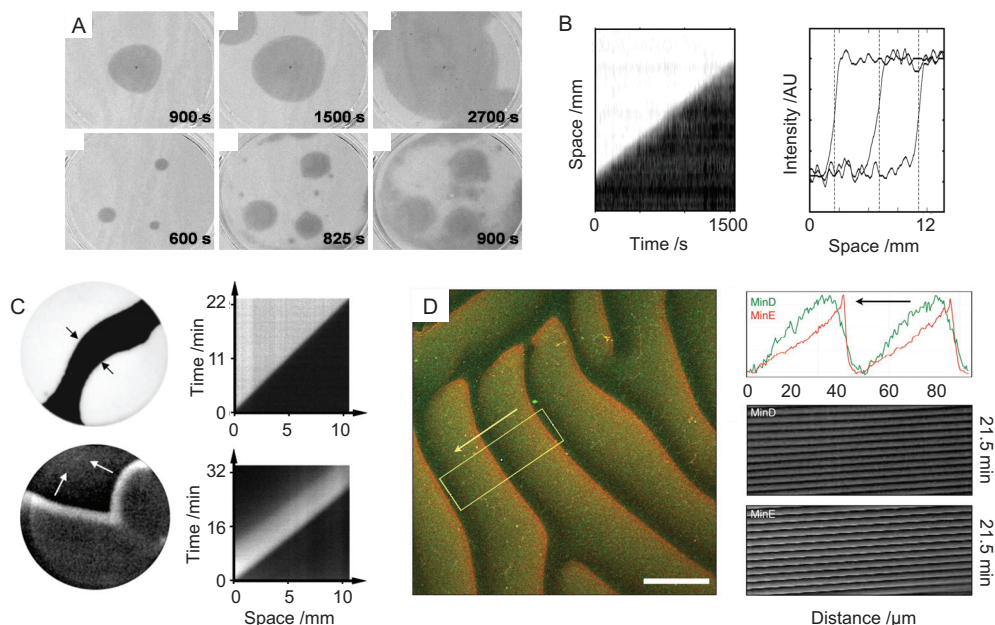


Figure 1.9 Spatiotemporal patterns in enzymatic systems. (A) Propagating pH waves of constant velocity in the urea-urease reaction. Waves travel radially outwards and are visualized using pH indicator cresol red at high (dark) and low (light) pH. (B) Kymograph and intensity-space plots showing constant velocity and front sharpness. Image taken from ref. 36. (C) Propagating pH fronts (top) and pulses (bottom) in the glucose oxidase reaction using a quasi-2D setup (left). Corresponding kymographs displaying constant velocity of the wave front and pulse, respectively (right). Colors correspond to presence of ferrocyanide (dark) and ferricyanide (light). Image taken from ref. 37. (D) Confocal fluorescent images of surface waves caused by MinD and MinE, labeled with BODIPY (green) and Alexa647 (orange) fluorophores, respectively. Graph depicts intensity profile in the area indicated in the image, while kymographs show constant velocity and wavelength of propagating waves. Image taken from ref. 38.

A related single-enzyme system was explored by Epstein and co-workers³⁷, who employed the ping-pong mechanism of glucose oxidase as a source of autocatalysis. Glucose oxidase (GOx) in oxidized form catalyses the oxidation of glucose to glucono- δ -lactone, while the reduced form of GOx reduces ferricyanide to ferrocyanide, producing two protons and returning to its original oxidized state. Because the GOx catalytic rate constant is pH-dependent and protons are produced, the overall enzymatic reaction displays autocatalysis, as evident by an accelerating decrease in pH from pH 7 to 3 as the reaction progresses. Similar to the work of Taylor and colleagues³⁶, a propagating pH front with constant velocity develops in thin slabs of polyacrylamide gel, both in a quasi-1D setup and a quasi-2D continuously fed unstirred reactor (Figure 1.9C).

Subsequently, negative feedback was introduced using H_2O_2 , which restores the enzymatically reduced ferrocyanide into ferricyanide, raising the pH and giving rise to a propagating pulse with constant velocity (Figure 1.9C). Both experimentally and by simulations, it was shown that tuning the H_2O_2 concentration, led to either spatial bistability or excitability in the form of a propagating pulse. Importantly, kinetic simulations revealed that the velocity of the wave front is proportional to the square root of both the substrate and enzyme concentration. This relationship is often encountered in the general theory of chemical waves and is known as the Fisher-Kolmogorov relation. Because the urea-urease reaction and the GOx system involve only a single enzyme and an

intrinsic feedback mechanism, they represent some of the simplest enzymatic systems capable of spatial pattern formation, such as propagating waves and pulses in an unstirred reactor. This simplicity can be used in a bottom-up fashion to understand the fundamental mechanisms behind higher-order behavior observed in complex natural systems.

A more advanced system was reported by the group of Schwille³⁸, who investigated the *in vitro* behavior of the Min protein system, responsible for determining the position of the division septum at the centre of bacterial cells during mitosis. The adenosine triphosphatase MinD binds to a lipid membrane as a dimer, while MinE forms a complex with MinD to induce ATP hydrolysis, releasing both enzymes from the membrane. *In vivo*, a MinD patch forms at one of the cell's poles, while MinE forms a ring around the MinD patch, leading to disassembly of MinD and reassembly of a new patch at the other cell pole. This oscillating behavior results in a time-averaged concentration minimum of MinD in the middle of the cell. The distribution of a third enzyme MinC directly maps onto the MinD pattern and inhibits the membrane attachment of FtsZ, whose polymerization is responsible for the initiation of division septum formation at the bacterial cell poles. The authors reconstituted purified MinD and MinE with ATP *in vitro* in the presence of a supported lipid bilayer and observed spontaneous symmetry breaking and initiation of oscillating waves of constant velocity on the lipid surface (Figure 1.9D). The average front velocity increased while the average wavelength decreased upon increasing MinE concentration, similar to the studies discussed above^{36,37}. Importantly, confocal images of the spatial patterns of fluorescently-labeled MinD and MinE revealed a small lag between the sharp trailing edge of MinD and the trailing edge of MinE (Figure 1.9D). Similar to *in vivo* observations, this indicates MinE complexation with MinD, followed by subsequent release of the complex from the membrane. MinD then reassembles onto the membrane behind the trailing edge, resulting in the build-up of shallow MinD and MinE gradients as leading edges of the next oscillation.

To understand the mechanism by which the surface waves are generated, a simple reaction-diffusion model was used which could qualitatively reproduce the shape of the surface waves by taking into account cooperative effects during MinD and MinE binding. However, the wavelength and the sawtooth-like appearance of—and the phase difference between—MinD and MinE waveforms could not be well reproduced. Meanwhile, theoretical studies have provided a more accurate mechanism that is based on the polymerization of MinD-MinE heterotetramers³⁹. Nonetheless, the approach elegantly shows the application of a relatively simple, dissipative two-enzyme system *in vitro*, displaying complex spatiotemporal behavior that resembles the *in vivo* situation.

The previous studies illustrate how simple enzymatic systems containing few components can display various higher-order functionalities when starting from homogeneous initial conditions. However, cells are often exposed to time-varying signals, such as gradients and oscillations, and spatial patterns. The modulation of a spatiotemporal input signal to produce a relevant intracellular output is an important feature cells use to sense and accordingly react to their environment. This is for example apparent in the ultrasensitive or switch-like response to chemical gradients in chemotaxis or signal transduction pathways.

Huck, De Greef and co-workers⁴⁰ investigated the possibility of spatial modulation using a bottom-up strategy, starting with a simple one-enzyme system consisting of the protease β -trypsin, a

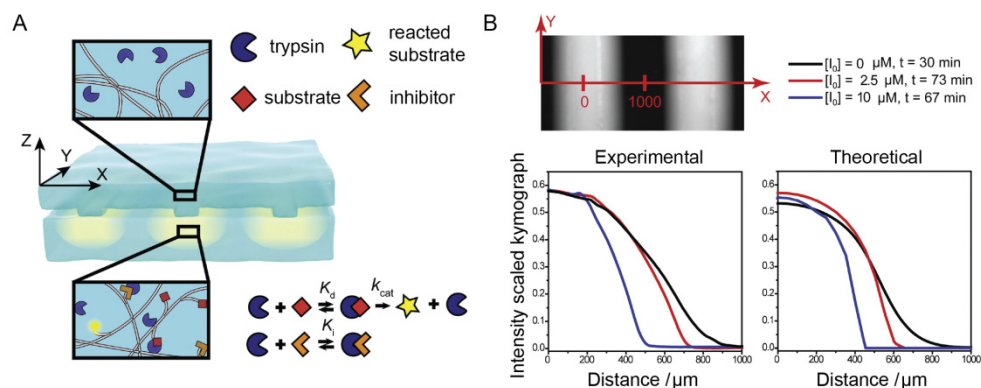


Figure 1.10 Ultrasensitivity through molecular titration using a strong, reversible inhibitor. (A) Experimental setup of wet stamping experiments. (B) Intensity profiles in the polyacrylamide gel at various inhibitor concentrations, in experiments (left) and simulations (right). Image taken from ref. 40.

fluorogenic substrate that becomes fluorescent upon cleavage, and a strong, reversible inhibitor. Similar to the work of Grzybowski *et al.*,³⁵ fluorogenic substrate and inhibitor were immobilized in a polyacrylamide hydrogel, while freely-diffusing trypsin was introduced using micropatterned agarose stamps (Figure 1.10A). The immobilization of the fluorogenic substrate is important because it allows the direct read-out of trypsin activity in the hydrogel. In the absence of inhibitor, trypsin propagates through the hydrogel *via* diffusion and generates a broad product front which overlaps with the propagating trypsin front. Upon addition of a strong, reversible inhibitor the system displays ultrasensitivity by spatially extended molecular titration. The molecular titration effect is based on the fact that the inhibitor binds more strongly to the enzyme compared to its substrate resulting in sequestration of enzyme activity until the total enzyme concentration surpasses the inhibitor concentration. This results in a sharp increase in product conversion and as a result, the spatial propagation of enzyme activity is modulated, significantly sharpening the product front (Figure 1.10B). A system of partial differential equations, taking into account enzymatic activity, inhibition and diffusion of the reaction components, was solved on a representative two-dimensional geometry (Figure 1.10B). The simulations corresponded well with the experimental data and revealed how the steepness of product fronts relates to the concentration and diffusion constant of the inhibitor.

A similar approach was explored by Mrksich and colleagues⁴¹, using a system based on Abelson tyrosine kinase (Abl), which contains a catalytic domain that phosphorylates tyrosine residues and a Sarcoma homology 2 (SH2) domain that binds the phosphorylated substrate. Due to local proximity effects as a result of binding of the SH2 domain to the product, the enzymatic reaction proceeds autocatalytically on a self-assembled monolayer (SAM) consisting of immobilized peptide substrates (Figure 1.11A). By patterning the SAM with regions of substrate peptide and product phosphopeptide followed by exposing the surface to a uniform solution of Abl kinase and ATP, the reaction spatially propagates towards regions of unreacted substrate. Subsequently, patterns with different boundary-to-surface ratios were fabricated and the space-averaged initial reaction rate was measured. The authors concluded that the overall rate increases

molecular titration to develop a system which is able to sense the spatial distribution of enzymes. Extending the work by de Greef & Huck *et al.* (*vide supra*)⁴⁰, positive feedback was added to the trypsin-inhibitor system by employing trypsinogen, a zymogen that is autocatalytically cleaved into trypsin. A biochemical analog-to-digital converter was engineered by introducing trypsin *via* a micropatterned agarose stamp to a polyacrylamide gel containing immobilized fluorogenic substrate and freely-diffusing trypsinogen and inhibitor (Figure 1.11C). Depending on the width of the embossed pillars, the trypsin diffusion front either adopts a radial or one-dimensional character. For narrow pillars, diffusion of trypsin into the gel occurs from a point-source and the inhibitor can completely suppress the autocatalytic activation of trypsinogen. In contrast, wider pillars generate a one-dimensional diffusion front with a smaller boundary-to-surface ratio resulting in a failure of the inhibitor to inactivate trypsin, thus inducing a traveling wave of trypsin activity throughout the gel (Figure 1.11C). This experiment exhibits threshold sensing, *i.e.* the combination of analog input signals (trypsin concentration, pillar width and diffusion time) is transformed into a digital output, *i.e.* propagation or complete inhibition of enzyme activity depending on a threshold level of the input. Next, the authors extended the system with a second polyacrylamide gel layer that only contains freely-diffusing trypsinogen and inhibitor and demonstrated that such a system is capable of responding to the density of pillars on the agarose stamp (Figure 1.11D).

Isalan and co-workers⁴³ demonstrated the power of *in vitro* synthetic biology by engineering a minimal biochemical system capable of generating expression patterns analogous to the complex gene system responsible for *Drosophila* embryonic development. Spatiotemporal control over gene activity is essential for embryonic development and involves concentration gradients of a large number of morphogens that activate or repress transcriptional and translational activity. Specifically, the Bicoid morphogen determines the anterior-posterior orientation of the embryo through spatial concentration gradients from the cell's pole and affects members of the gap gene system in a concentration-dependent manner to induce pattern-formation. In the experimental setup, magnets were used to spatially localize paramagnetic beads coated with the DNA of three genes A, B and C, encoding for zinc finger-based transcription repressor proteins (Figure 1.12A-C). Gene A represses gene B and C and is located at the edges of the reaction chamber, while gene B (repressing gene C) and gene C are present ubiquitously in the setup. T7 polymerase activates expression of gene A and B and is injected at the edges of the chamber to generate a spatial concentration gradient, emulating the function of Bicoid. Finally, Gene C is activated by homogeneously distributed SP6 polymerase that serves as a nonexplicit model for other interactions in the gap gene system (Figure 1.12A). As a result, patterns of expression levels are formed in the reaction chamber (Figure 1.12D). Gene A and B are both activated by T7 polymerase at the chamber edges and therefore display low expression levels in the chamber's centre. Because gene B is repressed by protein A, the expression of protein B at the edges is suppressed. Expression of protein C is exclusively seen in the middle of the chamber, due to repression by proteins A and B. Although significantly simplified, this system shows that complex natural behavior such as transcriptional pattern-formation can arise in a rationally designed *in vitro* system constructed from minimal components.

The above studies clearly illustrate how simple enzymatic reaction networks are combined with spatial patterning or delivery of reaction components to modulate spatiotemporal input patterns. However, modularity and programmability of these networks is generally low, because the rational

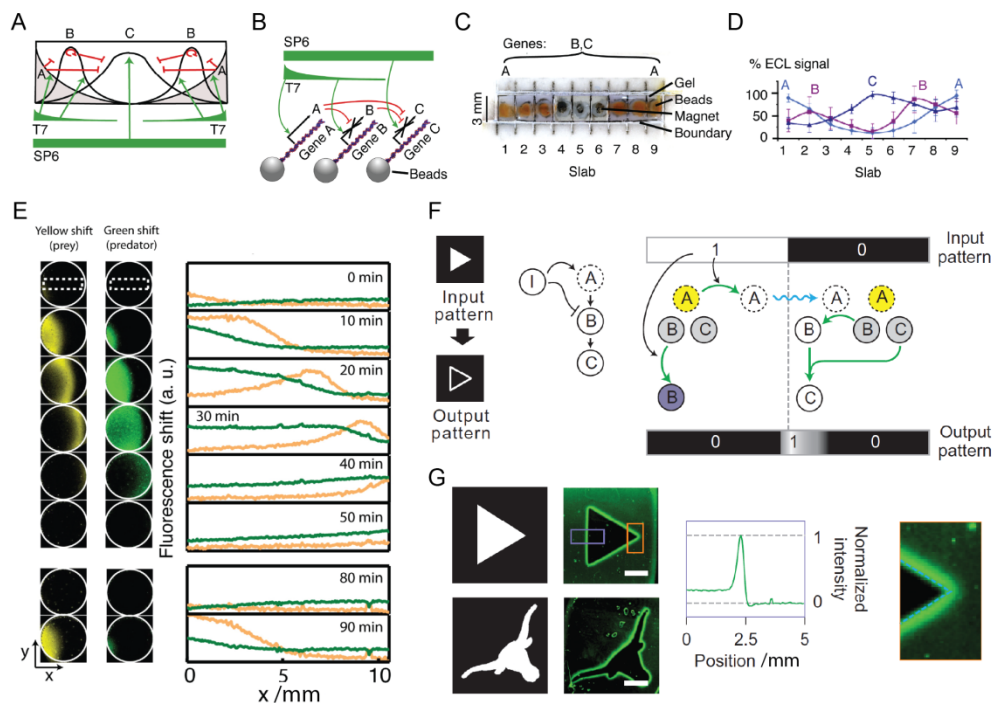


Figure 1.12 Spatial patterns in biochemical reaction networks involving nucleic acids. (A) Schematic overview of network interactions mimicking embryonic expression patterns in *Drosophila*. (B) Schematic overview of the implementation of the artificial reaction network. (C) Compartmentalized experimental setup. (D) Normalized expression levels measured by enhanced chemiluminescence (ECL) from Western blots, data of four replicate experiments plus one standard deviation. Image taken from ref. 43. (E) Fluorescent images of DNA-based predator-prey waves (green and yellow, respectively) and corresponding profiles in the region indicated by the dashed rectangle. Image taken from ref. 45. (F) Principle of edge detection via a DNA-based incoherent feed-forward loop. Colors indicate caged (yellow), inactive (grey), ablated (purple) and active (white) species. (G) Fluorescent images showing edge detection after selective UV illumination. Scale bars, 3 mm. Image taken from ref. 46.

redesign of enzymes can be extremely challenging. In contrast, synthetic nucleic acids such as DNA and RNA are easily programmable, allowing rational design of sequence, length and interactions. As a result, tuning of reaction parameters such as rate constants, diffusion coefficients and network connectivity over several orders of magnitude is relatively straightforward. Similar to enzymatic reaction networks, coupling of reaction to diffusion in spatially extended systems based on nucleic acids can lead to important insights into the complex spatiotemporal ordering in living systems. An early example is the work of McCaskill and Bauer⁴⁴ who, in their study on *in vitro* evolution, used the Q β replicase enzyme to amplify short RNA strands in a long thin capillary. Because Q β replicase can use both the plus and minus strand of a double-stranded RNA as a template, replication occurs autocatalytically and traveling concentration waves of constant velocity were observed.

More recently, Estévez-Torres, Rondelez and colleagues⁴⁵ employed the PEN toolbox to engineer a reaction network *de novo* displaying predator-prey oscillations in a quasi-two-dimensional closed reactor. Predator-prey dynamics are classically described by Lotka-Volterra type

reaction-diffusion equations and involve prey growth by consuming food and predator growth by consuming prey, leading to oscillatory spatiotemporal behavior. In this study, prey growth was implemented by a template species (*i.e.* the ‘food’) onto which prey can hybridize. Subsequently, two copies of prey are released after extension and nicking. Prey can bind to predator species, after which polymerase turns the prey into a predator. Finally, exonucleases degrade both prey and predator into unreactive products. When the three DNA fragments and three enzymes are homogeneously distributed in a spatially extended reactor, traveling waves of predator chasing prey were observed (Figure 1.12E). One-dimensional reaction-diffusion models agreed well with experimental data and the relation between front velocity and reaction parameters, such as the diffusion coefficient of the predator species and the concentration of polymerase, was found to obey the Fisher-Kolmogorov relation, similar to the results described above for enzymatic reaction networks³⁷.

The programmability of synthetic DNA oligonucleotides was elegantly utilized by Ellington, Chen and co-workers⁴⁶, who engineered an enzyme-free DNA-based molecular program capable of spatial edge detection. The system is based on an incoherent feed-forward loop consisting of an input signal I, intermediate DNA species A and B and output species C. In the reaction network, input I activates A but inactivates B, while A activates B, which in turn produces active species C (Figure 1.12F). A highly-crosslinked polyacrylamide hydrogel was used to limit the diffusion of chemical species and a mask was used to selectively illuminate parts of the hydrogel, serving as a spatially modulated input signal I. In illuminated regions of the hydrogel output C is not produced because B is inhibited, while in unilluminated areas C is also not produced, because A cannot activate B to produce C. Critically, near the boundary between regions species A can diffuse from light into dark regions, activating species B and subsequently producing output C (Figure 1.12F). Because species C is labeled with a fluorophore, a fluorescent signal is generated only at the edges of the UV illuminated region (Figure 1.12G). Furthermore, the authors show that by changing the connectivity of the network, for example by changing DNA sequences or adding new species, the behavior of the molecular program could be tuned in a modular fashion. In this study, the bottom-up engineering of rationally designed molecular circuits is combined with spatially modulated input signals to generate truly emergent behavior.

In the previous two sections, we have described enzymatic and nucleic acid-based reaction networks that emulate complex spatiotemporal ordering using a minimal set of components. The studies presented here demonstrate that *in vitro* synthetic biologists are capable of emulating life-like out-of-equilibrium behavior using relatively small networks⁴. In the next section, we discuss emerging applications of cell-free synthetic biology.

1.5 EMERGING APPLICATIONS

Due to the modularity and inherent biocompatibility of engineered biomolecular circuits, applications of this technology in the biomedical sciences are within reach. We will briefly highlight several emerging applications of this technology in the areas of cell-free genetic biosensors, biological nanofactories capable of reciprocal communication with living systems, and in the area of autonomous materials.

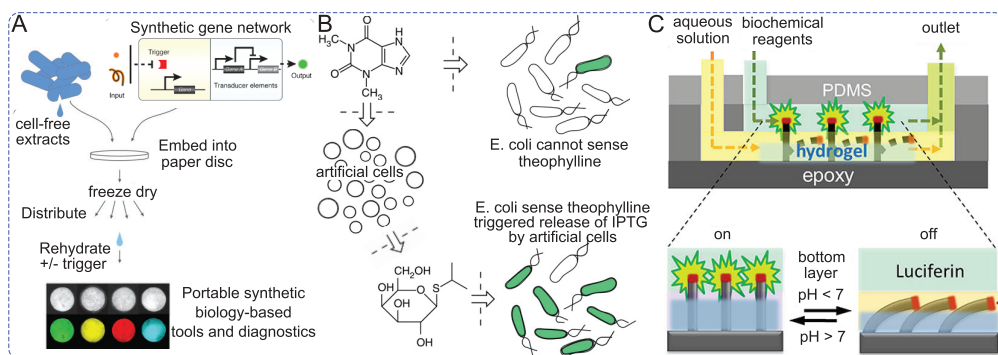


Figure 1.13 Selected emerging applications of engineered biomolecular circuits. (A) Usage of synthetic-biology-based technologies outside the laboratory is facilitated by paper-based technology where cell-free genetic networks are freeze-dried and, after distribution, reactivated by rehydration. Image taken from ref. 48. (B) Engineered, artificial, non-living cellular mimics activate (or repress) natural sensory pathways in living cells through chemical communication, allowing new opportunities in controlling cellular behavior without the use of genetically modified organisms. Image taken from ref. 50. (C) A signal converter translating chemical signals at the nanoscale (pH changes) to macroscopic optical outputs using chemo-mechanical feedback loops. Image taken from ref. 51.

A first emerging application comprises cell-free genetic biosensors. In contrast to genetically engineered biosensing cells, which might have limited detection ranges due to membrane impermeability of the analytes and which may be difficult to commercialize due to official regulations of genetically modified organisms, cell-free biosensors can be readily made using coupled *in vitro* transcription translation systems. Pellinen *et al.*⁴⁷ for instance used an *E. coli* extract to produce firefly luciferase as reporter in response to specific analytes. The authors showed that their cell-free approach allowed for improved sensitivity, wider detection range and faster assays combined with minimal preparation times. Recently, Collins *et al.*⁴⁸ have taken this approach to the next level by freeze drying cell-free synthetic gene networks onto paper and other porous materials (Figure 1.13A). The resulting abiotic materials are sterile and stable at room temperature, allowing for facile storage and distribution. Retaining their *in vitro* transcription and translation capability, these materials can be activated at will by simply adding water. As such, the paper-based platform allows for the safe use of engineered gene networks beyond the lab, *e.g.* in the clinic and industry, presumably at a low cost. The authors demonstrated their platform as glucose sensors and strain-specific Ebola virus sensors making use of a colorimetric output. Due to the modularity of the paper-based approach, arbitrary reaction networks from the ever growing toolbox of synthetic biology can be added resulting in paper-based biosensors for ultrasensitive multiplexed diagnostics. In addition, simple cameras present on many contemporary mobile phones can be used to quantitatively detect the optical readout. Therefore, this technique paves the way for the creation of low-cost biosensors that can be embedded ubiquitously into daily life. Finally, the time-to-result is important in all applications for biosensors. Reduction of the response times of genetic reporter circuits could be increased *e.g.* with the utilization of reporter proteins that require less to no maturation time, *i.e.* those that require less to no post-translational reaction steps.

A second emerging application of cell-free circuits consists of biological nanofactories that can communicate with living systems. This technology is based on the fact that cells sense and respond to their environment, and each other, using extracellular molecules. Bentley *et al.*⁴⁹ engineered

nanofactories, *i.e.* macromolecules consisting of modules that can target, sense and synthesize molecules, to trigger communication between different bacterial populations. Their design comprises an antibody to selectively target the outer-membrane of bacteria where it triggers a quorum sensing response by means of cell-surface synthesis and delivery of quorum signaling molecules.

In addition, the technology was used to trigger communication between two bacterial populations that otherwise are non-communicative. The idea of targeting existing cell signaling pathways has also been pursued by Mansy and coworkers⁵⁰ who engineered artificial, non-living cellular mimics to activate (or repress) existing sensory pathways of living cells through chemical communication. They expanded the senses of *E. coli* by adding liposomes containing a genetic network that converts a chemical message that *E. coli* cannot sense to a molecule that activates a natural cellular response (Figure 1.13B). This approach may allow for new opportunities in engineering cellular behavior without exploiting genetically modified organisms. Moreover, it paves the way to revolutionize the treatment of bacterial infections as molecules that interrupt or modulate bacterial communication rather than their viability exert less selective pressure to develop resistance.

Other interesting applications may also arise from the coupling of biochemical and mechanical processes. Mechano-chemical conversions are abundant in living systems, which are able to mechanically respond to a changing chemical environment. However, although being a highly desired property for next-generation materials and biosensors, mechano-chemical feedback is generally lacking in synthetic systems. Interesting in this respect is the development of Self-regulated Mechano-chemical Adaptively Reconfigurable Tunable Systems (SMARTS) by Aizenberg and coworkers⁵¹.

This technology utilizes catalyst-bearing microstructures embedded into a hydrogel which mechanically deforms upon a biochemical signal. In this way for instance a pH change can switch on/off an optical output producing biochemical process (Figure 1.13C). The vast variety of switchable biochemical reactions in combination with the wide variety of triggers that induce a hydrogel response (ranging from pH, glucose or other metabolic compounds, light to temperature) and the range of possible outputs (*i.e.* fluorescence, gas generation, visible color change) paves the way for a new generation of bio-responsive materials and biosensors that can autonomously function in and on the human body.

1.6 OUTLINE OF THIS THESIS

In this introductory Chapter, we have highlighted a diverse range of studies on chemical self-organization starting from seminal work on the BZ reaction to more recent endeavors employing biological components. The central element of these multidisciplinary research efforts is to understand and emulate complex kinetic networks of regulatory circuits in the living cell. To make advances in this field will require extensive collaboration between physical chemists and biologists and a firm integration of mathematical modeling of complex reaction networks into the design process. The bottom-up construction of CRNs will give us a unique approach to understand, predict, and possibly control the behavior of complex molecular systems.

The aim of this Thesis is to show that for the furtherance of—and developments within—the field of synthetic biology, it is imperative that one makes extensive use of mathematical modeling and computational tools to design, predict and analyze biochemical reaction networks in a modular fashion. We will demonstrate the necessity of *in silico* approaches within synthetic biology based on several studies. The central research question is thus to what extent we can employ computational techniques to design, predict and analyze the behavior of synthetic biochemical reaction networks in both time and space.

In Chapter 2, we develop an automated design procedure for PEN-based BRNs. This automated procedure integrates many aspects, ranging from mathematical modeling, techniques from computational biology and bioinformatics and notions from statistical mechanics and thermodynamics. We show that we are able to produce accurate descriptions of a previously engineered oscillating PEN-based network and apply the methodology to the design of a new type of PEN-based circuit, one that shows adaptation.

In Chapter 3, we show that experimental results create the need for a shift of a more *a priori* engineering role of mathematical modeling towards an *a posteriori* analytic role. Based on experiments, the models from Chapter 2 are extended and used to reevaluate descriptions of the adaptive BRN. We computationally find increased quality of the adaptive response. This reevaluation introduces the need to apply model reduction and use the reduced models to intuitively analyze the origin of the phenomenon responsible for the increase in quality of adaptation.

In Chapter 4, we demonstrate the dual role of *in silico* techniques in both an *a priori* design process and an *a posteriori* analysis procedure. In designing an enzymatic oscillatory BRN, we show how parameter sampling of a mathematical model can exclude certain choices in the design process of the oscillator. Based on experiments we compose a two parameter sets that, in consolidation with the model, optimally resembles the experimental data. Even though we base the optimization criterion on an empirical cost rather than a probability distribution, the model closely resembles the experimental data and we obtain insights into the identifiability of the parameters in the model.

The final study in Chapter 5 deals with *a posteriori* simulations of solvent evaporation of dilute solutions containing single-chain polymeric nanoparticles. Experimentally, the evaporative self-assembly of such solutions with polymeric nanoparticles results in interesting characteristic morphologies. To rationalize the mechanism by which these spatial patterns form, Monte Carlo simulations set to operate under conditions reminiscent of the nonequilibrium conditions in a closed chamber of a nonequilibrium evaporation process leads to simulated morphologies that are statistically compared to experimental data. In contrast to previous Chapters, there is no temporal data available, and the simulation technique employed is therefore a coarse-grained model that does not incorporate temporal dynamics. Nevertheless, such a phenomenological approach provides the necessary insights into the physical self-assembly mechanism.

1.7 REFERENCES

1. K. Kaneko (2006), *Life: An Introduction to Complex Systems Biology*. Springer.
2. B. Hess and A. Mikhailov (1994), Self-organization in living cells. *Science* **264**, 223–224.

3. U. S. Bhalla and R. Iyengar (1999), Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387.
4. A. Alon (2007), Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461.
5. W. A. Lim, C. M. Lee and C. Tang (2013), Design principles of regulatory networks: searching for the molecular algorithms of the cell. *Mol. Cell* **49**, 202–212.
6. O. Brandman and T. Meyer (2008), Feedback loops shape cellular signals in space and time. *Science* **322**, 390–395.
7. E. Karsenti (2008), Self-organization in cell biology: a brief history. *Nat. Rev. Mol. Cell Biol.* **9**, 255–262.
8. S. Soh, M. Byrská, K. Kandere-Grzybowska and B. A. Grzybowski (2010), Reaction-diffusion systems in intracellular molecular transport and control. *Angew. Chem., Int. Ed.* **49**, 4170–4198.
9. I. R. Epstein and J. A. Pojman (1998), *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos*. Oxford University Press.
10. R. F. Ludlow and S. Otto (2007), Systems chemistry. *Chem. Soc. Rev.* **37**, 101–108.
11. A. C. Forster and G. M. Church (2007), Synthetic biology projects *in vitro*. *Genome Res.* **17**, 1–6.
12. P. Schwiille and S. Diez (2009), Synthetic biology of minimal systems. *Crit. Rev. Biochem. Mol. Biol.* **44**, 223–242.
13. A. N. Zaikin and A. M. Zhabotinsky (1970), Concentration wave propagation in two-dimensional liquid-phase self-oscillating system. *Nature* **225**, 535–537.
14. K. J. M. Bishop and B. A. Grzybowski (2006), Localized chemical wave emission and mode switching in a patterned excitable medium. *Phys. Rev. Lett.* **97**, 128702.
15. Z. Noszticzius, W. Horsthemke, W. D. McCormick, H. L. Swinney and W. Y. Tam (1987), Sustained chemical waves in an annular gel reactor: a chemical pinwheel. *Nature* **329**, 619–620.
16. J. Horváth, I. Szalai and P. De Kepper (2009), An experimental design method leading to chemical Turing patterns. *Science* **324**, 772–775.
17. M. Toiya, V. K. Vanag and I. R. Epstein (2008), Diffusively coupled chemical oscillators in a microfluidic assembly. *Angew. Chem., Int. Ed.* **47**, 7753–7755.
18. A. F. Taylor, M. R. Tinsley, F. Wang, Z. Huang and K. Showalter (2009), Dynamical quorum sensing and synchronization in large populations of chemical oscillators. *Science* **323**, 614–617.
19. O. Steinbock, P. Kettunen and K. Showalter (1996), Chemical wave logic gates. *J. Phys. Chem.* **100**, 18970–18975.
20. A. Hjelmfelt, F. W. Schneider and J. Ross (1993), Pattern recognition in coupled chemical kinetic systems. *Science* **260**, 335–337.
21. M. Nakajima, K. Imai, H. Ito, T. Nishiwaki, Y. Murayama, H. Iwasaki, T. Oyama and T. Kondo (2005), Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. *Science* **308**, 414–415.
22. V. Noireaux, R. Bar-Ziv and A. Libchaber (2003), Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12672–12677.
23. J. Kim, K. S. White and E. Winfree (2006), Construction of an *in vitro* bistable circuit from synthetic transcriptional switches. *Mol. Syst. Biol.* **2**, 68.
24. J. Kim, I. Khetarpal, S. Sen and R. M. Murray (2014), Synthetic circuit for exact adaptation and fold-change detection. *Nucleic Acids Res.* **42**, 6078–6089.
25. A. Padiac, T. Fujii and Y. Rondelez (2012), Bottom-up construction of *in vitro* switchable memories. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3212–E3220.
26. H. W. H. van Roekel, L. H. H. Meijer, S. Masroor, Z. C. Félix Garza, A. Estévez-Torres, Y. Rondelez, A. Zagaris, M. A. Peletier, P. A. J. Hilbers and T. F. A. de Greef (2015), Automated design of programmable enzyme-driven DNA circuits. *ACS Synth. Biol.* **4**, 735–745.
27. J. Kim and E. Winfree (2011), Synthetic *in vitro* transcriptional oscillators. *Mol. Syst. Biol.* **7**, 465.
28. E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree and F. C. Simmel (2011), Timing molecular motion and production with a synthetic transcriptional clock. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E784–E793.
29. M. Weitz, J. Kim, K. Kapsner, E. Winfree, E. Franco and F. C. Simmel (2014), Diversity in the dynamical behavior of a compartmentalized programmable biochemical oscillator. *Nat. Chem.* **6**, 295–302.
30. K. Montagne, R. Plasson, Y. Sakai, T. Fujii and Y. Rondelez (2011), Programming an *in vitro* DNA oscillator using a molecular networking strategy. *Mol. Syst. Biol.* **7**, 466.
31. T. Fujii and Y. Rondelez (2013), Predator–prey molecular ecosystems. *ACS Nano* **7**, 27–34.
32. S. N. Semenov, A. S. Y. Wong, R. M. van der Made, S. G. J. Postma, J. Groen, H. W. H. van Roekel, T. F. A. de Greef and W. T. S. Huck (2015), Rational design of functional and tunable oscillating enzymatic networks. *Nat. Chem.* **7**, 160–165.
33. H. Niederholtmeyer, V. Stepanova and S. J. Maerkl (2013), Implementation of cell-free biological networks at steady state. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15985–15990.
34. E. Karzbrun, A. M. Taylor, V. Noireaux and R. H. Bar-Ziv (2014), Programmable on-chip DNA compartments as artificial cells. *Science* **345**, 829–832.

35. Y. Wei, P. J. Wesson, I. Kourkine and B. A. Grzybowski (2010), Measurement of protein–ligand binding constants from reaction–diffusion concentration profiles. *Anal. Chem.* **82**, 8780–8784.
36. M. M. Wrobel, T. Bansagi, S. K. Scott, A. F. Taylor, C. O. Bounds, A. Carranza and J. A. Pojman (2012), pH wave-front propagation in the urea-urease reaction. *Biophys. J.* **103**, 610–615.
37. D. G. Míguez, V. K. Vanag and I. R. Epstein (2007), Fronts and pulses in an enzymatic reaction catalyzed by glucose oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 6992–6997.
38. M. Loose, E. Fischer-Friedrich, J. Ries, K. Kruse and P. Schwille (2008), Spatial regulators for bacterial cell division self-organize into surface waves *in vitro*. *Science* **320**, 789–792.
39. V. Ivanov and K. Mizuuchi (2010), Multiple modes of interconverting dynamic pattern formation by bacterial cell division proteins. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8071–8078.
40. S. N. Semenov, A. J. Markvoort, W. B. L. Gevers, A. Piruska, T. F. A. de Greef and W. T. S. Huck (2013), Ultrasensitivity by molecular titration in spatially propagating enzymatic reactions. *Biophys. J.* **105**, 1057–1066.
41. X. Liao, R. T. Petty and M. Mrksich (2011), A spatially propagating biochemical reaction. *Angew. Chem., Int. Ed.* **50**, 706–708.
42. S. N. Semenov, A. J. Markvoort, T. F. A. de Greef and W. T. S. Huck (2014), Threshold sensing through a synthetic enzymatic reaction–diffusion network. *Angew. Chem., Int. Ed.* **53**, 8066–8069.
43. M. Isalan, C. Lemerle and L. Serrano (2005), Engineering gene networks to emulate *Drosophila* embryonic pattern formation. *PLoS Biol.* **3**, e64.
44. J. S. McCaskill and G. J. Bauer (1993), Images of evolution: origin of spontaneous RNA replication waves. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4191–4195.
45. A. Padirac, T. Fujii, A. Estévez-Torres and Y. Rondelez (2013), Spatial waves in synthetic biochemical networks. *J. Am. Chem. Soc.* **135**, 14586–14592.
46. S. M. Chirieleison, P. B. Allen, Z. B. Simpson, A. D. Ellington and X. Chen (2013), Pattern transformation with DNA circuits. *Nat. Chem.* **5**, 1000–1005.
47. T. Pellinen, T. Huovinen and M. Karp (2004), A cell-free biosensor for the detection of transcriptional inducers using firefly luciferase as a reporter. *Anal. Biochem.* **330**, 52–57.
48. K. Pardee, A. A. Green, T. Ferrante, D. E. Cameron, A. DaleyKeyser, P. Yin and J. J. Collins (2014), Paper-based synthetic gene networks. *Cell* **159**, 940–954.
49. R. Fernandes, V. Roy, H.-C. Wu and W. E. Bentley (2010), Engineered biological nanofactories trigger quorum sensing response in targeted bacteria. *Nat. Nanotechnol.* **5**, 213–217.
50. R. Lentini, S. P. Santero, F. Chizzolini, D. Cecchi, J. Fontana, M. Marchioretto, C. Del Bianco, J. L. Terrell, A. C. Spencer, L. Martini, M. Forlin, M. Assfalg, M. D. Serra, W. E. Bentley and S. S. Mansy (2014), Integrating artificial with natural cells to translate chemical messages that direct *E. coli* behavior. *Nat. Commun.* **5**, 4012.
51. X. He, R. S. Friedlander, L. D. Zarzar and J. Aizenberg (2013), Chemo-mechanically regulated oscillation of an enzymatic reaction. *Chem. Mater.* **25**, 521–523.

Chapter 2

Automated Design of Programmable Enzyme-Driven DNA Circuits

Part of this work has been published:

Hendrik W. H. van Roekel, Lenny H. H. Meijer, Saeed Masroor, Zandra C. Félix Garza, AndréEstévez-Torres, Yannick Rondelez, Antonios Zagaris, Mark A. Peletier, Peter A. J. Hilbers and Tom F. A. de Greef (2015), Automated design of programmable enzyme-driven DNA circuits. *ACS Synthetic Biology* **4**, 735–745.

ABSTRACT

Molecular programming is inspired by the information processing tasks that biomolecular networks in living cells perform, and involves the engineering of systems capable of performing preprogrammed tasks of similar complexity. Specifically, molecular programming allows for the bottom-up engineering of biochemical reaction networks in a controlled *in vitro* setting. These engineered biochemical reaction networks yield important insight in the design principles of biological systems and can be applied to molecular diagnostic systems. The DNA polymerase–nickase–exonuclease (PEN) toolbox has recently been used to program oscillatory and bistable biochemical networks using a minimal number of components. Previous work has reported the automatic construction of *in silico* descriptions of biochemical networks derived from the PEN toolbox, paving the way for generating networks of arbitrary size and complexity *in vitro*. Here, we report an automated approach that further bridges the gap between an *in silico* description and *in vitro* realization. A biochemical network of arbitrary complexity can be globally screened for parameter values that display the desired function and combining this approach with robustness analysis further increases the chance of successful *in vitro* implementation. Moreover, we present an automated design procedure for generating optimal DNA sequences, exhibiting key characteristics deduced from the *in silico* analysis. Our *in silico* method has been tested on a previously reported oscillatory network, the Oligator, and has also been applied to the design of a reaction network capable of displaying adaptation in one of its components.

2.1 INTRODUCTION

Biological systems are hierarchically organized networks in which each level displays internalized self-regulation while coupled to the external environment¹. Specifically, interactions between the hierarchical layers are a key feature of the reaction networks underlying replication, protein production, energy metabolism, and organelle control². Attempts to unravel the hierarchy and the exact mechanisms governing such biological processes can roughly be classified as top-down or bottom-up approaches³. The former tries to infer full biological networks in their natural context qualitatively at the expense of accuracy, while the latter aims at describing and understanding components of smaller size quantitatively at the expense of system complexity^{4,5}. The sophisticated character of biological systems is to a great extent regulated through complex biochemical reactions networks (BRNs). The molecular components of such reaction networks interact in a dynamic manner to sense, to respond and to adapt to external signals, for example, *via* signaling cascades⁶. Recently, the programmability of BRNs has been demonstrated with genetic engineering of bacterial colonies^{7–9}. However, engineering these BRNs in a cellular environment leads to cross-talk with the host housekeeping functions, *e.g.* unforeseen interactions with metabolic systems present in the host reduce the predictability. Such interactions are not always trivial to circumvent; therefore, the study of minimal biological systems in an isolated or *in vitro* fashion is an ongoing area of research^{10,11}.

Employing a bottom-up approach to synthetic *in vitro* systems enables study of BRNs in a modular fashion. DNA-based BRNs have proven to be highly programmable and amenable for such bottom-up design strategies^{12,13}. Moreover, isolation of the molecular machinery for DNA replication and transcription–translation machinery in a test tube^{14,15}, followed by compartmentalization of these processes in droplets and liposomes, has opened the door toward functional artificial cells¹⁶. Harnessing the dissipative transcription–translation machinery in an open system (CSTR) enabled the programmable construction of an *in vitro* oscillating BRN in a defined environment¹⁷. The dynamics of such systems can be further tuned by considering shared resource usage and negative autoregulation to match output and demand^{18,19}. Establishing predefined dynamic behaviors has also been achieved with enzymatically enriched RNA and DNA-based biochemical reactions based on transcription and replication^{20,21}. Recent work has revealed that interconnecting DNA templates (or genelets) with RNA transcripts as dynamical species can yield switches^{22,23}, oscillators^{24,25}, and adaptation²⁶. The successful compartmentalization of these dissipative biochemical reaction networks in water-in-oil microdroplets is further paving the way toward the design of artificial cells capable of molecular communication with living cells or other protocells^{27–29}. Networks that are based on DNA replication, nicking, and degradation (*i.e.*, the polymerase–exonuclease–nickase (PEN) toolbox) can show stable oscillations³⁰, multistability³¹, traveling waves^{32,33}, symbiosis, and chaotic dynamics³⁴, further demonstrating the bottom-up programmability of fundamentally different types of network topologies. Indeed, the reactions of this PEN toolbox are computationally³⁵ and experimentally well characterized³⁶, making the toolbox an excellent choice to study and design new types of network topologies and their associated dynamics.

Here, we focus on systematic bottom-up design of BRNs employing the PEN toolbox. We report the automated design of biochemical networks constructed from the PEN toolbox displaying

PEN toolbox

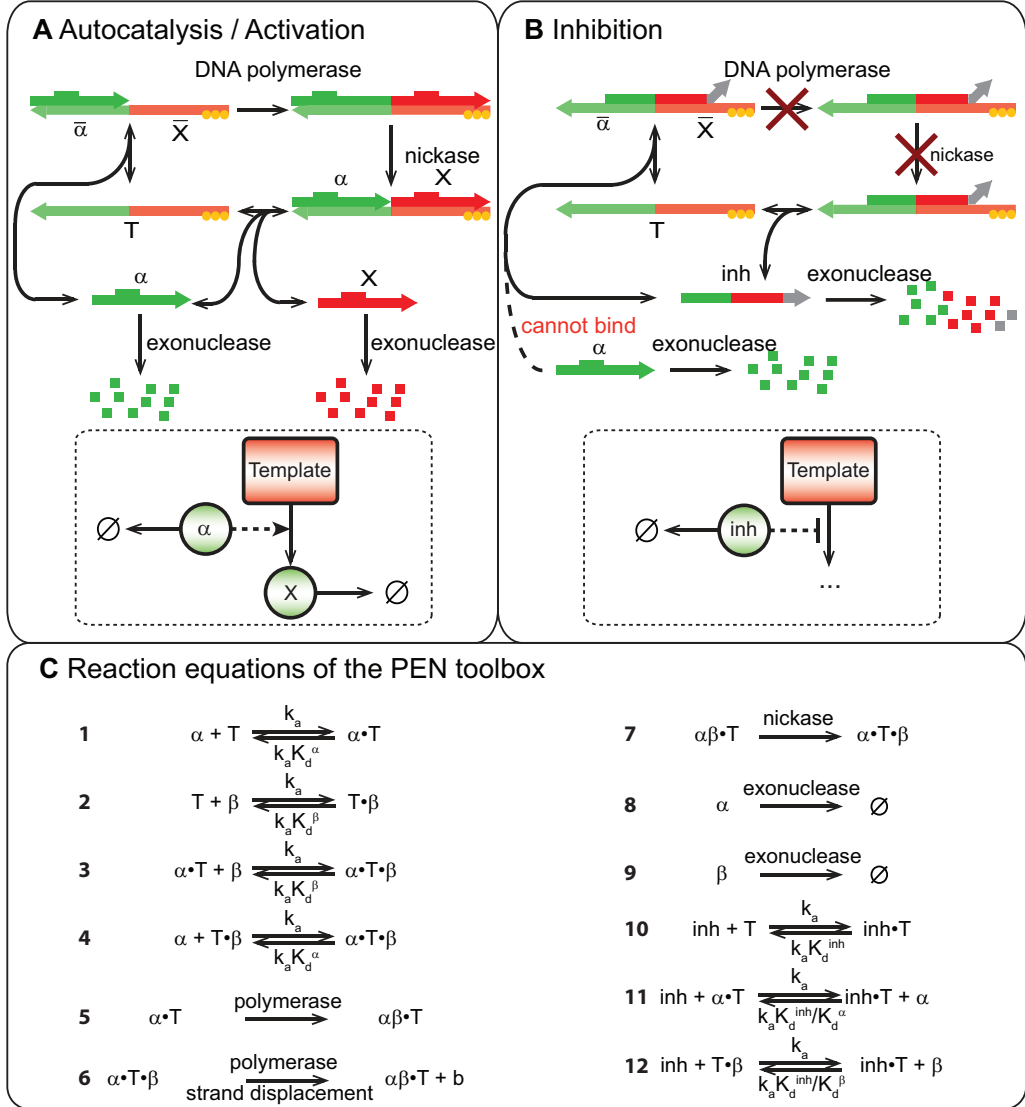


Figure 2.1 Bottom-up engineering of *in vitro* biochemical reaction networks using the polymerase–exonuclease–nickase (PEN) toolbox. (A) Activation of a protected template: oligomer α binds to its complementary part $\bar{\alpha}$ on template T . DNA polymerase then extends the primed template with sequence X complementary to \bar{X} resulting in fully hybridized dsDNA. Oligomer α contains a nicking enzyme recognition site. Because the temperature is close to the melting temperature of the primer–template complex, nicking of the fully hybridized dsDNA results in two oligomers α and X that can readily dissociate from the template. Depending on the choice of the sequence of oligomer X , the template can function as an autocatalytic node, that is, $X = \alpha$, a delay node, that is, $X = \beta$ or an inhibitory node, that is, $X = \text{inh}$. The yellow dots on the template strands indicate a phosphorothioate modified backbone at the 5' end, protecting the template from degradation by exonuclease. (B) Inhibition of a template: an inhibitory oligomer inh binds strongly to a template, rendering it inactive. Enzymes cannot act on the inactive template, as the nicking enzyme recognition site is not fully displayed, and a two base mismatch at the 3' end prevents DNA polymerase from extending oligomer inh . (C) Reaction equations of the PEN toolbox, where primer α drives the production of signal β . These reactions are inhibited by inh . The forward rate constant of

hybridization k_a is assumed to be sequence independent³⁷ and fixed at a previously determined value of $0.06 \text{ nM}^{-1} \text{ min}^{-1}$.³¹ The backward rate constant of hybridization is expressed as the product of the forward rate constant k_a and the equilibrium dissociation constant K_d . The forward rate constant of inhibitor binding to an occupied template, $\alpha \cdot T$ or $T \cdot \beta$, via a toehold, k_{aTH}^{inh} , is chosen equal to k_a , because the toehold length is 6 or 8 bases. The forward rate constants k_{aTH}^{α} and k_{aTH}^{β} are lower, because they initiate strand displacement reactions via shorter toeholds of 3 bases. See Section 2.2.2 for details. Polymerase, nickase, and exonuclease activities are described assuming Michaelis–Menten kinetics.

predefined dynamics. Specifically, an *in silico* approach is presented in which the multidimensional parameter space of dynamical models representing two key network motifs is explored, resulting in parameter regions in which the molecular circuits show either sustained oscillations or adaptation^{38–41}. These dynamical models are derived from the relevant reaction equations in the form of ordinary differential equations (ODEs) and are automatically generated by rule-based software (*vide infra*).

The design of the systems should be optimal in the sense that the target behaviors are retained when the networks are implemented in *in vitro* experiments. Therefore, the robustness to perturbations in the parameters of each circuit is assessed. Indeed, by selecting the most robust parameter set, the chance of successful *in vitro* implementation is increased. The information that is obtained from this computational analysis is subsequently used to automatically design optimal DNA sequences having hybridization energies that closely adhere to the design criteria as obtained from the robustness analysis. We perform this procedure on a network topology capable of oscillatory dynamics (the Oligator)³⁰ and on a PEN implementation of a type-1 incoherent feedforward loop (IFFL)⁴², which is shown to display adaptation⁴¹.

2.2 METHODS

2.2.1 General

The PEN toolbox is based on two fundamental types of reactions as building blocks, that is, activation (Figure 2.1A) and inhibition (Figure 2.1B). These reactions are triggered by short ssDNA strands, which act as regulating signals on longer ssDNA templates resulting in production of an output signal. Activation of templates by short ssDNA input strands enables DNA polymerase to extend the oligomer–template pair resulting in fully hybridized dsDNA strands, which are subsequently nicked resulting in double-stranded complexes consisting of templates, input strands, and newly synthesized output strands. Because reactions are performed at temperatures close to the melting temperature of the partial duplexes, input and output strands readily dissociate from their templates. Templates are inhibited by ssDNA strands that are complementary to part of the template’s sequence but lack a nickase recognition site. An inhibitor further possesses a two base mismatch at its 3’ end, preventing extension of the partial duplex thus rendering the template strand inactive. Signal and inhibition strands are degraded over time by exonucleases, while each template strand is protected against degradation by means of phosphorothioate modifications at its 5’ end. Figure 2.1C provides an overview of all types of reactions that occur in the PEN toolbox with respect to a single DNA template T . Because each template strand in a PEN derived BRN is equivalent to a functional node in a network motif, the regulation of a template’s activity by another template through a signal ssDNA is equivalent to an edge in a network motif. Using this approach, the PEN toolbox can be used as a versatile tool to design network topologies capable of oscillatory or adaptive (Figure 2.2A) behavior. The temporal behavior of such a system is described by ODEs

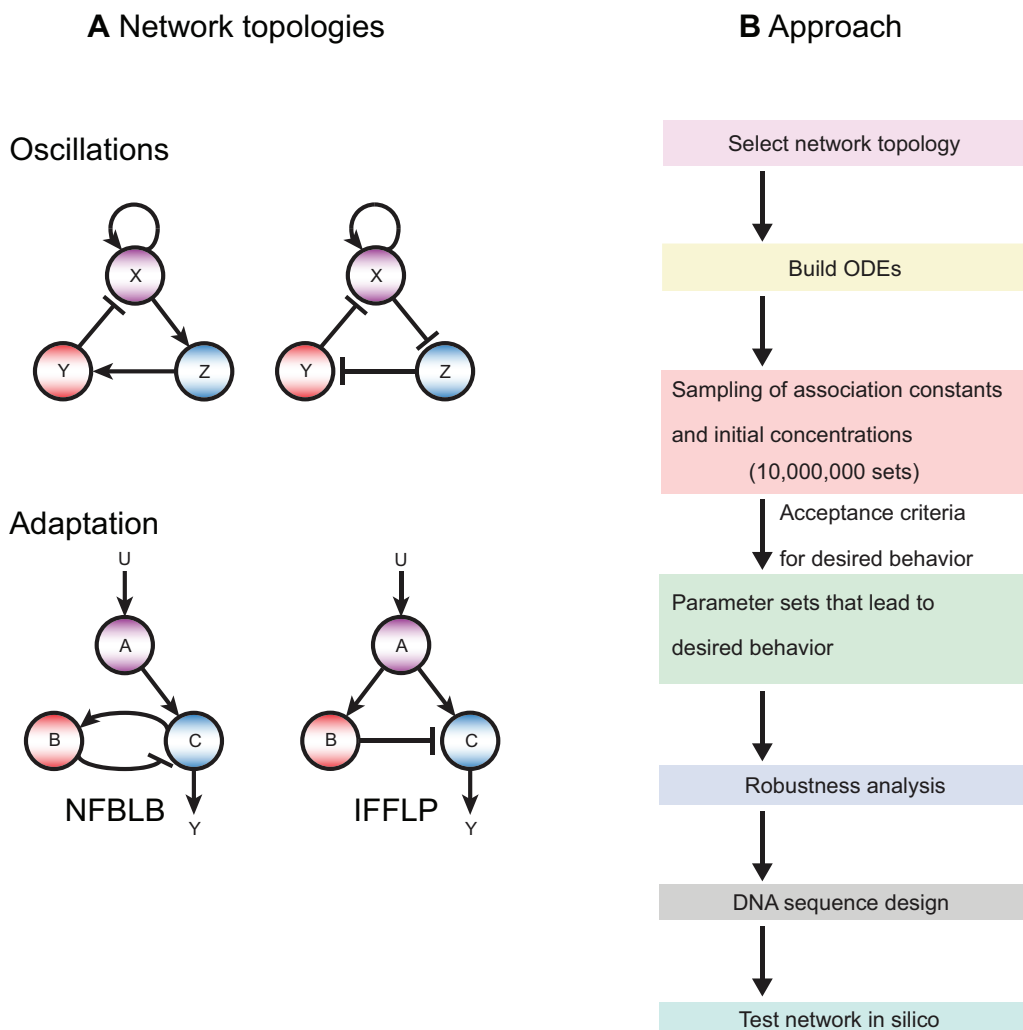


Figure 2.2 *In silico* design of *in vitro* biochemical reaction networks using the polymerase-exonuclease-nickase (PEN) toolbox. (A) Network topologies. (Top) Short-range activation combined with delayed inhibition is a common design motif to construct oscillatory biochemical circuits^{17,24,36,43}. (Bottom) The negative feedback loop with a buffer node (NFBLB) and the incoherent feedforward loop with a proportioner node (IFFLP) can display adaptation⁴¹. (B) The *in silico* approach. First, the ODEs for the network are derived, and subsequently, sampling all dissociation constants K_d and relevant initial concentrations between realistic values (e.g., 10 million randomly generated sets) is then used to optimally cover the parameter space of the model. The K_d values correspond to hybridization energies, ΔG_{AB}^0 , that are matched by choosing the correct base sequence (Section 2.2.3). Responses of the network that meet certain criteria are accepted and analyzed further. The corresponding parameter sets are subjected to robustness analysis, and the parameter set that is characterized as most robust is used as a target for the DNA sequence design algorithm. Thermodynamic parameters obtained from the DNA sequence design algorithm will slightly deviate from the optimal set found using robustness analysis, and therefore, the network is simulated *in silico* again to validate the response.

corresponding to the reactions. An arbitrary network topology is implemented with the PEN toolbox by correctly attributing the exact functionality of oligomers α , β , and *inh* with respect to other templates.

The first two crucial steps of the *in silico* design approach (Figure 2.2B) are selecting a network topology that is known to display a specific dynamic behavior followed by automatically generating the set of ODEs. The set of ODEs is then used to guide circuit design by characterizing the functional parameter space corresponding to the desired behavior (Section 2.2.2). This characterization is performed by randomly generating key parameters of the system within certain ranges, that is, the initial concentrations of DNA templates and trigger strands (*vide infra*), and the hybridization dissociation constants that are tunable via changes in the base sequence. In this work, we have kept the Michaelis–Menten parameters and concentrations of exonuclease, nickase, and polymerase constant at previously determined literature values³¹. Although the concentrations of the three enzymes could be varied *in silico*, an incomplete mechanistic understanding of the isothermal amplification of partial duplexes by extension and nicking precludes the formation of a detailed kinetic model⁴⁴. Therefore, the dissociation constants and the concentrations are the model parameters that are translated to experimental parameters.

Generally, the desired dynamic behavior is observed for a subset of the sampled parameter sets, called viable sets. The robustness of each of these viable sets is quantified in terms of a robustness measure^{45,46} (Section 2.2.3), and the set that exhibits the highest robustness is used as the target set for the automated DNA strand design algorithm.

2.2.2 ODEs, sampling procedure and response classification

We demonstrate the methodology on a network motif that has been previously shown to display oscillatory dynamics, the Oligator, corresponding to the left topology in Figure 2.1C. Figure 2.3A depicts the reaction scheme of the Oligator. Template T_1 serves as an autocatalytic module that produces oligomer α while reactions on template T_2 result in a delay in the activation of T_3 via ssDNA oligomer β . The product of T_3 , oligomer inh , in turn, inhibits the autocatalytic activity of T_1 . We also apply the procedure to investigate if a PEN implementation of a type-1 IFFL (Figure 2.3B) is able to display adaptation. Adaptation is the ability of a system to respond to a change in stimulus after which it returns to prestimulus activity while the stimulus level stays elevated. Sensitivity and precision are measures for the relative response to change in stimulus and the relative difference in steady state values, respectively (*vide infra*).

Types of reactions and reaction rate constants

The full sets of differential equations for the Oligator and IFFL circuit (Figure 2.3) are automatically derived as described in the following subsection, with the use of the reaction types as shown in Figure 2.2A. However, initial scoping simulations have indicated that substrate concentrations can grow quite large in comparison to availability of the enzymes; hence using linear kinetics for enzymatic activity is insufficient to adequately incorporate saturation kinetics. Therefore, we model enzymatic activity using Michaelis-Menten kinetics with simple competition, *i.e.* competition can only occur between substrates that are processed by the enzyme, and unproductive binding is not described³¹ (*vide infra*). All forward reaction rate constants are assumed to be sequence independent, while reaction rates are determined by the dissociation constant (K_d), *i.e.* the ratio of the backward reaction rate constant (k_d) and forward reaction rate constant (k_a), defined as $K_d = k_d/k_a$.³⁷ Reaction 10 of Figure 2.1C is an inhibition reaction, depicted in more detail in Figure 2.4A.

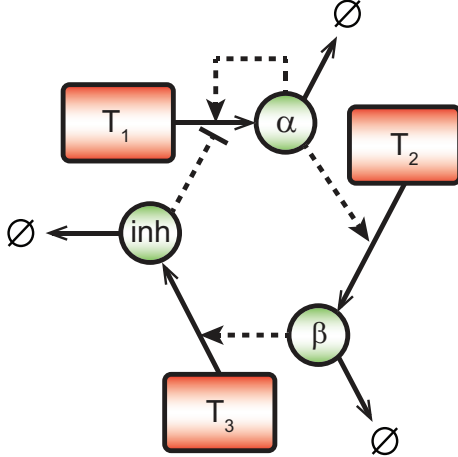
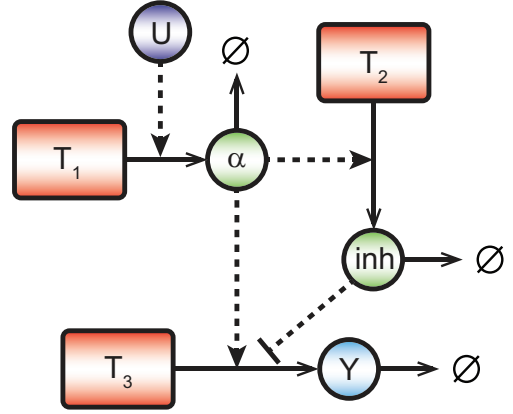
A Oligator**B** IFFL circuit

Figure 2.3 Schematic representation of network motifs as implemented with the PEN-toolbox. (A) Network topology of an oscillatory motif as implemented by the PEN toolbox, *i.e.* the Oligator. (B) Network topology of a PEN-toolbox implementation of a type-I incoherent IFFL. Input trigger U is protected from degradation by exonuclease through multiple phosphorothioate modifications at the 5' end of its backbone. To characterize the effect of these backbone modifications on the kinetics of DNA polymerase when U acts as the trigger on T_1 , we experimentally determined $V_{\max, \text{pol}U}$ and $K_{M, \text{pol}U}$ and used these in the model (Chapter 3).

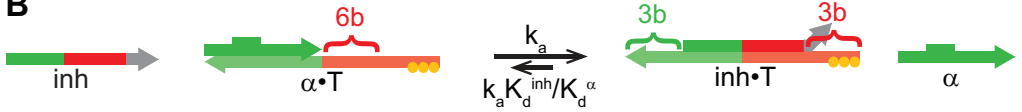
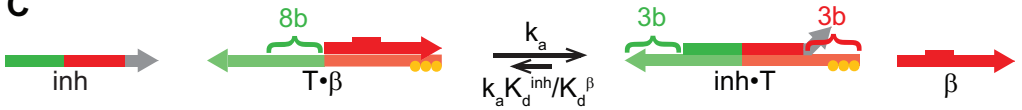
A**B****C**

Figure 2.4 Inhibition of a template and toehold exchange reaction. (A) Simple binding of the inhibitor inh to a template T , (reaction 10 in Figure 2.1C). The inhibitor uses 8 bases to bind to the part on T complementary to α (green), 6 bases to bind to the part on T complementary to β (red), and 2 bases to prevent polymerase from extending the inhibited template (gray). (B) Toehold exchange reaction where the inhibitor inh binds to the 6-base toehold on the template (red parts) and displaces ssDNA α (green ssDNA). In turn, ssDNA α can theoretically displace inh , but is due to the high dissociation of α as compared to inh and the short toehold (3 bases, green) not likely. This part of the Figure corresponds to reaction 11 in Figure 2.1C. (C) Toehold exchange reaction where the inhibitor inh binds to the 8-base toehold on the template (green parts) and displaces ssDNA β (red ssDNA). In turn, ssDNA β can theoretically displace inh , but is due to the high dissociation of β as compared to inh and the short toehold (3 bases, red) not likely. This part of the Figure corresponds to reaction 12 in Figure 2.1C.

Reactions 11 and 12 at the bottom of Figure 2.1C require some elaboration, as they are toehold exchange reactions, that is, reactions where DNA strands are displaced by other DNA strands.

Similar to previously reported PEN systems, an inhibitor binding to a template-input or template-output complex makes use of six or eight bases as its toehold. In these systems, an inhibitor inh initiates a toehold-mediated strand displacement reaction of α on the 3'-end of a template T with a toehold of length of 6 (Figure 2.4B) or a toehold-mediated strand displacement reaction of β on the 5' with a toehold of length 8 (Figure 2.4C). At 42° C, slightly above the melting temperature T_m of the partial duplexes consisting of α and template or β and template, the primers readily dissociate from their templates. Therefore, a toehold consisting of six or eight bases is always expected to be on display, and the rate constant of inh binding to a toehold is assumed to be equal to the hybridization rate for simple hybridization reactions, *i.e.* $k_{aTH}^{inh} = k_a^{inh} = 0.06 \text{ nM}^{-1} \text{ min}^{-1}$. Conversely, the association rate of α or β binding to the template when inhibitor is bound, is less than $0.06 \text{ nM}^{-1} \text{ min}^{-1}$. Zhang and Winfree³⁷ report this forward rate to be approximately 2 to 4 orders of magnitude lower than $0.06 \text{ nM}^{-1} \text{ min}^{-1}$, because an inhibitor bound to a template leaves three bases that can serve as a toehold on both ends. In previous work, this has either been modeled by multiplying the forward rate constant with 10^{-2} ,³¹ or by multiplying the rate with the ratio between the dissociation constants of the bound species, *i.e.* $k_{aTH}^{\alpha} = k_a^{inh} K_d^{inhT} / K_d^{\alpha}$ and $k_{aTH}^{\beta} = k_a^{inh} K_d^{inhT} / K_d^{\beta}$,³⁰ which generally both take small values. To account for general differences between α and β , we chose to adhere to the latter way of modeling.

Automated generation of sets of ODEs based on the PEN toolbox

As network complexity increases, capturing its behavior in a set of differential equations by hand becomes cumbersome and error-prone. Indeed, extensive bookkeeping and the associated difficulties of implementing Michaelis-Menten kinetics along with the desire to engineer ever larger network topologies introduce the need to quickly and without error convert networks of arbitrary complexity to a set of ODEs that can be numerically integrated. Recently, the PySB language was developed, which offers a flexible, high-level environment for rule-based programming in Python⁴⁷. It draws on the powerful numerical and symbolic Python libraries, to quickly build rule-based models. For the purpose of easily generating PEN toolbox based networks, the *modtoml* package was designed, which is based on PySB. This package is used to read specific input files and build a PySB model, which is subsequently enriched with Michaelis-Menten kinetics and converted to a Matlab® ODE m-file. The package can be extended to generate ODE files that are syntactically suitable for alternative programming languages, but extension of *modtoml* is beyond the scope of this Thesis.

The input file that *modtoml* requires is a txt-file that is structured in a specific manner. The first two lines of this file are reserved for the name of the model and additional arguments for the ODE solver. The rest of the file contains three sections, one for the reactions, one for the initial conditions and one for the reaction constants. Upon initialization, the dedicated class opens and reads each line in this file sequentially, while generating a dictionary in which the name as a string and the additional arguments is stored. The initial conditions and reaction rates are saved in separate dictionaries. Finally, the reactions are stored by type of reactions in a final dictionary. For each reaction that is marked as proceeding via Michaelis-Menten kinetics, dummy first order rate constants are added to the dictionary with the reaction rates.

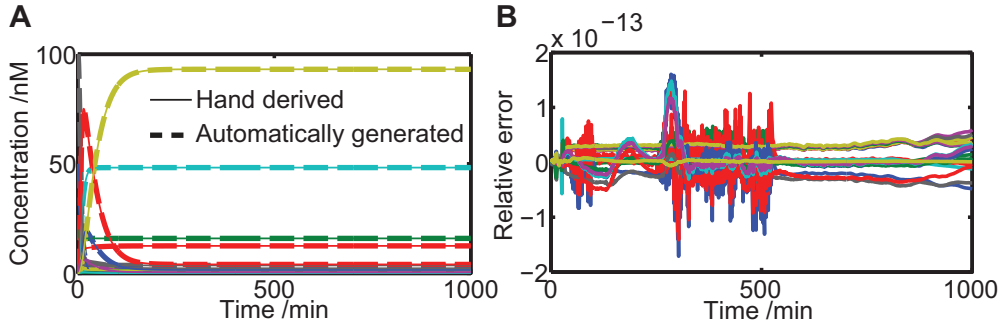


Figure 2.5 Comparison between sets of ODEs derived by hand and automatically generated by *modtoml* for the IFFL circuit. (A) A comparison of two simulations that show the temporal progress of the 20 states of the hand-derived IFFL circuit (solid lines) and the automatically generated set of ODEs (dashed lines). The parameters corresponding to these responses are $K_d^U = 2$ nM, $K_d^\alpha = 2 \times 10^3$ nM, $K_d^{inhT2} = 2 \times 10^{-3}$ nM, $K_d^\gamma = 1.67$ nM, $K_d^{inhT3} = 5.9 \times 10^{-2}$ nM, $[U]_0 = 4$ nM, $[T_1] = 2$ nM, $[T_2] = 50$ nM, and $[T_3] = 100$ nM. (B) The relative error of each of the 20 states with respect to the hand-derived set of ODEs, when simulations are required to report state values at fixed time points, *i.e.* each minute. The maximum relative deviation is approximately 2×10^{-13} , and is a result of a slightly different implementation of the ODEs. The automated generated set requires the reverse rates k_d to be given as explicit parameter values, while the hand-derived set calculates the reverse rates internally.

Table 2.1 Reaction class symbols for the input file

| Symbol | Reaction class |
|--------|--------------------|
| -> | Synthesis |
| - | Inhibition |
| -X | Degradation |
| -0 | Pseudo-degradation |

The three classes of reactions in the PEN DNA toolbox form the basis of the syntax of the input file. To incorporate competitive Michaelis-Menten kinetics and to allow for protected signal ssDNAs, pseudo-degradation and secondary association were also added as reaction classes. For example, the autocatalysis of input α on template T_1 (Figure 3.1A with $X = \alpha$), can be described with two statements, ‘a -> a + a, T1’ and ‘Inh -| a + a, T1’. Here, the ‘->’ symbol indicates synthesis and the ‘-|’ symbol indicates inhibition. Each reaction class has its own unique symbol, as shown in Table 2.1. Python is instructed to parse each line of the segment in the input file that contains the reactions and subsequently find the associated reaction symbol and invoke the correct unpacking function. This is accomplished with relative ease, since the input file is constructed such that each line contains a single reaction. Appendix section A.1, Figure A.1 shows the required construction of the input file that is used to generate the IFFL circuit. The invoked unpacking function will then store the input, output, template and the optional list of competitors for the different enzymes in the appropriate dictionary. The function will also check the file for all necessary constants and raise an error if a constant is missing in the file. To incorporate competitive Michaelis-Menten kinetics and to allow for protected signal ssDNAs, pseudo-degradation and secondary association were also added as reaction classes. A pseudo-degradation reaction will have a positive K_M , but V_{max} will be set to zero. This choice means the species itself will not be degraded, but it will bind competitively to exonuclease, thus influencing the degradation of other species. Figure 2.5 displays a comparison between sets of ODEs derived by hand and automatically generated by *modtoml* for the IFFL circuit.

Network motifs and parameter sampling

We automatically generate the sets of ordinary differential equations (ODEs) for the Oligator and the IFFL with *modtoml*. The ODEs describing the Oligator are shown in equations (A.1)-(A.4) Appendix section A.1, and those that describe the IFFL circuit are displayed in equations (A.5)-(A.8). The dissociation constants in these ODE models can be varied to assess when the system is able to display oscillatory dynamics. The values of the forward rate constant and the enzymatic activities are taken from literature^{30,31,34} and summarized in Appendix section A.1, Table A.1.

In order to screen the Oligator and the IFFL circuit we performed large-scale parameter sampling of the remaining parameters in the model, *i.e.* the dissociation constants, as well as the initial concentrations of templates and oligomer α , between realistic values based on literature. Latin Hypercube Sampling was performed in logarithmic space, and 10 million parameter sets were drawn from the ranges shown in Appendix section A.1, Table A.2. The choice of 10 million sets is based on the observation that for the motifs explored in this Chapter, there are at least 1000 parameter sets displaying the desired behavior. For each parameter set, the system of ODEs was numerically integrated in MATLAB® R2010a with compiled MEX files^{48,49} using numerical integrators from the SUNDIALS CVode package⁵⁰. Note that the value of K_d^{inhT3} (K_d^{inhT2}) is always lower than K_d^{inhT1} (K_d^{inhT3}), as the inhibitor must contain a two-base mismatch with T_1 (T_3) for the Oligator (IFFL circuit) to prevent extension on an inhibited template.

For the Oligator, simulations of 7000 minutes were carried out, and the output of each simulation is classified as responsive when the total dsDNA concentration *i.e.* $[dsDNA]_{tot}$, exceeds a threshold concentration of 1 nM within 100 minutes of simulation time; otherwise the output is termed unresponsive and not considered further. Simulations of 7000 minutes are performed because this time span is much longer than that of a typical experiment (*i.e.* ~1800 minutes). In this manner, we can unambiguously assessed if an oscillatory response is sustained. The response is classified as a sustained oscillation in case $[dsDNA]_{tot}$ exhibits at least 15 peaks within these 7000 minutes of simulation time (that is, the experimental period of an oscillation may not exceed 467 minutes for practical purposes), and the amplitude of the 15th peak is greater than or equal to 97.5% of the amplitude of the 14th peak. Figure 2.6A shows a visual representation of these criteria. The output of the Oligator is $[dsDNA]_{tot}$ since this concentration can be measured directly using EvaGreen³⁰.

The output of the IFFL circuit is the free concentration of ssDNA strand Y . Simulations of 1000 minutes are performed, as this is an experimentally feasible time span for the system to settle into a steady state. If the system reaches a steady state within these 1000 minutes, the concentration of input primer U is increased to be twice the initial concentration of input primer U , *i.e.* $[U](t = 1000) = 2[U](t = 0)$. The response is evaluated in terms of sensitivity S and precision P , defined in equation (2.1), which is accompanied by Figure 2.6B.

$$S = \left| \frac{(Y_{peak} - Y_1) / Y_1}{(U_2 - U_1) / U_1} \right|, P = \left| \frac{(Y_2 - Y_1) / Y_1}{(U_2 - U_1) / U_1} \right|^{-1} \quad (2.1)$$

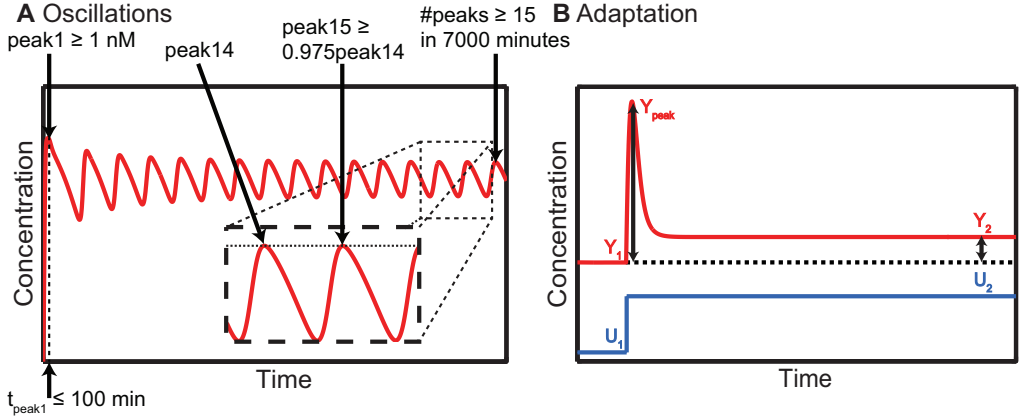


Figure 2.6 Response classification. (A) Criteria for an oscillatory response in total dsDNA concentration for the Oligator model. The first peak of total dsDNA concentration should exceed 1 nM, and take place within 100 minutes of simulation time. Furthermore, the dsDNA concentration is required to exhibit at least 15 peaks within 7000 minutes of simulation time, and the amplitude of the 15th peak must be greater than or equal to 97.5% of the amplitude of the 14th peak. (B) Sensitivity and precision. A response Y that reaches steady state Y_2 from steady state Y_1 as a result of an input increase from U_2 to U_1 , while reaching a peak value Y_{peak} , can be characterized with its sensitivity S and precision P . Such a response is defined as adaptive if $\log_{10}(S) \geq -0.5$ and $\log_{10}(P) \geq 1$.

The response of the IFFL circuit is considered adaptive when $\log_{10}(S) \geq -0.5$ and $\log_{10}(P) \geq 1$. That is, the relative peak of the transient in the response compared to the relative increase in input should be high, while the relative change in steady state after the transient compared to the relative change in input should be low.

2.2.3 Robustness analysis

Approach

From the parameter sampling of the biochemical reaction network models, a collection of parameter sets that results in a certain desired behavior (*i.e.* oscillations or adaptation) is obtained. More specifically, a portion of the viable parameter space is being covered through sampling, but we do not know whether naïve sampling covers the whole viable space. Several studies have previously been reported to efficiently characterize viable parameter spaces in combination with robustness analysis^{45,46}. Following Wagner *et al.*⁴⁵, an iterative procedure is guided with principal component analysis (PCA) from the parameter sets obtained through sampling, greatly expanding the viable space (Figure 2.7A, **I**). When this iterative procedure has converged (*i.e.* if the principal variances in two consecutive iterations deviate less than 10% from each other), Monte Carlo (MC) integration is performed (Figure 2.7A, **II**) to assess this viable space. MC integration is performed by the construction of a hyperbox around the viable space obtained by the directed search. Dense uniform sampling is performed within this hyperbox and this sampling procedure results in a subset of viable parameter sets. Generally, this procedure increases the number of parameters leading to the desired behavior, and embedded with the right constraints, will cover the region of the viable space that is experimentally feasible. Finally, the resulting accepted parameter sets are subjected to local analyses (Figure 2.7A, **III**), by perturbing each parameter in each set independently and simultaneously and characterizing the fraction of accepted parameter sets. The parameter set that retains the desired

behavior in most perturbations is classified as the most robust parameter set, and has the highest chance of succeeding in an *in vitro* experiment.

Details of the directed search method

The accepted parameter vectors resulting from naïve sampling, denoted collectively in matrix Θ_1 , will be used for a directed search. Based on Wagner *et al.*⁴⁵, PCA is performed on Θ_1 , and the principal axes of this PCA are the eigenvectors of the $N \times N$ covariance matrix Σ_1 , where N is the number of parameters in the model. This matrix can in turn be used to perform a Gaussian process with zero mean and covariance matrix Σ_1 . The N -dimensional Gaussian process is used to generate a set S_2 of parameter vectors, of which a subset Θ_2 of parameter vectors shows the desired behavior. This is generally expressed in equation (2.2).

$$S_{i+1} = \{\tilde{\theta}_j = \langle \Theta_i \rangle + \lambda_i \tilde{\epsilon}_j \mid j = 1, \dots, L\}, \quad i = 1, \dots, n \quad (2.2)$$

Here, $\tilde{\theta}_j$ is the j -th parameter vector resulting from the j -th realization of this Gaussian process, denoted as $\tilde{\epsilon}_j$. The element-wise mean of the accepted parameter vectors of the previous iteration is denoted as $\langle \Theta_i \rangle$, λ_i is a scaling factor for the standard deviations of the distribution along the PCA axes and n is the last iteration, either preset by the user or reached by convergence of the procedure. Because $\lambda_i > 1$, the sampling space is taken to be broader than expected from Σ_i , ensuring that sampling reaches a portion of the space beyond the bounds set by the previous iteration alone. L is the number of samples (taken to be 10^6). We identify the subset $\Theta_{i+1} \subset S_{i+1}$ of parameter vectors that retain our desired behavior, and move to the next iteration if Θ_{i+1} contains 1000 accepted parameter sets. The value of λ_i decreases linearly with i from 2.5 to 1.5 in the first 5 iterations, and remains 1.5 for $i > 5$. The directed search has converged if the maximum change in standard deviation along the PCA directions compared to the previous iteration is less than 10%, as shown in equation (2.3). The choice of 10% is based on the empirical notion that a larger percentage leads to convergence too fast, while a lower percentage does not converge fast enough.

$$\max_k \left| \frac{\sqrt{(\Sigma_{i+1})_{kk}}}{\sqrt{(\Sigma_i)_{kk}}} - 1 \right| < 0.1, \quad k = 1, \dots, N \quad (2.3)$$

Monte Carlo integration

A hyperbox is constructed with sides parallel to the principal axes of the converged accepted set of parameter vectors Θ_n . These sides have a length determined by the minimum and maximum value of each model parameter in PCA space. That is, performing a PCA on Θ_n allows us to represent each parameter vector $\tilde{\theta}_m \in \Theta_n, m \in \{1, \dots, 1000\}$ with respect to a basis determined by the principal axes. Therefore, there exist a minimum value and a maximum value resulting from this operation on each principal axes, because Θ_n is finite. A dense uniform sampling is performed in this hyperbox (10^6 parameter vectors) and the model is assessed if the desired behavior is displayed for each of the parameter vectors. The resulting MC integrated set of parameter vectors is denoted as Θ_{MC} .

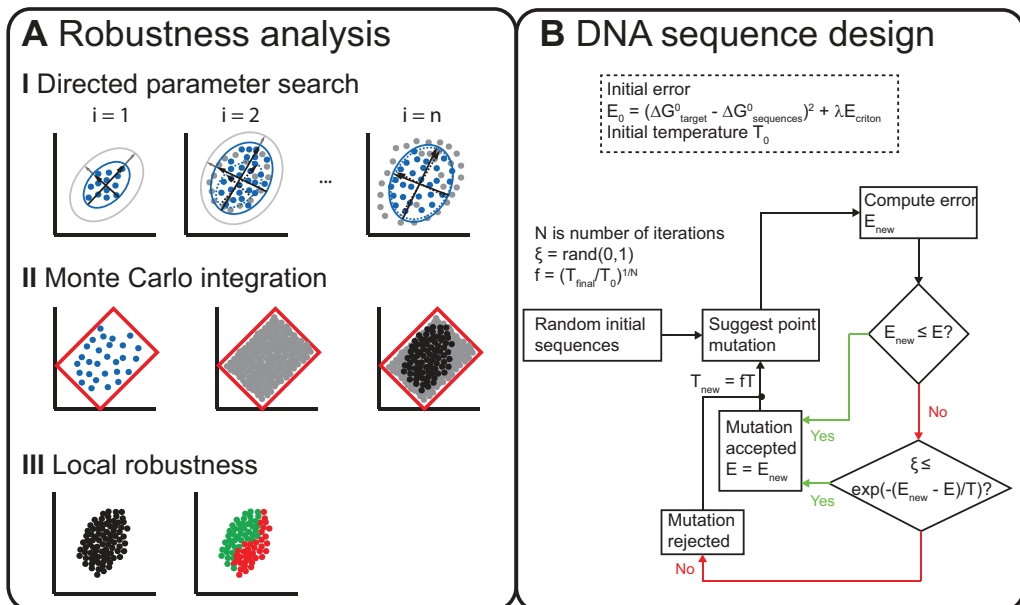


Figure 2.7 Robustness analysis and DNA sequence design. (A) Robustness analysis consisting of three steps. **I** Directed parameter search. ($i = 1$) The parameter sets resulting from the naïve sampling procedure (blue) are subjected to PCA (blue ellipse and black arrows). Directed sampling is now performed in the directions of the principal axes (gray ellipse and gray extensions). ($i = 2$) These newly sampled parameter sets are either accepted (blue dots) or rejected (gray dots). PCA is performed on the newly accepted sets (solid blue ellipse and black arrows) and compared to the PCA of the previous iteration (dashed blue ellipse and dashed black arrows). This procedure is repeated (gray ellipse and gray extensions) until the PCA on the accepted parameter sets of the current iteration does not deviate significantly from the PCA on the accepted parameter sets of the previous iteration, and the directed search has converged (depicted at $i = n$). **II** A hyperbox is constructed around the accepted parameter sets from the directed search procedure (red). Within this hyperbox, a dense uniform sampling is performed (gray dots), and the parameter sets that show the desired behavior are shown in black. **III** The resulting large portion of parameter sets describing the desired behavior is subjected to local perturbations, assessing how well each parameter set can withstand experimental noise. The parameter set that retains the desired behavior in most perturbations is the most robust parameter set. (B) Simulated annealing procedure for automated DNA sequence design⁵¹. The parameters resulting from the robustness analysis are used to calculate the optimal Gibbs free energy of DNA hybridization, and these are subsequently used as target Gibbs free energies for automated DNA sequence design. DNA hybridization Gibbs free energies are calculated by NUPACK^{52,53} and should match the targets as closely as possible. Next to optimization of the Gibbs free energies of hybridization, self-complementarity and unwanted inter- and intramolecular hybridization are penalized.

Local perturbations

The set of parameter vectors in Θ_{MC} consists of concentrations and dissociation constants, the latter representing hybridization of primers to templates. As concentrations can be controlled fairly well in an *in vitro* experiment, we will not perturb concentrations, but focus on the dissociation constants describing DNA hybridization. The dissociation constants are converted to standard Gibbs free energies of hybridization using equation (2.4).

$$\Delta G_{AB}^0 = RT \ln(K_d^{AB}) - RT \ln([H_2O]) \quad (2.4)$$

The dimensionality of K_d^{AB} is corrected for the molarity of water $[H_2O]$ because for dilute solutions of DNA the concentration of all DNA molecules is negligible compared to the concentration of water³⁷. Hence, the mole fractions of the DNA species in the system are roughly the molarity of water multiplied by their respective molarities. Therefore, correction with the molarity of water ensures that the standard Gibbs free energies of hybridization with respect to the dissociation constants are calculated from mole fractions, consistent with NUPACK^{52,53}. In equation (2.4), R is the gas constant and T the temperature. For each parameter vector in Θ_{MC} , the standard Gibbs free energy of hybridization is calculated from its corresponding dissociation constant. The Gibbs free energies in the model are perturbed with 100 realizations of an $(N - \#C)$ -dimensional Gaussian process with zero mean and standard deviation 1 kcal/mol, where $N - \#C$ is the number of dissociation constants in the model. These perturbations are used to assess the robustness of each realization of the desired behavior in Θ_{MC} against changes in the standard Gibbs free energy of hybridization. In designing the optimal DNA sequences (*vide infra*), NUPACK calculates the standard Gibbs free energy values, but these calculations are not ideal, because of measurement errors in the database, and because of the introduction of phosphorothioate modifications of the backbones of certain DNA strands. Therefore, it is important to assess how well the model employed with a certain parameter vector is able to withstand slight deviations in the standard Gibbs free energy values. Hence, the fraction of accepted sets after perturbing the standard Gibbs free energies is a measure for the overall local robustness of the parameter vector.

Finally, each of the dissociation constants in the most robust parameter vector is assessed on its individual contribution to the vector's overall local robustness. The most robust parameter vector is taken from the simultaneous perturbation, and each individual standard Gibbs free energy of hybridization is perturbed with 1000 realizations of a Gaussian process with zero mean and a standard deviation of 2 kcal/mol, while keeping the other parameters constant at their optimal values. The fraction of these perturbations for which the system displays the desired behavior is an indication of the importance of this individual hybridization energy to the functionality of the network. For hybridization between DNA strands A and B with dissociation constant K_d^{AB} , this fraction leads to the parameter robustness ρ_{AB} . This value is used to weigh the contribution of each hybridization reaction in the automated DNA sequence design (*vide infra*).

2.2.4 Automated DNA sequence design

General approach

Scientific literature contains a wealth of methodologies for automated design of DNA or RNA sequences. Each of these methodologies are designed for a specific purpose, *e.g.* for the design of branched structures⁵⁴⁻⁵⁸ or linear sequences^{59,60}. The automated DNA design algorithm we developed is specific for the PEN toolbox, and incorporates two types of optimization criteria. First, the difference between the target standard Gibbs free energy values as obtained by robustness analysis and the standard Gibbs free energy of hybridization of the DNA duplexes as obtained by NUPACK must be minimal. Second, the DNA sequences should be assessed on sequence quality. That is, cross-hybridizations have to be minimized, including self-complementarity. Furthermore, intramolecular folding into hairpin loops should be prevented, since they reduce the hybridization efficiency. Secondary structures other than the nicked state must also be excluded. Repeating bases must be minimally present for polymerase to efficiently polymerize the strands⁶¹, and repeating

guanines in particular are penalized preventing the formation of G-quadruplexes⁶². Finally, each oligomer should consist of a unique sequence, whenever possible. If these criteria are met, the network has the highest probability of showing the desired behavior *in vitro* with the obtained sequences. We present an approach using simulated annealing⁶³ to design DNA sequences that match these criteria. This approach makes use of the target standard Gibbs free energy values, uses NUPACK to calculate the standard Gibbs free energy values of hybridization in each round for comparison, and evaluates DNA strand quality by assessing the uniqueness of 3-base critons in the strands. The procedure is outlined schematically in Figure 2.7B. Simulated annealing has been employed before in the automated design of DNA sequences^{51,64}.

Matching the target Gibbs free energies of hybridization

NUPACK^{52,53} can predict the standard Gibbs free energy of unpsuedoknotted structures with reasonable accuracy with the nearest-neighbor⁶⁵ mod using experimentally determined thermodynamic parameters^{66–68}. The Gibbs free energy values ΔG_{AB}^0 of DNA hybridization obtained from the most robust parameter set are used as targets for automated DNA sequence design. Furthermore, intramolecular folding can be minimized by minimizing ssDNA Gibbs free energy of self-hybridization. Hence, NUPACK can be used to minimize equation (2.5). In equation (2.5), ρ_{AB} is the individual robustness of the dissociation constant K_d^{AB} in the model corresponding to target standard free energy of hybridization $\Delta G_{AB,\text{target}}^0$, and $\Delta G_{AB,\text{NUPACK}}^0$ is the free energy value of the ordered complex that NUPACK reports for the hybridization. This value should be corrected with the intramolecular interactions of each contributing DNA strand (*i.e.* $\Delta G_{A,\text{NUPACK}}^0$ and $\Delta G_{B,\text{NUPACK}}^0$), therefore, it is the term $(\Delta G_{AB,\text{NUPACK}}^0 - \Delta G_{A,\text{NUPACK}}^0 - \Delta G_{B,\text{NUPACK}}^0)$ that must be optimized. For hybridizations that are not present in the model (cross-hybridizations), $\rho_{AB} = 1$ and $\Delta G_{AB,\text{target}}^0 = 0$. To minimize hairpin loop formation, equation (2.5) also incorporates minimization of the $\Delta G_{A,\text{NUPACK}}^0$ values multiplied by a factor of $\eta = 10$ to obtain numerically equal contributions to the Gibbs error of both terms.

$$E_{\text{Gibbs}} = \sum_{A=1}^{N_{\text{strands}}} \sum_{B=A}^{N_{\text{strands}}} \left(\frac{1}{\rho_{AB}} (\Delta G_{AB,\text{target}}^0 - (\Delta G_{AB,\text{NUPACK}}^0 - \Delta G_{A,\text{NUPACK}}^0 - \Delta G_{B,\text{NUPACK}}^0)) \right)^2 + \left(\eta \sum_{A=1}^{N_{\text{strands}}} \Delta G_{A,\text{NUPACK}}^0 \right)^2 \quad (2.5)$$

Evaluation of the strand quality using the criton concept

To achieve strand uniqueness, the above procedure is enriched with an algorithm that determines the uniqueness of critons—a concept introduced by Seeman⁵⁴, and used in previous work dealing with automated DNA design^{55,58}. A proper sequence for a desired DNA structure must fulfill the following criteria:

1. The sequence of each criton not belonging to a nicking endonuclease recognition site must be unique.

2. Anti-critons located elsewhere than their supposed locations as a result of 1 are not allowed.
3. Self-complementary sequences are not allowed.

The algorithm determines all 3-base critons in each individual DNA strand (templates and oligomers), as well as determines all 3-base critons present in the entire system. Each criton contributes to the sum in equation (2.6).

$$E_{\text{criton}} = \left(\sum_{i=1}^{N_{\text{strands}}} \left(\nu(RB_i + GC_i) + RC_i \right) + RCs_c \right)^2 \quad (2.6)$$

Here, RB_i is a count of the repeated base critons, *i.e.* the sum of critons in each DNA strand that are “AAA”, “TTT”, “GGG” or “CCC”. GC_i is a count of the critons that are “GGG” or “CCC” in a DNA strand, and because we want to prevent sequences that can exhibit G-quadruplexation⁶² are given extra weight. RC_i is a count of the duplicated critons that do not consist of repeated bases in each DNA strand. RCs_c is a count of duplicates in each criton in the entire system, which could be missed by counting the number of duplicated critons in each DNA strand. The value of $\nu = 2$ reflects the importance of penalizing repeated base critons compared to non-uniqueness of non-repeated base critons. Finally, duplicated critons resulting from the incorporation of nickase recognition sites in the sequences are not counted.

Implementation details of simulated annealing

By combining the two error functions in equations (2.5) and (2.6), the cost function for automated DNA strand design is defined as equation (2.7).

$$E = E_{\text{Gibbs}} + \lambda E_{\text{criton}} \quad (2.7)$$

Here, we found that $\lambda = 10$ provides a feasible incorporation of the evaluation of the strand quality based on critons compared to standard free energy optimization.

Simulated annealing is initiated by DNA sequences with a preset length that are randomly generated—except for the nickase recognition sites. The algorithm proceeds by introducing a random point mutation in an arbitrarily selected DNA strand. All DNA strands which are complementary to the one in which the point mutation is introduced, are recursively adjusted at their appropriate locations. The latter is followed by an analysis of the new sequences using the NUPACK executable `complexes`, parallel to an evaluation of the critons, hence providing a value of the error E as defined in equation (2.7). If the computed error value is less than the error value from the previous iteration, the mutation is accepted and the generated sequences are saved. If the condition is not met, the mutation is accepted with a certain acceptance probability $P_{\text{acc}} = \min(1, e^{-\Delta E/T})$ where ΔE is the error difference that is introduced with the point mutation, and T the temperature of optimization. This process is performed for a preset number of iterations. The additional inputs for the NUPACK executable `complexes` are provided in Appendix section A.1, Table A.3. The `dangles` option is set to `some`, since we do not have any *a priori* knowledge about

the effect of phosphorothiate modifications on the Gibbs free energy values, and the concentrations of nicked species are small compared to partial duplexes. Hence, we set dangles to the recommended option provided.

2.3 RESULTS AND DISCUSSION

2.3.1 Automated Design of Programmable PEN Toolbox Networks

The functional parameter space of the Oligator that leads to an oscillatory response as obtained from sampling comprises is depicted in the form of box plots in Figure 2.8A, where the sampling ranges are indicated by bars⁶⁹. Approximately 0.2% of the sample sets are classified as oscillatory. From the box plots, it is observed that the optimal concentration of template T_2 is significantly lower than of the optimal concentrations of templates T_1 and T_3 , suggesting that high concentrations of template T_2 tend to destroy oscillatory behavior. This trend is supported by the original experiments³⁰ and can be rationalized by the fact that high concentrations of T_2 results in rapid production of inhibitor *via* template T_3 , which makes the time-delay over the negative feedback loop too short. Importantly, once the inhibitor is formed, it should strongly inhibit T_1 (low value of $K_d^{inhT_1}$) as a rigorous shutdown of the positive feedback loop is necessary to generate sustained oscillations. Finally, the delicate balance between contributions from the positive and negative feedback loop in the Oligator becomes apparent from the optimal median value of K_d^α . Sustained oscillations only appear if oligomer α rapidly dissociates from templates T_1 and T_2 (*i.e.*, K_d^α adopts high values). We attribute this to the fact that rapid dissociation of oligomer α prevents saturation of template T_1 , and hence, a fast exponential increase of α is guaranteed. On the other hand, fast dissociation of α from T_2 also contributes to the increased time-delay over the negative feedback loop. A typical oscillatory response is depicted in Figure 2.8B where the total base-pair concentration is shown as a function of time, along with the temporal concentration profiles of the individual oligomers α , β , and *inh*. We verify that there is a single limit periodic trajectory in Figure 2.8C in the $[\alpha]$ -*versus*- $[\beta]$ -plane and that, regardless of the initial condition, the response spirals toward that same trajectory. The *in silico* analysis of the Oligator is thus able to identify ranges of thermodynamic parameters and template concentrations that give rise to oscillatory responses and thus greatly facilitates successful experimental implementation. Moreover, these results corroborate with experimental findings by Rondelez and co-workers.

The procedure is repeated to investigate if the IFFL circuit is able to display adaptation. We have chosen the type-1 IFFL because previous work has shown that this pulse-generating motif can control gene expression *in vivo*⁷⁰, and adaptation in type-1 IFFL network motifs is more robust compared to adaptation arising in NFBLB network motifs *in silico*⁴¹. Furthermore, alternative implementations of DNA based IFFL circuits show adaptive responses of similar quality compared to the circuit shown in Figure 2.3B; however, the concentration of output ssDNA is lower (data not shown). In the type-1 IFFL circuit considered here, activation occurs on template T_1 by input strand U that is resistant to degradation. Activation of T_1 by U initiates production of oligomer α that in turn activates both T_2 and T_3 . Activation of T_3 leads to the production of output oligomer Y , while activation of T_2 leads to the production of inhibitor *inh*, which renders T_3 inactive. To explore the phase space of this circuit, simulations were performed in which the system is first equilibrated and the response in the concentration of free Y to a doubling of the concentration of U is evaluated. The

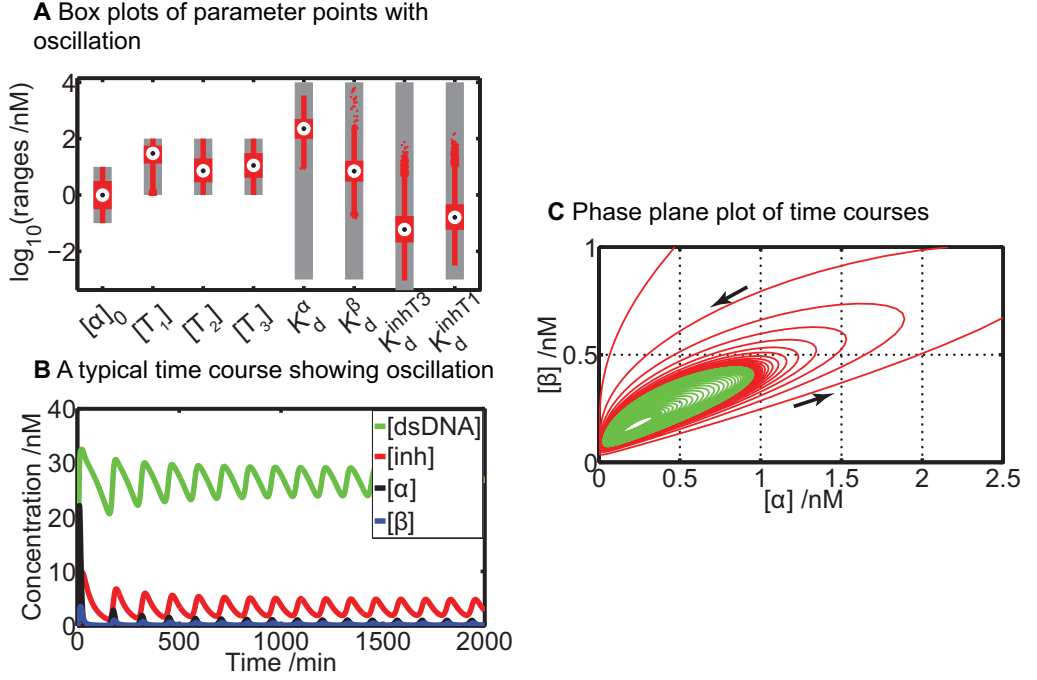
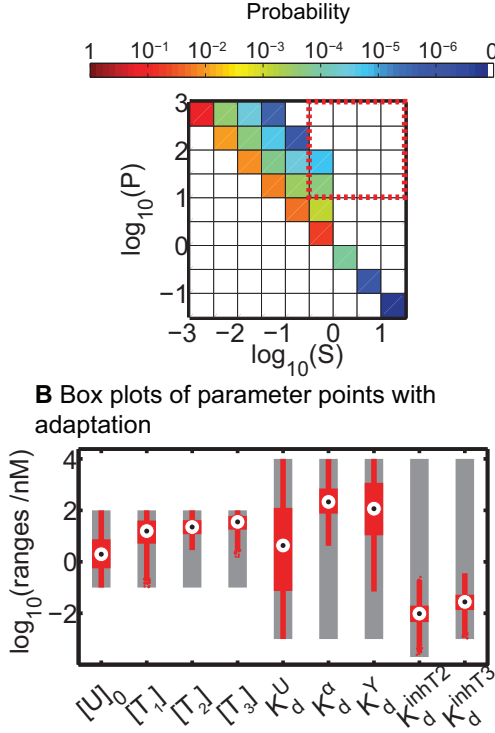


Figure 2.8 Design of a biochemical oscillator using the PEN toolbox. (A) Box plots of parameter points with oscillation. The gray bars indicate the sampling ranges for each parameter, and the red box plots correspond to parameter values leading to sustained oscillations (hit rate 0.2%). Notably, the median of template T_2 is lower than that of T_1 and T_3 , indicating that high concentrations of template T_2 tend to destroy oscillatory behavior, because the delay time becomes insufficiently short. Furthermore, oscillatory behavior is also critically dependent on strong inhibition of T_1 . (B) A typical oscillatory response of the Oligator, showing trajectories of the total dsDNA concentration, and the individual signal oligomers inh , α , and β . For this system, sampling resulted in $K_d^\alpha = 127$ nM, $K_d^\beta = 3.7$ nM, $K_d^{inhT3} = 0.63$ nM, $K_d^{inhT1} = 0.9$ nM, $[\alpha]_0 = 5$ nM, $[T_1] = 16$ nM, $[T_2] = 3$ nM, and $[T_3] = 16$ nM. (C) Trajectories plotted in the $[\alpha]$ -versus- $[\beta]$ plane. Numerically, there is a single limit cycle that attracts all trajectories. The red trajectory is the response shown in B where the amplitudes of the oscillations decrease toward the limit cycle. The green trajectory results from initial conditions close to the equilibrium and shows oscillations of increasing amplitude converging toward the limit cycle.

response of the network measured by the concentration of free Y is classified as adaptive if the sensitivity (S) and precision (P) exceed preset threshold values, that is, $\log_{10}(S) \geq -0.5$ and $\log_{10}(P) \geq 1$ (Figure 2.9A). Figure 2.9B shows box plots that indicate ranges in the concentration and hybridization constants for which the IFFL circuit is able to display an adaptive response. Approximately 0.05% of the responses are defined as adaptive. The IFFL circuit displays adaptive dynamics when the hybridization dissociation constants of oligomers α and Y to their respective templates are high compared to the other hybridization constants, reflecting the fact that fast activation of T_2 and T_3 ensures rapid synthesis of free Y . The low affinity of output Y for T_3 prevents sequestration of Y by its template, which is critical to achieve a response with a high sensitivity. In order for the adaptive response to display a high precision, that is, the free concentration of Y returns to prestimulus levels, inhibition of T_3 by inh must be very strong, which is reflected in low values of the dissociation constant K_d^{inhT3} (and K_d^{inhT2}). The concentration of template T_3 adopts a high value because the response in Y must be noticeable even though inhibition

A Sensitivity and precision heatmap and definitions



C A typical time course showing adaptation

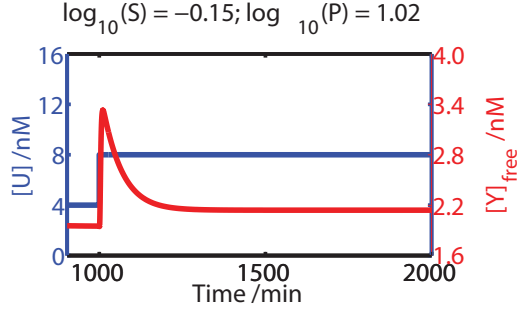


Figure 2.9 Design of an adaptive biochemical circuit using the PEN toolbox. (A) Heatmap of the probability of finding a certain combination of sensitivity and precision values, along with the definitions of sensitivity and precision. The adaptive regime ($\log_{10}(S) \geq -0.5$ and $\log_{10}(P) \geq 1$) is highlighted by a red dashed box. Notably, the probability of finding an adaptive response through sampling is low (hit rate is 0.05%). (B) Box plots as a result of parameter sampling of the IFFL motif. The gray bars indicate the sampling ranges for each parameter, and the red box plots correspond to parameter values leading to adaptation. The analysis shows that the concentrations of input oligomer U and templates T_1 , T_2 , and T_3 can be tuned such that the network displays an adaptive response in the concentration of free Y . The dissociation constants for oligomers α and Y are near their upper sampling bounds, which is explained by the need for fast activation of T_2 and T_3 ensuring rapid synthesis of output Y . The low affinity of Y for T_3 prevents sequestration of Y by its template. In order to overcome the transient increase in Y , that is, matching the steady state value of Y after the transient increase to the previous steady state as closely as possible and thus maximizing the precision of the response, inhibition of T_3 must be very strong. This explains the low values found for dissociation constant K_d^{inhT3} (and hence also K_d^{inhT2}). The concentration of T_3 is high because the response in Y must be noticeable when inhibition is strong. (C) Adaptation as observed in the concentration of oligomer Y in its free form. The parameters corresponding to this response are $K_d^U = 14.6 \times 10^3$ nM, $K_d^\alpha = 1.8 \times 10^3$ nM, $K_d^{inhT2} = 0.037$ nM, $K_d^Y = 5.8 \times 10^3$ nM, $K_d^{inhT3} = 0.042$ nM, $[U]_0 = 4$ nM, $[T_1] = 24$ nM, $[T_2] = 12$ nM, and $[T_3] = 95$ nM.

is strong. The heatmap in Figure 2.9A depicts the probability of finding a response with a certain sensitivity and precision. An important observation from this heatmap is that $\log_{10}(S)$ does not exceed the value 0 (or S does not exceed 1), indicating fold-change in the transient response in Y is unable to exceed fold-changes in the concentration of input U . The automated circuit design is able to identify an adaptive responses in the DNA-based IFFL (Figure 2.9C), however; the sensitivity of the network is always lower than one. Hence, we have shown, using two distinct network

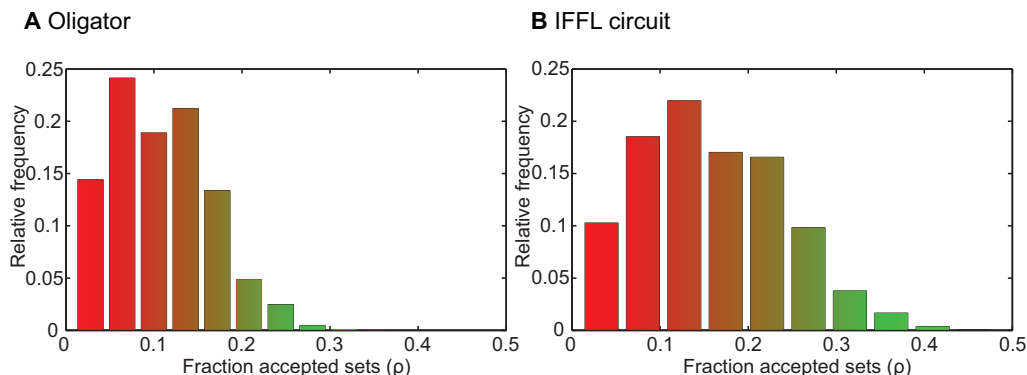


Figure 2.10 Overall robustness versus relative frequency for (A) the Oligator and (B) the IFFL circuit. Robustness increases as the relative frequency decreases for both models. Color gradient indicates fragility (red) and robustness (green) of parameter sets. The IFFL circuit appears to have a higher maximum overall robustness value. A higher overall robustness increases the probability of successful *in vitro* realization.

topologies, that an automated approach is successful in finding functional dynamical behavior in DNA-based enzymatic CRNs.

2.3.2 Robustness Analysis and Automated DNA Sequence Design

In silico design of synthetic biochemical circuits consists of a performance and a robustness classification phase⁷¹. In the previous sections, we have classified the performance of two DNA-based CRNs resulting in parameter ranges where the desired response can be observed. However, this procedure does not yield insight in the robustness of the response, that is, the effect of small perturbations on the performance of the circuit. We have used the “glocal” robustness concept as introduced in Section 2.2.3 to compute a robustness measure for each dissociation constant for the two PEN-based circuits (Figures 2.3A and 2.3B) and have used this information in the design of template sequences⁴⁵. Because the number of viable parameter sets resulting in the desired dynamical behavior is small for both circuits (0.05–0.2%), we first vastly increased the number of viable parameter sets by a directed parameter search of the parameter space close to the viable sets. All resulting parameter sets are locally perturbed, leading to a robustness measure for each parameter set. Figure 2.10 depicts histograms of the overall robustness (fraction accepted sets) ρ of parameter sets that allow the system to be perturbed of the Oligator and the IFFL, *i.e.* those parameter sets for which $\rho > 0$ holds. The maximum overall robustness for the Oligator is $\rho \approx 0.3$, and for the IFFL circuit this is $\rho \approx 0.4$, while the hit rate from naïve sampling is lower for the IFFL circuit (0.05%) than for the Oligator (0.2%). However, comparing the hit rates of the models is unfair as the number of dimensions in the sampling procedures differ in the models (8 for the Oligator and 9 for the IFFL circuit). A comparison is best assessed by the normalized viable volume⁴⁵, which is indeed larger for the IFFL circuit than for the Oligator. A larger normalized viable volume means that a single perturbation of a parameter vector is more likely to generate another viable parameter vector.

Next, the dissociation constants of the set with the highest robustness measure are perturbed individually such that every dissociation constant in the circuit is associated with an individual

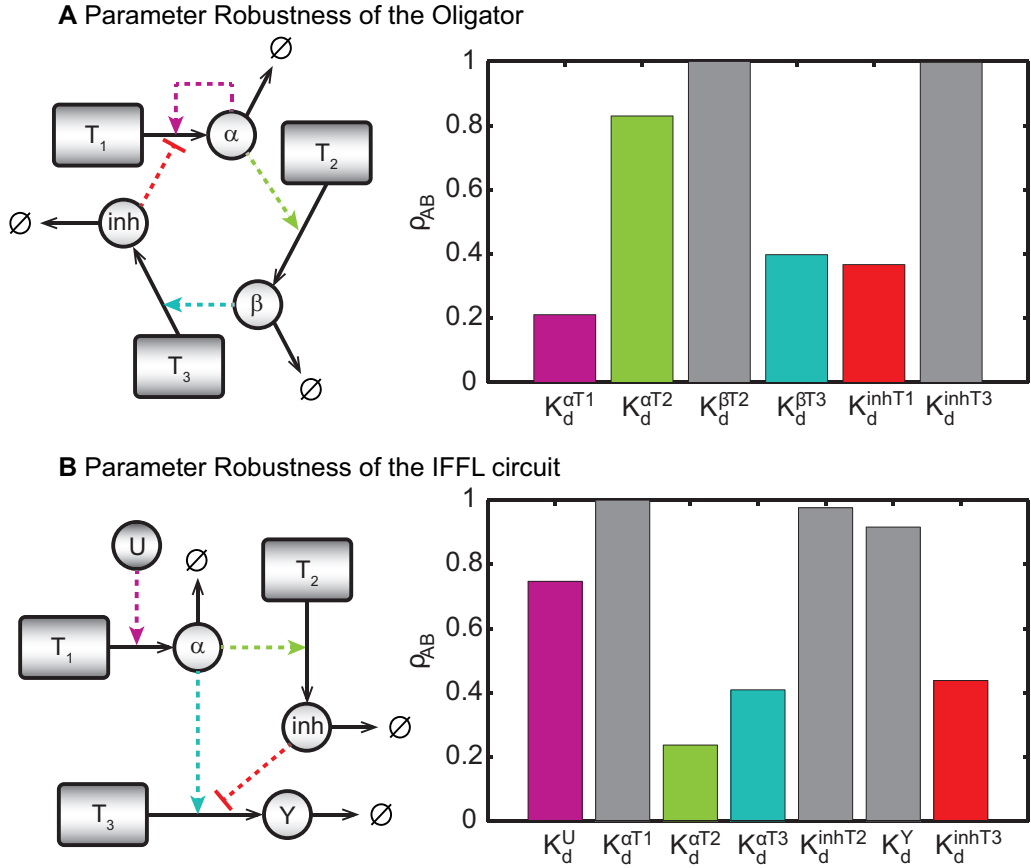


Figure 2.11 Individual parameter robustness of (A) the Oligator and (B) the IFFL circuit. Frail interactions are matched by color in the motifs and in the bar graphs, while relatively robust interactions are not highlighted in the motifs and indicated by gray bars.

robustness measure ρ_{AB} between 0 and 1. Figure 2.11 shows the individual robustness measures of the dissociation constants in the models, and also displays color correspondence with the interaction in the schematic of the network motif. The hybridization Gibbs free energy values are calculated from the optimal dissociation constants and are subsequently used as target Gibbs free energies for automated DNA sequence design. We then perform sequence optimization with a simulated annealing approach. Eventually, the Gibbs free energy values converge toward their respective target values and unique DNA sequences are ensured by incorporating the criton concept in the cost function. We show this convergence to the target Gibbs free energies for the IFFL circuit in Figure 2.12. The iteration that displays the minimum total error corresponds to the sequences of choice for that is suitable for *in vitro* characterization (Chapter 3). The target Gibbs free energy values that follow from the robustness analysis ($\Delta G_{AB, target}^0$) and the Gibbs free values corresponding to the optimal DNA sequences ($\Delta G_{AB, NUPACK}^0$) are shown in Table 2.2. The DNA sequences resulting from the optimization procedure, with the necessary modifications to retain functionality of the circuit, are depicted in Table 2.3. Finally, comparisons between the responses of the target models and the responses of the models embedded with the optimized Gibbs free energy values are shown in Figure

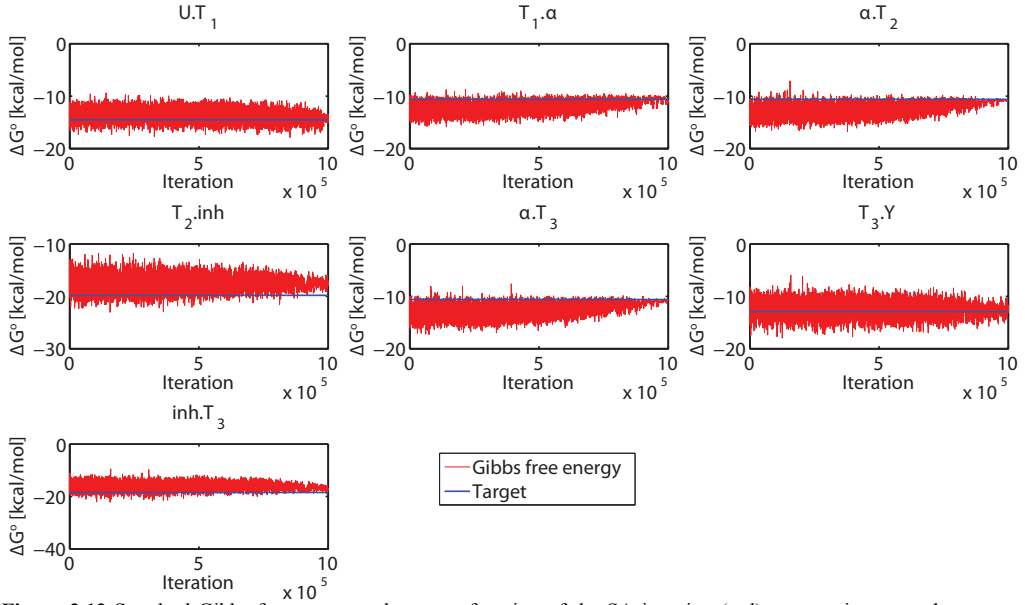


Figure 2.12 Standard Gibbs free energy values as a function of the SA iteration (red), converging towards a preset target standard Gibbs free energy value obtained from robustness analysis (blue) for all duplexes in the IFFL circuit.

Table 2.2 Designed Gibbs free energies of hybridization of the Oligator and the IFFL circuit ($\Delta G_{AB,target}^0$) following from robustness analysis, and the optimized values after DNA sequence design ($\Delta G_{AB,NUPACK}^0$), all expressed in kcal/mol. Simulations of these values are compared in Figure 2.11.

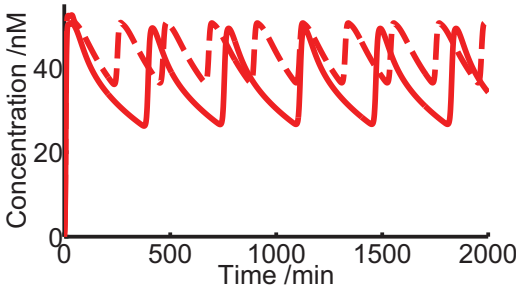
| <i>Oligator</i> | | | <i>IFFL circuit</i> | | |
|--------------------|--------------------------|--------------------------|---------------------|--------------------------|--------------------------|
| dsDNA pair | $\Delta G_{AB,target}^0$ | $\Delta G_{AB,NUPACK}^0$ | dsDNA pair | $\Delta G_{AB,target}^0$ | $\Delta G_{AB,NUPACK}^0$ |
| $\alpha \cdot T_1$ | -12.03 | -11.87 | $U \cdot T_1$ | -14.50 | -14.35 |
| $T_1 \cdot \alpha$ | -12.03 | -11.87 | $T_1 \cdot \alpha$ | -10.61 | -11.68 |
| $\alpha \cdot T_2$ | -12.03 | -11.68 | $\alpha \cdot T_2$ | -10.61 | -11.56 |
| $T_2 \cdot \beta$ | -13.88 | -13.55 | $\alpha \cdot T_3$ | -10.61 | -11.21 |
| $\beta \cdot T_3$ | -13.88 | -13.77 | $T_2 \cdot inh$ | -19.80 | -17.97 |
| $T_3 \cdot inh$ | -17.88 | -17.59 | $T_3 \cdot Y$ | -12.89 | -12.13 |
| $inh \cdot T_1$ | -17.35 | -16.08 | $inh \cdot T_3$ | -18.44 | -17.32 |

2.13. The approach has been empirically optimized to design PEN based systems, *i.e.* for the design of short oligomers. However, it is easily extended to systems with longer DNA molecules or even RNA. In the sequence optimization, the balance between the contributions to the error function (*i.e.* the regularization parameters η , ν and λ in equations (2.6), (2.5) and (2.7), respectively) could also be automatically optimized by considering L-curves⁷².

Table 2.3 DNA Sequences for the Oligator and the IFFL circuit resulting from the automated design procedure. Marked bases (*) are protected from degradation by exonuclease through phosphorothioate modifications on their 5' backbones. Bold bases belong to the nickase recognition sites, and *italic* bases are the inhibitory part on the inhibitor strand and corresponding template strands.

| <i>Oligator</i> | | <i>IFFL circuit</i> | |
|-----------------|---|---------------------|-------------------------------------|
| Oligomer | Sequence | Oligomer | Sequence |
| T_1 | C*A*T*T GACTCTG - CATT GACTCTG | T_1 | A*T*T*AGACTCAC- AAACGACTCCTA |
| T_2 | T*A*C*CGACTCCA- CATTGACTCTG | T_2 | A*T*C*CCAACATTAGACT- ATTAGACTCAC |
| T_3 | A*C*A*CTCTGCATTGACT- TACCGACTCCA | T_3 | A*A*A*TACCCAAC- ATTAGACTCAC |
| α | CAGAGTCAATG | U | T*A*G*GAGTCGTTT |
| β | TGGAGTCGGTA | α | GTGAGTCTAAT |
| <i>inh</i> | AGTCAATGCAGAGTGT | Y | GTTGGGTATTT |
| | | <i>inh</i> | AGTCTAATGTTGGGAT |

A Oligator



B IFFL circuit

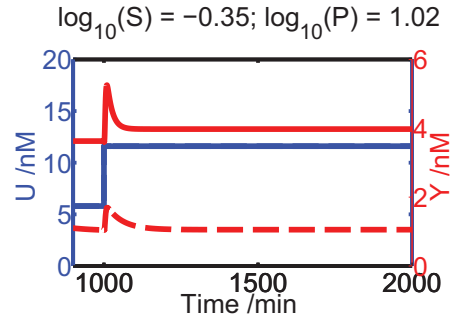


Figure 2.13 Circuit response generated by employing hybridization Gibbs free energies corresponding to the most robust parameter set (dashed red lines) and hybridization Gibbs free energies obtained after DNA sequence design (solid red lines). Initial concentrations of all species are the same for both responses, but Gibbs free energies of hybridization are different (Table 2.1). (A) The responses of the Oligator in terms of dsDNA concentration. The initial concentrations for these simulations are $[\alpha]_0 = 0.9$ nM, $[T_1] = 42$ nM, $[T_2] = 6$ nM and $[T_3] = 9$ nM. (B) The responses of the IFFL circuit in terms of output ssDNA Y . The initial concentrations for these simulations are $[U]_0 = 5.8$ nM, $[T_1] = 22$ nM, $[T_2] = 11$ nM and $[T_3] = 182$ nM, and the Gibbs free energies obtained from robustness analysis and from DNA sequence optimization are depicted in the right column of Table 2.1.

2.4 CONCLUSION

Herein we have shown the successful automated design of BRNs constructed from the PEN toolbox. The procedure is fully automated, from establishing the sets of ODEs to the optimization of the DNA sequences. An important asset of this approach is that it couples a directed exploration of the phase space and calculated robustness measures of parameters to a physical implementation of the desired circuit that is optimized to account for unexpected perturbations. Indeed, this coupling ensures that the expected functionality of the circuit is preserved after optimization of the DNA sequences. The automated design methodology has shown to successfully reproduce the Oligator system and is applied to the design of a new type of PEN-based circuit based on a type-1 IFFL

motif. This approach can easily be extended such that it can be applied to different types of *in vitro* synthetic toolboxes used for engineering reaction networks, e.g. those that are based on RNA, on transcription-translation machinery or only on enzymes as discussed in Chapter 1. Moreover, the associated analyses also reveal critical interactions in the designed networks. In the next Chapter, we thoroughly analyze the IFFL circuit based on new insights that are obtained through experimental characterization of this network.

2.5 REFERENCES

1. S. Mann (2012), Systems of creation: the emergence of life from nonliving matter. *Acc. Chem. Res.* **45**, 2131–2141.
2. S. M. Rafelski and W. F. Marshall (2008), Building the cell: design principles of cellular architecture. *Nat. Rev. Mol. Cell Biol.* **9**, 593–602.
3. B. R. Fritz, L. E. Timmerman, N. M. Daringer, J. N. Leonard and M. C. Jewett (2010), Biology by design: from top to bottom and back. *J. Biomed. Biotechnol.* **2010**, 232016.
4. S. N. Semenov, A. J. Markvoort, W. B. L. Gevers, A. Piruska, T. F. A. de Greef and W. T. S. Huck (2013), Ultrasensitivity by molecular titration in spatially propagating enzymatic reactions. *Biophys. J.* **105**, 1057–1066.
5. S. N. Semenov, A. J. Markvoort, T. F. A. de Greef and W. T. S. Huck (2014), Threshold sensing through a synthetic enzymatic reaction–diffusion network. *Angew. Chem., Int. Ed.* **53**, 8066–8069.
6. M. Behar and A. Hoffmann (2010), Understanding the temporal codes of intra-cellular signals. *Curr. Opin. Genet. Dev.* **20**, 684–693.
7. M. B. Elowitz and S. Leibler (2000), A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338.
8. T. S. Gardner, C. R. Cantor and J. J. Collins (2000), Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342.
9. J. J. Tabor, H. M. Salis, Z. B. Simpson, A. A. Chevalier, A. Levskaya, E. M. Marcotte, C. A. Voigt and A. D. Ellington (2009), A synthetic genetic edge detection program. *Cell* **137**, 1272–1281.
10. P. Schwillie and S. Diez (2009), Synthetic biology of minimal systems. *Crit. Rev. Biochem. Mol. Biol.* **44**, 223–242.
11. P. L. Luisi and P. Stano (2011), Synthetic biology: minimal cell mimicry. *Nat. Chem.* **3**, 755–756.
12. D. Soloveichik, G. Seelig and E. Winfree (2010), DNA as a universal substrate for chemical kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5393–5398.
13. Y.-J. Chen, N. Dalchau, N. Srinivas, A. Phillips, L. Cardelli, D. Soloveichik and G. Seelig (2013), Programmable chemical controllers made from DNA. *Nat. Nanotechnol.* **8**, 755–762.
14. A. C. Forster and G. M. Church (2007), Synthetic biology projects *in vitro*. *Genome Res.* **17**, 1–6.
15. C. E. Hodgman and M. C. Jewett (2012), Cell-free synthetic biology: thinking outside the cell. *Metab. Eng.* **14**, 261–269.
16. V. Noireaux, Y. T. Maeda and A. Libchaber (2011), Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3473–3480.
17. H. Niederholtmeyer, V. Stepanova and S. J. Maerkl (2013), Implementation of cell-free biological networks at steady state. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15985–15990.
18. D. Siegal-Gaskins, Z. A. Tuza, J. Kim, V. Noireaux and R. M. Murray (2014), Gene circuit performance characterization and resource usage in a cell-free ‘breadboard’. *ACS Synth. Biol.* **3**, 416–425.
19. E. Franco, G. Giordano, P.-O. Forsberg and R. M. Murray (2014), Negative autoregulation matches production and demand in synthetic transcriptional networks. *ACS Synth. Biol.* **3**, 589–599.
20. A. Padirac, T. Fujii and Y. Rondelez (2013), Nucleic acids for the rational design of reaction circuits. *Curr. Opin. Biotechnol.* **24**, 575–580.
21. B. Yordanov, J. Kim, R. L. Petersen, A. Shudy, V. V. Kulkarni and A. Phillips (2014), Computational design of nucleic acid feedback control circuits. *ACS Synth. Biol.* **3**, 600–616.
22. J. Kim, K. S. White and E. Winfree (2006), Construction of an *in vitro* bistable circuit from synthetic transcriptional switches. *Mol. Syst. Biol.* **2**, 68.
23. P. Subsoontorn, J. Kim and E. Winfree (2012), Ensemble Bayesian analysis of bistability in a synthetic transcriptional switch. *ACS Synth. Biol.* **1**, 299–316.
24. J. Kim and E. Winfree (2011), Synthetic *in vitro* transcriptional oscillators. *Mol. Syst. Biol.* **7**, 465.
25. E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree and F. C. Simmel (2011), Timing molecular motion and production with a synthetic transcriptional clock. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E784–E793.
26. J. Kim, I. Khetarpal, S. Sen and R. M. Murray (2014), Synthetic circuit for exact adaptation and fold-change detection. *Nucleic Acids Res.* **42**, 6078–6089.

27. J. W. Szostak, D. P. Bartel and P. L. Luisi (2001), Synthesizing life. *Nature* **409**, 387–390.
28. P. M. Gardner, K. Winzer and B. G. Davis (2009), Sugar synthesis in a protocellular model leads to a cell signalling response in bacteria. *Nat. Chem.* **1**, 377–383.
29. M. Weitz, J. Kim, K. Kapsner, E. Winfree, E. Franco and F. C. Simmel (2014), Diversity in the dynamical behavior of a compartmentalized programmable biochemical oscillator. *Nat. Chem.* **6**, 295–302.
30. K. Montagne, R. Plasson, Y. Sakai, T. Fujii and Y. Rondelez (2011), Programming an *in vitro* DNA oscillator using a molecular networking strategy. *Mol. Syst. Biol.* **7**, 466.
31. A. Padirac, T. Fujii and Y. Rondelez (2012), Bottom-up construction of *in vitro* switchable memories. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3212–E3220.
32. A. Padirac, T. Fujii, A. Estévez-Torres and Y. Rondelez (2013), Spatial waves in synthetic biochemical networks. *J. Am. Chem. Soc.* **135**, 14586–14592.
33. K. Hasatani, M. Leocmach, A. J. Genot, A. Estévez-Torres, T. Fujii and Y. Rondelez (2013), High-throughput and long-term observation of compartmentalized biochemical oscillators. *Chem. Commun.* **49**, 8090–8092.
34. T. Fujii and Y. Rondelez (2013), Predator–prey molecular ecosystems. *ACS Nano* **7**, 27–34.
35. N. Aubert, C. Mosca, T. Fujii, M. Hagiya and Y. Rondelez (2014), Computer-assisted design for scaling up systems based on DNA reaction networks. *J. R. Soc. Interface* **11**, 20131167.
36. A. Baccouche, K. Montagne, A. Padirac, T. Fujii and Y. Rondelez (2014), Dynamic DNA-toolbox reaction circuits: a walkthrough. *Methods* **67**, 234–249.
37. D. Y. Zhang and E. Winfree (2009), Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131**, 17303–17314.
38. W. Ma, A. Trusina, H. El-Samad, W. A. Lim and C. Tang (2009), Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760–773.
39. N. Barkai and S. Leibler (1997), Robustness in simple biochemical networks. *Nature* **387**, 913–917.
40. A. B. Artyukhin, L. F. Wu and S. J. Altschuler (2009), Only two ways to achieve perfection. *Cell* **138**, 619–621.
41. D. Muzzey, C. A. Gómez-Urbe, J. T. Mettetal and A. van Oudenaarden (2009), A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* **138**, 160–171.
42. U. Alon (2007), Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461.
43. B. Novák and J. J. Tyson (2008), Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9**, 981–991.
44. J. Van Ness, L. K. Van Ness and D. J. Galas (2003), Isothermal reactions for the amplification of oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4504–4509.
45. M. Hafner, H. Koeppl, M. Hasler and A. Wagner (2009), ‘Glocal’ robustness analysis and model discrimination for circadian oscillators. *PLoS Comput. Biol.* **5**, e1000534.
46. E. Zamora-Sillero, M. Hafner, A. Ibig, J. Stelling and A. Wagner (2011), Efficient characterization of high-dimensional parameter spaces for systems biology. *BMC Syst. Biol.* **5**, 142.
47. C. F. Lopez, J. L. Muhlich, J. A. Bachman and P. K. Sorger (2013), Programming biological models in Python using PySB. *Mol. Syst. Biol.* **9**, 646.
48. J. Vanlier, C. A. Tiemann, P. A. J. Hilbers and N. A. W. van Riel (2012), An integrated strategy for prediction uncertainty analysis. *Bioinformatics* **28**, 1130–1135.
49. J. Vanlier, C. A. Tiemann, P. A. J. Hilbers and N. A. W. van Riel (2012), A Bayesian approach to targeted experiment design. *Bioinformatics* **28**, 1136–1142.
50. A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker and C. S. Woodward (2005), SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Software* **31**, 363–396.
51. G. Rodrigo, J. Carrera and A. Jaramillo (2007), Genetdes: automatic design of transcriptional networks. *Bioinformatics* **23**, 1857–1858.
52. R. Dirks, J. Bois, J. Schaeffer, E. Winfree and N. Pierce (2007), Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.* **49**, 65–88.
53. J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks and N. A. Pierce (2011), NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173.
54. N. C. Seeman (1982), Nucleic acid junctions and lattices. *J. Theor. Biol.* **99**, 237–247.
55. N. C. Seeman and N. R. Kallenbach (1983), Design of immobile nucleic acid junctions. *Biophys. J.* **44**, 201–209.
56. N. C. Seeman (1990), *De novo* design of sequences for nucleic acid structural engineering. *J. Biomol. Struct. Dyn.* **8**, 573–581.
57. U. Feldkamp, S. Saghafi, W. Banzhaf and H. Rauhe (2002), DNASequencesGenerator: a program for the construction of DNA sequences. In *DNA Computing* (eds. N. Jonoska and N. C. Seeman), pp. 23–32. Springer.
58. K. A. Afonin, E. Bindewald, A. J. Yaghoubian, N. Voss, E. Jacovetty, B. A. Shapiro and L. Jaeger (2010), *In vitro* assembly of cubic RNA-based scaffolds designed *in silico*. *Nat. Nanotechnol.* **5**, 676–682.
59. L. Kaderali, A. Deshpande, J. P. Nolan and P. S. White (2003), Primer-design for multiplexed genotyping. *Nucleic Acids Res.* **31**, 1796–1802.
60. A. Kick, M. Bönsch and M. Mertig (2012), EGNAS: an exhaustive DNA sequence design algorithm. *BMC*

Bioinformatics **13**, 138.

61. J. Qian, T. M. Ferguson, D. N. Shinde, A. J. Ramírez-Borrero, A. Hintze, C. Adami and A. Niemz (2012), Sequence dependence of isothermal DNA amplification via EXPAR. *Nucleic Acids Res.* **40**, e87.
62. T. Simonsson (2001), G-quadruplex DNA structures variations on a theme. *Biol. Chem.* **382**, 621–628.
63. S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi (1983), Optimization by simulated annealing. *Science* **220**, 671–680.
64. J. S. Kim, J.-W. Lee, Y.-K. Noh, J.-Y. Park, D.-Y. Lee, K.-A. Yang, Y. G. Chai, J. C. Kim and B.-T. Zhang (2008), An evolutionary Monte Carlo algorithm for predicting DNA hybridization. *Biosystems* **91**, 69–75.
65. K. J. Breslauer, R. Frank, H. Blöcker and L. A. Marky (1986), Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3746–3750.
66. J. SantaLucia Jr. and D. Hicks (2004), The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415–440.
67. J. SantaLucia Jr., H. T. Allawi and P. A. Seneviratne (1996), Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**, 3555–3562.
68. S. Bommarito, N. Peyret and J. SantaLucia Jr. (2000), Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* **28**, 1929–1934.
69. G. Cellière, G. Fengos, M. Hervé and D. Iber (2011), The plasticity of TGF- β signaling. *BMC Syst. Biol.* **5**, 184.
70. S. Basu, R. Mehreja, S. Thiberge, M.-T. Chen and R. Weiss (2004), Spatiotemporal control of gene expression with pulse-generating networks. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6355–6360.
71. M. Hafner, T. Petrov, J. Lu and H. Koepl (2011), Rational design of robust biomolecular circuits: From specification to parameters. In *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology* (eds. H. Koepl, G. Setti, D. Densmore and M. di Bernardo), pp. 253–279. Springer.
72. C. L. Lawson and R. J. Hanson (1995), *Solving Least Squares Problems*. SIAM.

Chapter 3

Adaptation Dynamics in a PEN-Based Incoherent Feedforward Loop

This work has been published:

Hendrik W. H. van Roekel, Lenny H. H. Meijer, Saeed Masroor, Zandra C. Félix Garza, AndréEstévez-Torres, Yannick Rondelez, Antonios Zagaris, Mark A. Peletier, Peter A. J. Hilbers and Tom F. A. de Greef (2015), Automated design of programmable enzyme-driven DNA circuits. *ACS Synthetic Biology* **4**, 735–745.

ABSTRACT

The sharing of catalytic resources among substrates in biochemical reaction networks can significantly affect the predicted response of such a system. In this Chapter, we first experimentally characterize the enzymatic platform of the PEN toolbox as it is intended to be incorporated in the incoherent feedforward loop (IFFL) circuit introduced in the previous Chapter. Importantly, this characterization reveals that the exonuclease enzyme, besides attaching to and degrading single-stranded DNA, also attaches to 5' phosphorothioate protected single-stranded DNA. Since the enzyme lacks the ability to degrade these protected strands, the protected template and input strands are considered as competitive inhibitors for the activity of exonuclease. This observation is incorporated in the computational model describing the IFFL circuit. The strong nonlinearities in the degradation of active components caused by this unintended cross-coupling are shown computationally to have a positive effect on adaptation quality. We mathematically analyze this effect with the use of toy models and present a novel mechanism responsible for increased circuit response.

3.1 INTRODUCTION

Catalytic resources in living cells are dedicated to the production of proteins which ensure the cell's survival and development by providing structural stability, transport of molecules and renewal of enzymes¹. Over the last decades, it has become apparent that protein synthesis is often a tightly regulated link in an orchestrated genetic network² responsible for *e.g.* the processing of external information and decision making in the cell cycle^{3,4}. Regulation occurs on many levels, including transcriptional^{5,6}, translational^{7,8} and post-translational^{9–11}. In such complex regulatory systems, proteins function as the computational elements¹², a feature of these molecules that is exploited in molecular programming and synthetic biology (Chapter 1).

Man-made synthetic biochemical reaction networks, like their biological inspiration, also often rely on catalytic resources to maintain a far-from-equilibrium state. In many cases, purified enzymes take the role of the catalytic platform. The introduction of enzymes to biochemical reaction networks (BRNs) comes with benefits as well as disadvantages. The use of enzymes is beneficial in the sense that they exhibit high selectivity for substrates and specificity in their catalytic action^{13–15}, and their saturation kinetics (*i.e.* Michaelis-Menten)¹⁶ allows for non-linear dynamical behaviors such as zero-order ultrasensitivity^{17,18}, vastly increasing the variety of dynamics that can be observed as a result of the interplay of enzymatic activities. Contrarily, the fact that enzymes are difficult to unambiguously characterize, both in terms of their kinetic parameters as well as in coupling effects that result from the enzymes' affinities for different substrates^{19,20} or substrate-like molecules—especially those unforeseen—can be a disadvantage.

In the previous Chapter, we have discussed the automated design of synthetic biochemical systems. Particularly, we described the engineering of BRNs using the PEN toolbox that operates with information-bearing molecules (DNA) and is enriched with a catalytic platform in the form of three enzymes. Whereas DNA hybridization reactions are modeled using mass-action kinetics, enzymatic activity is modeled to operate with saturation kinetics, as is common practice when describing enzyme-catalyzed reactions¹⁶. The resulting kinetic model is applied as a predictive tool. Even though kinetic models can differ in type, style and application, many models that are applied as predictive tools usually consist of large numbers of parameters that are either estimated from experimental data or obtained from literature. The desire where predictive models are concerned is that they generate reliable, robust predictions. However, in this Chapter we also introduce two intuitive toy models, but rather than utilize these as predictive tools, we employ them to help us rationalize experimental and computational findings in an accessible and analytical manner.

In this Chapter, we first experimentally explore the IFFL circuit from Chapter 2. Based on the results, the large kinetic model that is generated in Chapter 2 (equations (A.5)-(A.8) in Appendix section A.1) for the PEN-based IFFL circuit is extended to account for unproductive sequestration of the exonuclease enzyme. We reveal and characterize a previously unexplored reaction in the degradation kinetics of the components caused by unproductive sequestration of exonuclease by phosphorothioate modified single-stranded DNA (ssDNA) strands. Through simulation and mathematical analysis we demonstrate that the additional load on the degradation machinery results in an efficient amplification mechanism that increases the sensitivity of the computed adaptive response and enhances the robustness of the circuit under parametric uncertainty. Finally, we argue

that the deviation from the prediction obtained in Chapter 2 as a result of loading the catalytic resources does not impair the rational design of *in vitro* biomolecular networks.

3.2 MATERIALS AND METHODS

3.2.1 Experimental characterization of the PEN toolbox-based IFFL circuit

Reagents

Oligonucleotides were obtained from Integrated DNA Technologies (IDTDNA) and were purified using High Performance Liquid Chromatography (HPLC). All experiments for the enzymatic characterizations as described in this section were performed by L. H. H. Meijer, MSc. Three phosphorothioate backbone modifications at the 5' ends of the templates and input strand *U* of the IFFL circuit prevent their degradation by exonuclease. Furthermore, since 3'-OH can be extended by DNA polymerase, the templates were obtained with a phosphate modification at their 3'-ends to prevent circuit leakage²¹. Concentrations of DNA were verified using NanoDrop Spectrophotometer ND-100.

Enzymatic degradation of non-protected primers is induced by ttRecJ^{22,23}, a thermophilic equivalent of the RecJ enzyme from *Thermus thermophilus*. This enzyme has the advantage that it does not require a high excess of trehalose for stabilization as does the RecJ_f exonuclease that was used in previous work²⁴. A batch of this 5' → 3' processive exonuclease was obtained from Ryoji Masui. The batch was diluted in diluent D (New England Biolabs, *i.e.* NEB) to form a stock solution with a concentration of 2.5 μM.

Bst. polymerase was obtained from NEB. The working stock of this DNA polymerase was prepared by dissolving the original stock of 8000 U mL⁻¹ in diluent A complemented with 0.1% Triton X-100 to form a stock solution of 256 U mL⁻¹ (36.2 nM).

Nicking endonuclease Nt.BstNBI (*i.e.* nickase) was obtained from NEB and characterized as described before²⁵. Since its kinetics did not show significant deviation from literature values, we chose to adhere to the original reported values. Hence, kinetic characterization of nickase is not described in this Thesis.

Experimental characterization of the kinetics of Bst. polymerase

The action of polymerase, shown in Figure 3.1, is described by Michealis-Menten kinetics and can be monitored by measuring the increase in fluorescence from the double-strand specific intercalator EvaGreen. During the studies on the kinetics of Bst. polymerase, we used a master mix that contains 20 mM Tris-HCl, 10 mM KCl, 50 mM NaCl, 10 mM (NH₄)₂SO₄, 8 mM MgSO₄, 0.1% Triton X-100, 400 μM of each deoxyribonucleotide triphosphate (dNTP; NEB), 0.1% Synperonic F108 (Sigma Aldrich), 2 μM Netropsin (Sigma Aldrich), 500 μg mL⁻¹ Bovine Serum Albumine (BSA; NEB), 4 mM dithiothreitol (DTT) and 1x EvaGreen (1.33 μM; VWR) at a pH of 8.8.²⁶ Nuclease free water and DNA strands in TE buffer, in minimal volumes, were added to the eppendorfs (white real-time Polymerase Chain Reaction (PCR) tube strips, Eppendorf). The primer concentration was in excess (1000 nM) compared to the template which was varied over a range of 10-340 nM. To

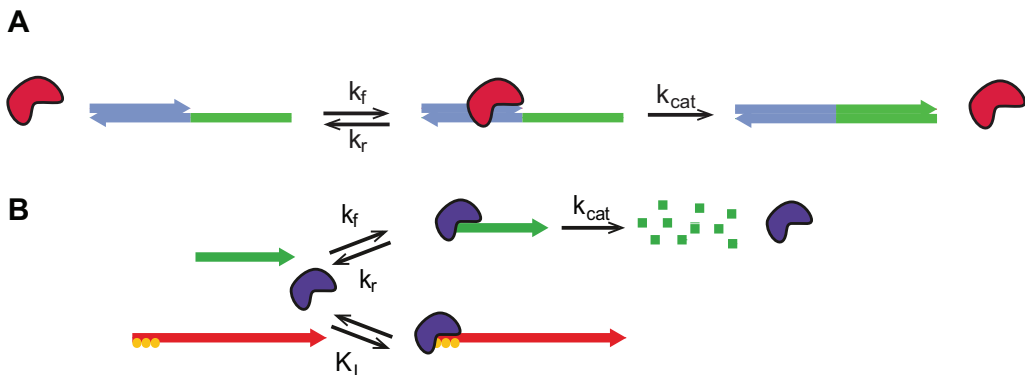


Figure 3.1 Schematic representation of enzyme kinetics in this study. (A) The action of polymerase. Primer (gray) bounded to template (gray-green) serves as substrate for polymerase (red). The polymerization of primer-template substrates can be monitored using the fluorescent dye EvaGreen. Parameters k_f , k_r and k_{cat} are enzymatic rate constants. (B) The action of exonuclease. The degradation of a ssDNA (green) by exonuclease ttRecJ (purple) is shown in the top pathway. This reaction is accompanied by the association of exonuclease ttRecJ to protected oligomer (red) which acts as a competitive inhibitor (bottom pathway). Phosphorothioate modifications at the 5' end are indicated by yellow circles. Parameters k_f , k_r and k_{cat} are enzymatic rate constants and K_i is the dissociation constant of exonuclease with protected oligomer.

disrupt possible aggregates, the DNA strands were annealed from 70°C to room temperature. Subsequently, master mix containing all components (except for dNTPs) with polymerase (0.64 U mL^{-1}) was added, and the eppendorfs were placed at 42°C for at least 10 minutes to equilibrate the mixtures. Finally, giving a total volume of 50 μL , dNTP solution was added to obtain a final concentration of 400 μM of each dNTP. The eppendorfs were placed in the preheated PCR system (CFX96 from Bio-Rad) at a temperature of 42°C and with a lid temperature of 70°C to prevent condensation. The relative fluorescence units (RFUs) were measured over cycles of 12 seconds. After each assembling step the solution was vorticed for at least 10 seconds, and mixtures in eppendorfs were spun down afterwards.

Experimental characterization of the kinetics of exonuclease ttRecJ

Exonuclease ttRecJ processively hydrolyzes ssDNA, therefore, we can describe the action of ttRecJ, shown in Figure 3.1B, by Michaelis-Menten kinetics. The ssDNA strands with backbone phosphorothioate modifications are not degraded by ttRecJ, but experiments indicated that they do serve as a substrate for ttRecJ. Hence, phosphorothioate modified DNA strands act as competitive inhibitors for the degradation of non-protected oligonucleotides, which is quantified by a dissociation constant, K_i . Although EvaGreen is a double-strand specific fluorescent intercalator, the fluorescence induced by single-stranded oligonucleotides is sufficient to observe variations in their concentrations. Hence, we were able to monitor the action of ttRecJ by measuring fluorescence over time.

During the kinetic studies on exonuclease kinetics, the master mix (*vide supra*) with 2x EvaGreen (2.66 μM ; VWR) and 1x ROX reference dye was used at a pH of 8.8.²⁶ First, nuclease free water and a solution of protected ssDNA (competitor) in TE buffer, in minimal volumes, were added to the eppendorfs. To disrupt possible aggregates, solutions were annealed from 70 °C to

room temperature. Then, master mix and 25 nM exonuclease ttRecJ were added, and the mixtures were equilibrated for 10 minutes at 42°C. Lastly, the non-protected oligonucleotide (~2000 nM) was added leading to a final volume of 20 µL. After all components had been assembled, the eppendorfs were placed in the PCR system which was preheated to 42°C with a lid temperature of 70°C to prevent condensation. The RFUs were measured over cycles of 19 seconds. Samples were prepared at least in triplet and the ROX reference dye was used to normalize the fluorescent signal from EvaGreen. After each assembling step solutions were vorticed for at least 10 seconds, and mixtures in eppendorfs were spun down afterwards.

3.2.2 Incorporation of unproductive sequestration of exonuclease by protected ssDNA strands into the IFFL model

We investigated how unproductive sequestration of the degradation machinery by protected ssDNA strands influences the adaptive response of the DNA-based IFFL. Previous reports have quantified how resource competition coupled with enzyme saturation in genetic networks can lead to ultrasensitive responses²⁷, allows for predictable tuning of bistable switches^{28,29} and presents a versatile mechanism to synchronize and tune oscillating circuits^{30–33}. However, the influence of resource competition on adaptation dynamics has, to the best of our knowledge, not been reported. We expand the ODE model describing the PEN implementation of the IFFL circuit to explicitly include exonuclease kinetics. However, all association reactions with exonuclease are modeled using mass-action kinetics rather than Michealis-Menten kinetics. We incorporated explicit exonuclease kinetics, as the concentration of this enzyme reflects its direct availability, rather than an estimate based on the Michaelis-Menten ratio. The reaction equations that are added to the model are depicted in equation (3.1). Here, the forward reaction rate $k_{f,exoN}$ is assumed to be equal for all forward reactions. The reverse rate constants follow from the forward rate constant $k_{f,exoN}$ and the measured dissociation constants (K_I) and Michealis-Menten constants (K_M) (Table 3.1), allowing for adequate mass-action modeling of ttRecJ exonuclease activity to be incorporated in the detailed kinetic model. Moreover, we found that the parameter $k_{f,exoN}$ is not critical for the response of the model, provided the reverse rate constants do not become negative. The resulting ODEs are straightforwardly derived and depicted in equation (A.9) in Appendix section A.2.

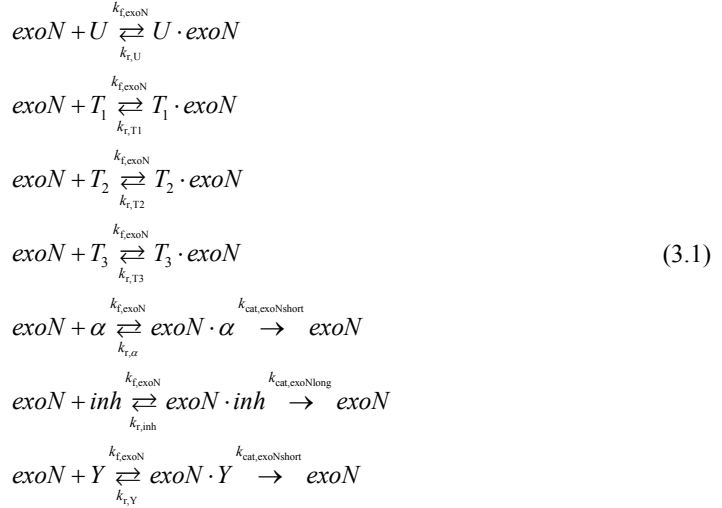


Table 3.1 Expressions for the parameters as they are incorporated in the extension of the model, equation (3.1).

| Parameter expression |
|---|
| $k_{r,U} = k_{f,\text{exoN}} K_{I,U}$ |
| $k_{r,T1} = k_{f,\text{exoN}} K_{I,T1}$ |
| $k_{r,T2} = k_{f,\text{exoN}} K_{I,T2}$ |
| $k_{r,T3} = k_{f,\text{exoN}} K_{I,T3}$ |
| $k_{r,\alpha} = k_{r,Y} = k_{f,\text{exoN}} K_{M,\text{exoNshort}}$ |
| $k_{r,\text{inh}} = k_{f,\text{exoN}} K_{M,\text{exoNlong}}$ |

3.2.3 Toy models to analyze the effect of unproductive sequestration of exonuclease

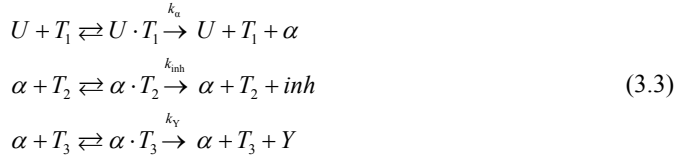
The ODE models presented earlier are kinetically detailed and provide stringent criteria for DNA sequence design, but their level of complexity is too high to enable a conceptual understanding of the nonlinear dynamics underlying the adaptation mechanism. To that end, we constructed two toy models lacking detailed kinetics but allowing one to explore and rationalize qualitative behavior. The first toy model (TMI) is a compact representation of the original design of the IFFL circuit (Chapter 2), whereas the second one (TMII) additionally includes resource competition by unproductive sequestration of the protected, nondegradable input strand by exonuclease. It should be noted that the three-node networks analyzed by Tang and co-workers¹⁸—contrary to our three-node circuits—are based on reactions that describe enzymatic conversions from a conserved pool, that is, each node of the circuit relates to a constant total concentration of enzyme that is interconverted between an active and an inactive form. The simplified ODE models analyzed in this Chapter are fundamentally different as the interacting species α , inh , and Y are constantly produced (by polymerase from dNTPs) and degraded (by exonuclease) over time.

Toy model I

TMI is described in equation (3.2), where the production of the activator (α), inhibitor (inh), and output (Y) is modeled by Michaelis–Menten kinetics and their degradation by first-order kinetics^{25,26}. In these equations, U represents the total concentration of input strand, whereas α , inh , and Y are the concentrations of the activating, inhibiting, and output species in the IFFL circuit. T_1 , T_2 , and T_3 are the concentrations of the template strands, while the parameters K_U , K_α , and K_{inh} are Michaelis-Menten constants—roughly equal to the hybridization dissociation constants K_d^U , K_d^α , and $K_d^{inhT2} \approx K_d^{inhT3}$ in the large ODE model, respectively. The rate constants k_α , k_{inh} , and k_Y are catalytic rates depending nontrivially on nickase and polymerase enzyme kinetics, while D represents the first-order degradation rate constant.

$$\begin{aligned}
 \frac{d\alpha}{dt} &= \frac{k_\alpha T_1 U}{K_U \left(1 + \frac{U}{K_U}\right)} - D\alpha \\
 \frac{dinh}{dt} &= \frac{k_{inh} T_2 \alpha}{K_\alpha \left(1 + \frac{\alpha}{K_\alpha}\right)} - Din h \\
 \frac{dY}{dt} &= \frac{k_Y T_3 \alpha}{K_\alpha \left(1 + \frac{\alpha}{K_\alpha} + \frac{inh}{K_{inh}}\right)} - DY
 \end{aligned} \tag{3.2}$$

This toy model is based on three main functions, *i.e.* amplification, competitive inhibition, and degradation. Degradation is modeled with first-order kinetics. There are three amplification reactions producing primers α , inh , and Y , which comprise mechanism (3.3).



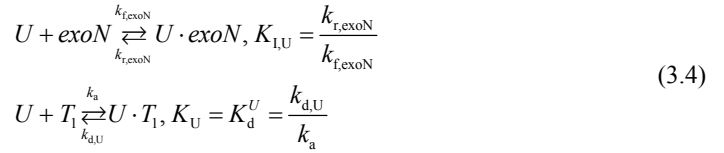
Similar to Rondelez and coworkers^{25,26}, we describe these reactions by Michaelis-Menten kinetics in which the templates are treated as enzymes converting a substrate to a product, *e.g.* U to α . Although Michaelis-Menten kinetics is not formally correct here, it is a suitable empirical approximation. In the production rate of α in TMI, we have implicitly incorporated the concentration of the complex $U \cdot T_1$ using quasi-steady state approximation in the Michaelis-Menten expression $U \cdot T_1 = T_1 U / (K_U + U)$. In this expression, we have defined $U \stackrel{\text{def}}{=} U_{\text{tot}}$. Namely, it is $U \cdot T_1$ that drives the formation of α (analogously to the detailed model), and the quasi-steady state approximation employed by Michaelis-Menten kinetics allows $U \cdot T_1$ to be expressed in terms of total U .

The network is designed in such a way that both Y and inh can bind to T_3 . In terms of Michaelis-Menten kinetics, this results in competition between these two species and is modeled by

introducing an additional term into the denominator of the amplification term of Y in the form of competitive inhibition in which K_{inh} is the dissociation constant of inh binding to T_3 .

Toy model II

The second toy model, TMII, is a modification of TMI incorporating the observation that protected species are able to unproductively bind exonuclease, which results in enzyme-limited competition. Here, we only consider the binding between input strands and exonuclease, neglecting binding of protected templates. This is in agreement with TMI, where changes in template concentrations were not taken into account. Specifically, we incorporate unproductive sequestration by introducing competition between exonuclease and template T_1 for protected input strand (mechanism is depicted in equation (3.4)).



TMII uses the free concentration of input strand, that is, $U_{free} = U - U \cdot exoN - U \cdot T_1$, as a competitive inhibitor in the degradation terms. Amplification of α by T_1 is also initiated by the free form of input ssDNA.

Degradation in TMII

Apart from the primers α , inh , and Y , also the protected primer U and all three templates can attach to exonuclease. The degradation rate of e.g. α is given in equation (3.5).

$$r_{\alpha \triangleright \emptyset} = \frac{V_{\max,exoN}[\alpha]}{K_{M,exoN} \left(1 + \frac{[U]}{K_{I,U}} + \frac{[T_1] + [T_2] + [T_3]}{K_{I,T}} + \frac{[\alpha] + [Y] + [inh]}{K_{M,exoN}} \right)} \quad (3.5)$$

For the sake of simplicity, we neglect the effect of competition from templates in TMII and only consider the effect of sequestration of exonuclease by the input strand. Since $K_{M,exoN} \approx 200$ nM and $K_{I,U} \approx 1$ nM (see section 3.3), we can assume inequality (3.6)

$$\frac{[U]}{K_{I,U}} \gg \frac{[\alpha] + [Y] + [inh]}{K_{M,exoN}} \quad (3.6)$$

so that the degradation rate becomes (3.7).

$$r_{\alpha \triangleright \emptyset} = \frac{V_{\max,exoN}[\alpha]}{K_{M,exoN} \left(1 + \frac{[U]}{K_{I,U}} \right)} \quad (3.7)$$

Hence, TMII is described as in equation (3.8). We again note the difference to the simplified ODE models analyzed by Tang and colleagues¹⁸, as in TMII the input explicitly affects network dynamics by controlling the degradation rates due to sequestration of exonuclease activity.

$$\begin{aligned}
 \frac{d\alpha}{dt} &= \frac{k_a T_{1,\text{tot}} U_{\text{free}}}{K_U \left(1 + \frac{U_{\text{free}}}{K_U}\right)} - \frac{V_{\text{max,exoN}} \alpha}{K_{\text{M,exoN}} \left(1 + \frac{U_{\text{free}}}{K_{\text{I,U}}}\right)} \\
 \frac{dinh}{dt} &= \frac{k_{\text{inh}} T_2 \alpha}{K_a \left(1 + \frac{\alpha}{K_a}\right)} - \frac{V_{\text{max,exoN}} inh}{K_{\text{M,exoN}} \left(1 + \frac{U_{\text{free}}}{K_{\text{I,U}}}\right)} \\
 \frac{dY}{dt} &= \frac{k_Y T_3 \alpha}{K_a \left(1 + \frac{\alpha}{K_a} + \frac{inh}{K_{\text{inh}}}\right)} - \frac{V_{\text{max,exoN}} Y}{K_{\text{M,exoN}} \left(1 + \frac{U_{\text{free}}}{K_{\text{I,U}}}\right)}
 \end{aligned} \tag{3.8}$$

In these equations, $V_{\text{max,exoN}}$ is the maximum degradation rate and $K_{\text{M,exoN}}$ is the Michaelis–Menten constant of exonuclease, while $K_{\text{I,U}}$ is the dissociation constant of protected U unproductively binding to exonuclease. Sequestration is assumed to occur instantaneously; that is, U_{free} is obtained at each time-step by solving the algebraic equations corresponding to the steady state of equation (3.4). It was derived by S. Masroor, MSc that the free amount of U from these two reactions is the positive root of the third order polynomial $U_{\text{free}}^3 + aU_{\text{free}}^2 + bU_{\text{free}} + c = 0$, where a , b and c are given as

$$\begin{aligned}
 a &= K_U + K_{\text{I,U}} + T_1 + \text{exoN} - U \\
 b &= K_U (\text{exoN} - U) + K_{\text{I,U}} (\text{exoN} - U) + K_U K_{\text{I,U}} \\
 c &= -K_U K_{\text{I,U}} U
 \end{aligned} \tag{3.9}$$

in which $U \stackrel{\text{def}}{=} U_{\text{tot}}$.

3.3 RESULTS AND DISCUSSION

3.3.1 Enzyme kinetics

Kinetics of Bst. Polymerase

For every temporal datacurve corresponding to a different concentration of duplex the same procedure was followed. The initial part of the RFU *versus* time curve often contains an undershoot or an overshoot as a result of equilibration effects that occur in the samples. When such an artifact was observed, the initial data was removed, and an estimate of these datapoints was made by extrapolation. Thereafter, the RFU at the first timepoint, *i.e.* at $t = 0$, was subtracted from the datacurve, so that at $t = 0$ an RFU of zero was obtained. Then, the RFU was converted to the concentration of duplex by dividing the RFU datacurve with a factor obtained by dividing the final RFU by the initial concentration of template, which is equal to the final concentration of duplex as saturation occurs due to continuous activity of polymerase.

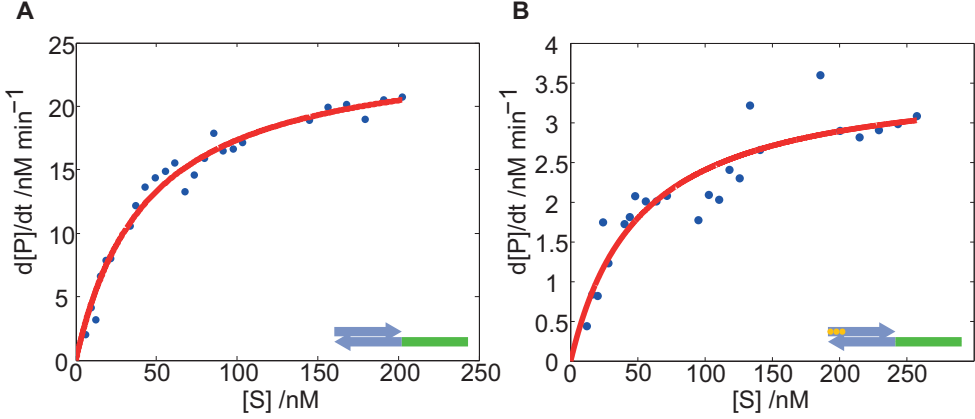


Figure 3.2 Experimental characterization of polymerase activity. Experiments were performed at 42 °C in master mix with 0.64 U mL⁻¹ Bst. Polymerase. RFU was measured every 12 seconds. Raw data was prepared as described in the text. The initial velocity ($d[P]/dt$) was determined by fitting a linear function to the RFU data curves, which was plotted versus the initial concentration of *primer:template* substrate (S) and fitted to the integrated Michaelis-Menten equation. The concentrations of *primer:template* were calculated using the thermodynamic parameters obtained from melting curves and concentrations of primer and template. (A) Polymerase activity using an unmodified primer (5'-TAGAGTCAAAC-3') hybridized to phosphorothioate modified template (5'-T*T*C*CCCCTCGA GTTTGACTCTA/3Phos/-3') as substrate (S). The Michaelis-Menten parameters V_{\max} and K_M obtained from least-squares fitting are 25.7 ± 0.8 nM min⁻¹ and 35.6 ± 3.1 nM, respectively. (B) Polymerase activity using a phosphorothioate modified primer (5'-C*A*T*GAGTCAGAT-3') hybridized to phosphorothioate modified template (5'-G*T*T*TGACUCTAATCTGACTCATG/3Phos/-3') as substrate. The Michaelis-Menten parameters V_{\max} and K_M obtained from least-squares fitting are 3.6 ± 0.2 nM min⁻¹ and 54.5 ± 8.3 nM, respectively.

We assume quasi-steady state conditions (Figure 3.1A) and Michaelis-Menten kinetics is used to interpret the obtained results. The initial slopes of the curves are determined by fitting a line and are plotted versus the concentrations of *primer:template*, which are calculated using the thermodynamic parameters obtained from melting curves and concentrations of primer and template (Figure 3.2). The resulting datapoints are used to fit the Michaelis-Menten equation (3.10).

$$\frac{d[P]}{dt} = -\frac{d[S]}{dt} = \frac{V_{\max}[S]}{K_M + [S]} \text{ with } K_M = \frac{k_r + k_{\text{cat}}}{k_f} \text{ and } V_{\max} = k_{\text{cat}}[E]_0 \quad (3.10)$$

To determine the parameters V_{\max} and K_M , optimization was performed using the Matlab routine `lsqnonlin` with a subpace trust-region method based on the interior-reflective Newton method. To prevent entrapment in local minima of the cost function, Latin Hypercube sampling was used to create twenty initial parameter vectors, with values in between an interval of 50% and 150% of the expected parameter value. From these twenty initial parameter values, we selected the parameter values that correspond to the lowest residual sum of squares. The lower bounds of the 95% confidence intervals and asymptotic standard errors were determined using the numerically estimated Fisher information matrix³⁴.

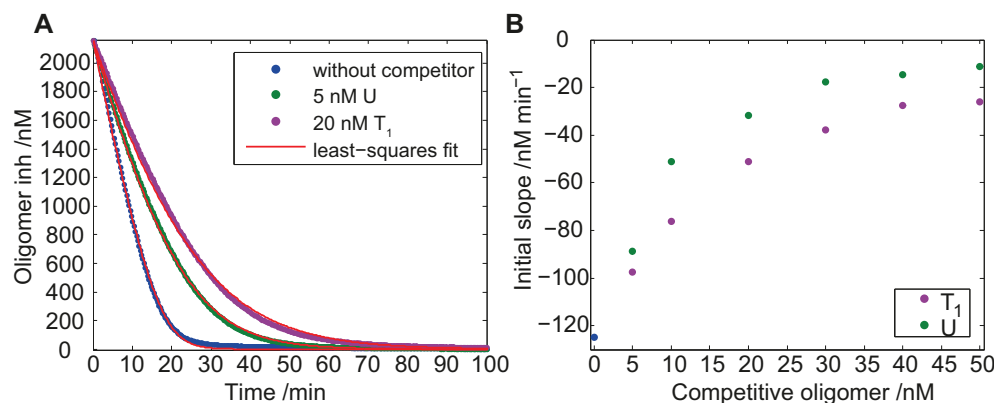


Figure 3.3 Experimental data of degradation of ssDNA by exonuclease was handled and analyzed as described to obtain the Michaelis-Menten parameters. Experiments were performed at 42°C in master mix (2x EvaGreen) with 25 nM exonuclease ttRecJ. RFU was measured every 19 seconds. (A) Degradation of oligomer *inh* in presence of competitors with different length and sequence. The data curves are the mean concentration at each timepoint of triplet of samples as described in the analysis procedure. Degradation of oligomer *inh* in absence of inhibiting template is shown as the blue line. Non-linear least squares optimization of equation (3.11) to the data was performed (red line) to obtain V_{\max} and K_M . Degradation of oligomer *inh* in presence of competitive oligomer **U optimal** (5 nM) and oligomer **T₁ optimal** (20 nM) are shown in green and purple respectively, and the corresponding optimized curves with equation (3.12) are shown in red. The dissociation constants of the competitive inhibitors **T₁ optimal** and **U optimal** with exonuclease, obtained from least-squares fitting, are 5.099 ± 0.021 nM and 2.415 ± 0.004 nM respectively. (B) Plot of initial slope of degradation of oligomer *inh* against the concentration of competitive oligomer **T₁ optimal** (purple) and **U optimal** (green). The blue point is the initial slope of degradation of oligomer *inh* in absence of competitive oligomer.

Kinetics of exonuclease ttRecJ

During experimental characterization ttRecJ exonuclease, we discovered that this enzyme binds strongly to phosphorothioate protected ssDNA strands thereby preventing binding of degradable unprotected ssDNA strands (Figure 3.1B). We believe that this binding is the result of strong metal-ligand interactions between the backbone and Mn^{2+} or Mg^{2+} sites present in the exonuclease binding pocket²². This unproductive sequestration can result in unintentional cross-coupling between network components, and therefore, the dissociation constant characterizing this binding event was quantified.

The fluorescent signal from EvaGreen was normalized using the signal from the reference dye ROX. Then, the RFU at the timepoint at which it has settled into a steady state, corresponding to the timepoint at which all non-protected primer is degraded by exonuclease, is subtracted from the datacurve, so that at $t = t_{\text{end}}$ an RFU of zero is obtained. Next, the RFU is converted to the concentration of ssDNA by dividing the RFU with a factor obtained by dividing the initial RFU by the initial concentration of degradable ssDNA. After preparing the temporal datacurves, the mean concentration at each timepoint of the triplet of samples was taken and used for fitting. This procedure was followed for all datacurves (Figure 3.3A).

Assuming quasi-steady state approximation, the Michaelis-Menten parameters V_{\max} and K_M are determined. The integrated Michaelis-Menten equation (3.11) is fitted to the datacurve obtained from the experiment without competitor by means of non-linear least squares optimization.

$$V_{\max} t = [S]_0 - [S](t) + K_M \ln \left(\frac{[S]_0}{[S](t)} \right) \quad (3.11)$$

Here, $[S]_0$ is the initial concentration of oligonucleotide, $[S](t)$ the concentration of oligonucleotide at a certain timepoint t , V_{\max} the maximum enzymatic rate and K_M the Michaelis-Menten constant. Subsequently, the dissociation constant K_I quantifying the unproductive binding of protected ssDNA templates to the exonuclease is determined by non-linear least squares optimization of the integrated Michaelis-Menten equation incorporating competitive inhibition.

$$V_{\max} t = [S]_0 - [S](t) + K_M \left(1 + \frac{[I]}{K_I} \right) \ln \left(\frac{[S]_0}{[S](t)} \right) \quad (3.12)$$

Here, $[I]$ is the concentration of non-complementary inhibitive oligomer and K_I the dissociation constant of the inhibitor. The Michaelis-Menten parameters V_{\max} and K_M obtained from non-linear least-square optimization of the data with equation (3.11) were used as constant values leaving K_I as the only parameter to be estimated. To determine the parameters V_{\max} , K_M and K_I , optimization was performed using the Matlab routine `lsqnonlin` with the same settings and procedure as described in the previous section.

In Figure 3.3A, the results of degradation of oligomer *inh* in the presence of competitive oligomers are shown, and the corresponding dissociation constants are displayed in Table 3.2. The competitive sequences are variations on the optimal sequences that were determined by the DNA sequence design algorithm. The effect of competitive oligomer on the activity of exonuclease is visualized more clearly in Figure 3.3B. Increasing the concentration of competitive oligomer from 0 nM to 20 nM causes a large decrease in enzymatic activity, while further increasing the concentration of competitive oligomer (from 20 nM to 50 nM) further decreases the enzymatic activity only slightly. This observation is explained by the low dissociation constant of competitive oligomer with exonuclease that causes exonuclease to be saturated with competitive inhibitor at relatively low concentrations of inhibitor. Furthermore, the decrease in initial slope is less for competitive oligomer T_1 compared to competitive oligomer U , which is explained by the lower dissociation constant of U compared to T_1 .

Table 3.2 Dissociation constants quantifying the unproductive binding of protected ssDNA oligomers to exonuclease. The dissociation constant was determined at 42°C in master mix for different designs of oligomer T_1 and U . The Michaelis-Menten parameters of the degradation of oligomer *inh* obtained from fitting the integrated Michaelis-Menten equation without competition are $V_{\max} = 169.97 \pm 1.80 \text{ nM min}^{-1}$ and $K_M = 529.76 \pm 17.43 \text{ nM}$, respectively. Sequences are displayed 5' \rightarrow 3'.

| Name | Sequence | K_I (nM) |
|----------------------|-----------------------------------|-------------------|
| T_1 optimal | A*T*T* AGA CTC ACA AAC GAC TCC TA | 5.099 ± 0.021 |
| T_1 modification 1 | G*T*T* AGA CTC ACA AAC GAC TCC TA | 5.558 ± 0.019 |
| T_1 modification 2 | C*T*T* AGA CTC ACA AAC GAC TCC TA | 3.492 ± 0.009 |
| U optimal | T*A*G* GAG TCG TTT | 2.415 ± 0.004 |
| U modification 1 | A*A*G* GAG TCG TTT | 0.831 ± 0.004 |
| U modification 2 | T*T*G* GAG TCG TTT | 1.699 ± 0.005 |
| U modification 3 | C*A*T* GAG TCA GAT | 0.555 ± 0.002 |

3.3.2 The extended model

Sampling and response comparison

The relevant expressions for the additional parameters in the extended model are depicted in Table 3.1, where the values are obtained from literature^{25,35} and experiments. Data from Fukuyama and co-workers³⁵ and the aforementioned measurements were used to determine the forward reaction rate $k_{f,exoN}$. This parameter, along with the others depicted Table 3.3, are kept at these constant values when assessing the extended model. The reverse rate constants follow from the forward rate constant $k_{f,exoN}$ and the measured dissociation constants (K_I) and Michealis-Menten constants (K_M) (Table 3.1).

Table 3.3 Parameter values that are kept at constant values in the extended model, based on literature and measurements.

| Parameter value |
|---|
| $k_{f,exoN} = 0.025 \text{ nM}^{-1} \text{ min}^{-1}$ |
| $K_{I,U} = 1.38 \text{ nM}$ |
| $K_{I,T1} = K_{I,T2} = K_{I,T3} = 4.72 \text{ nM}$ |
| $K_{M,exoNshort} = 190 \text{ nM}$ |
| $K_{M,exoNlong} = 530 \text{ nM}$ |

Exhaustive sampling is performed on the extended model (equation (A.9) in Appendix section A.2) to screen its ability for adaptive responses. As compared to the parameter screening performed in Chapter 2, here we include an additional parameter, *i.e.* the concentration of exonuclease [*exoN*]. The response heatmap (Figure 3.4A) of the improved model shows that the circuit is capable of adaptive responses characterized by high sensitivity values ($\log_{10}(S) > 0$) while maintaining precision ($\log_{10}(P) > 1$), that is, output signals that return to prestimulus activity show a transient response exhibiting a sensitivity value larger than one. The viable regions of the parameter space after exhaustive sampling are broader compared to the original model (Figure 3.4B depicts the corresponding box plots for adaptive response). However, the portion of the parameter space that

leads to sensitivity values larger than 1 (*i.e.* amplified exact adaptation), is quite narrow (Figure 3.4C).

The effect of the concentration of exonuclease

The parameters of a typical response of the original IFFL circuit that is shown in Figure 2.7C are used to investigate the extended model. Figure 3.5A shows a response of the improved model exhibiting amplified exact adaptation ($\log_{10}(S) > 0$ and $\log_{10}(P) > 1$) alongside the response of the original circuit with the same parameters as in Figure 2.7C. Figure 3.5B displays the effect of the concentration of exonuclease on the quality of the response. Low availability of exonuclease leads to accumulation of ssDNA strands, saturating the templates, thus reducing the relative responses of the system. As the exonuclease concentration increases, the response time and steady state concentration of Y decrease, while the peak value stays approximately constant, thus increasing sensitivity. The concentration of exonuclease reaches an optimal point at which the system displays its optimal response, after which further increase decreases the overall quality of the

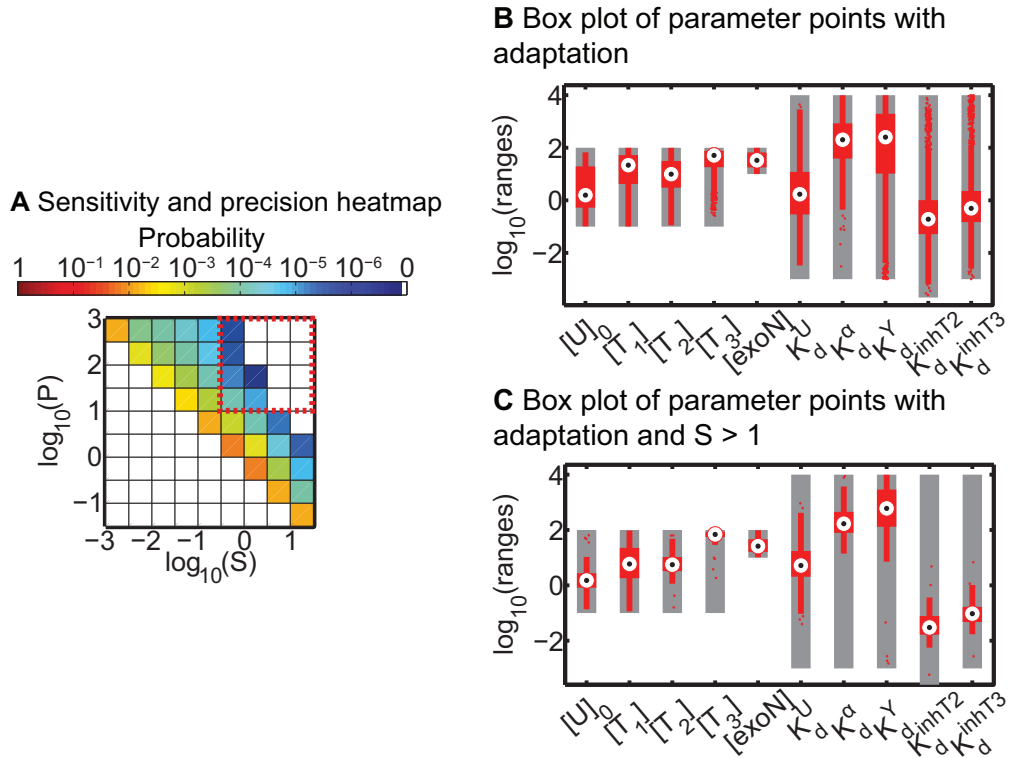


Figure 3.4 Parameter sampling of the extended model. (A) Heatmap of the probability of finding a certain combination of sensitivity and precision values (10^7 samples). The adaptive regime ($\log_{10}(S) > -0.5$ and $\log_{10}(P) > 1$) is highlighted by a red dashed box. Notably, responses that exhibit sensitivity values larger than 1 ($\log_{10}(S) > 0$) while maintaining high precision are found. The hit rate is 0.01%. (A) Box plots representing the parameter points that lead to adaptive responses in the model. (B) Box plots representing the parameter points that lead to adaptive responses that exhibit sensitivity values larger than 1 (amplified exact adaptation).

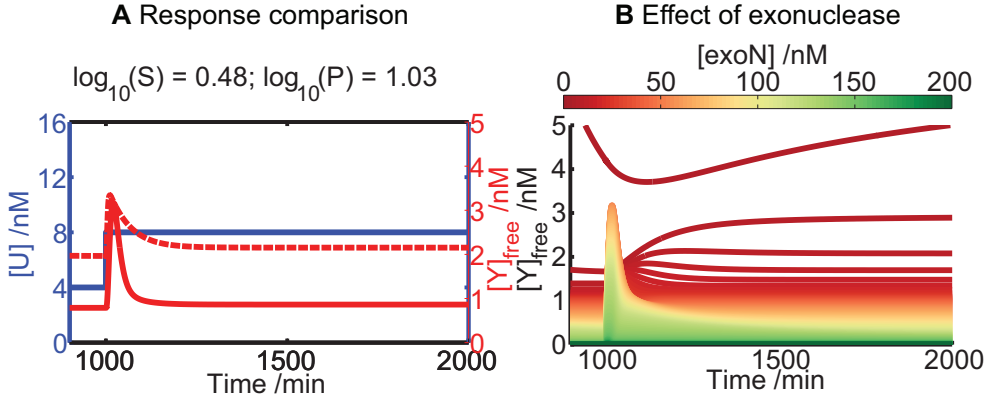


Figure 3.5 Elaboration of a response of the extended model. (A) Comparison of an adaptive response without (dashed lines, Figure 4D) and with (solid lines) competitive inhibition of exonuclease by phosphorothioate protected strands. Competitive inhibition of the degradation machinery results in an amplified adaptive response characterized by sensitivity larger than 1. The parameters corresponding to this response are $K_d^U = 14.6 \times 10^3$ nM, $K_d^\alpha = 1.8 \times 10^3$ nM, $K_d^{inhT2} = 0.037$ nM, $K_d^Y = 5.8 \times 10^3$ nM, $K_d^{inhT3} = 0.042$ nM, $[U]_0 = 4$ nM, $[T_1] = 24$ nM, $[T_2] = 12$ nM, $[T_3] = 95$ nM, and $[exoN] \approx 50$ nM. The lower steady state for the extended model is explained by less availability of DNA templates due to unproductive sequestration. (B) Effect of the exonuclease concentration in the model that incorporates competitive inhibition of exonuclease with the same parameters as in A. A concentration of $[exoN] \approx 50$ nM exhibits the best response. Very high exonuclease concentrations lead to extremely low ssDNA concentrations and no significant response in Y .

response because ssDNA strands are degraded too fast. Near the optimal exonuclease concentration, the nonlinearity introduced by high affinity of the enzyme for protected ssDNA allow these strands to function as a buffer for enzymatic activity, which is necessary for the circuit to exhibit $S > 1$.

Moreover, the parameter regimes in which exact adaptation ($\log_{10}(S) > -0.5$ and $\log_{10}(P) > 1$) occurs are significantly broadened due to unproductive exonuclease sequestration (Figure 3.4B), while the robustness measures of the dissociation constants increase (Figure 3.6A). While this is not reflected by the value of the normalized viable volume in a global sense, it should be noted, that values of initial concentrations do contribute to the global robustness, but not to the local robustness (see Chapter 2). Both the median and maximum relative occurrence of fraction accepted sets (*i.e.* overall local robustness) when perturbing dissociation constants, but not concentrations, are higher for the extended model than for the original model. Figure 3.6B displays the robustness measures of each of the individual dissociation constants, which—except for K_d^U being slightly smaller—are larger than those of the original model (Figure 2.9B). All in all, these observations indicate that incorporation of explicit exonuclease kinetics reduces the seemingly required high discrepancy in dissociation constants for the original IFFL circuit to show adaptive dynamics. Therefore, we can conclude that unproductive sequestration of the degradation machinery of the DNA-based IFFL circuit appears to greatly improve the quality and robustness of the adaptive response.

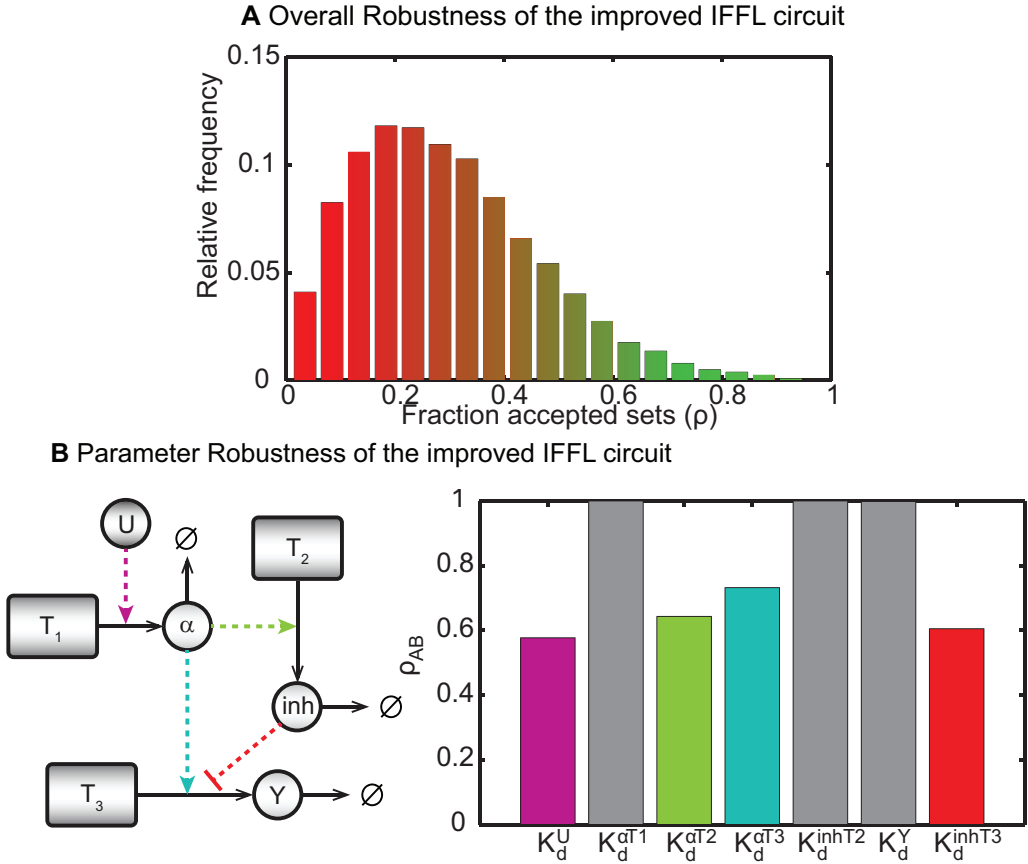


Figure 3.6 Local robustness analysis of the improved model. (A) Overall local robustness *versus* relative frequency for the improved IFFL circuit. Robustness increases as the relative frequency decreases. Color gradient indicates fragility (red) and robustness (green) of parameter sets. The improved IFFL circuit has a higher maximum overall robustness value than the original circuit (Figure 2.8B). (B) Individual parameter robustness of the improved IFFL circuit. Indeed, the robustness of individual parameters is higher in the extended circuit—except for K_d^U (see Figure 2.9B). We believe that the latter can be attributed to the increased number of reactions that input primer U participates in.

3.3.3 Analysis of adaptation dynamics in the toy models

Toy model I

In principle, competition between the network constituents for exonuclease should result in nonlinear degradation terms. However, the first-order approximation of equation (3.2) remains valid, as the individual concentration of each of these species is well below the Michaelis-Menten constant of exonuclease for unprotected strands ($K_{M, \text{exoN}} \sim 190 - 530 \text{ nM}$). It can be readily shown that TMI exhibits high precision but the sensitivity of its adaptive response is below one. The results in this section are obtained through close collaboration with S. Masroor, MSc, Dr. A. Zagaris and Prof. M. A. Peletier. Suppose that the system is in steady state at $t = 0$. As a result of a step increase in $U: U_1 \rightarrow U_2$, $\alpha(t)$ and $inh(t)$ always increase monotonically, *i.e.* $\alpha(t) > \alpha(0)$ and

$inh(t) > inh(0)$. This observation assists the estimation of the maximum value Y_p of Y , which we assume to be attained at some time t_p . Since the time derivative of Y equals 0 both at $t = t_p$ and at $t = 0^-$, we obtain equation (3.13) for Y_p ,

$$Y_p = \frac{k_Y T_3 \alpha(t_p)}{DK_a \left(1 + \frac{\alpha(t_p)}{K_a} + \frac{inh(t_p)}{K_{inh}} \right)} \quad (3.13)$$

and equation (3.14) for $Y(0^-)$

$$Y(0) = \frac{k_Y T_3 \alpha(0)}{DK_a \left(1 + \frac{\alpha(0)}{K_a} + \frac{inh(0)}{K_{inh}} \right)} \quad (3.14)$$

Now, we may set a bound for the value Y_p using monotonicity (equation (3.15)),

$$Y_p \leq \frac{k_Y T_3 \alpha(\infty)}{DK_a \left(1 + \frac{\alpha(0)}{K_a} + \frac{inh(0)}{K_{inh}} \right)} \quad (3.15)$$

hence, we obtain an upper bound for sensitivity (equation (3.16)).

$$\begin{aligned} S &= \left(\frac{Y_p - Y(0)}{Y(0)} \right) / \left(\frac{U_2 - U_1}{U_2} \right) \\ &\leq \left(\frac{\alpha(\infty) - \alpha(0)}{\alpha(0)} \right) / \left(\frac{U_2 - U_1}{U_2} \right) \\ &= \left(\frac{K_U}{(K_U + U_2)} \frac{(U_2 - U_1)}{U_1} \right) / \left(\frac{U_2 - U_1}{U_1} \right) \\ &= \frac{K_U}{(K_U + U_2)} \Rightarrow \\ S &\leq 1 \end{aligned} \quad (3.16)$$

This is in excellent agreement with the earlier observation that the DNA-based enzymatic IFFL circuit fails to amplify transient changes in its input in Chapter 2. Intuitively, the output dynamics of TMI are characterized by the total concentration of input strand. For low total concentrations of input, the kinetics of the enzymes in equation (3.2) operate in their linear regime, whereas high concentrations saturate them. The inhibitory effect of inh on Y is negligible in the former case, resulting in a steady state value \bar{Y} of the output that grows roughly proportionally with U ; that is, the network is imprecise. In the latter case, inh inhibits the output successfully, making \bar{Y} largely insensitive to changes in U , that is, the network exhibits high precision. This becomes obvious from the explicit formula for the steady state (equation (3.17)) which yields the approximations $\bar{Y} \approx \bar{Y}_{\max}$ and $\bar{Y} \approx \bar{Y}_{\min}$ for low and high concentrations U , respectively. The expressions for the constants K ,

K' , and K'' are given in equation (A.14)-(A.18) in Appendix section A.2. Summarizing, while TMI is able to generate adaptive responses with high precision, the sensitivity of the response is low ($S < 1$).

$$\bar{Y} = \bar{Y}_{\max} \frac{U^2 + KU}{U^2 + K'U + K''} \quad (3.17)$$

Toy model II

TMII displays an adaptive response with high precision under the same conditions as TMI. The mechanism enabling adaptation is similar to that in TMI, with the steady state values of the model components behaving differently for input strand concentration in excess of $U > \text{exo}N + T_1$ (Figure 3.7B). In contrast to TMI, however, TMII is capable of generating adaptive responses with high sensitivity for input concentrations around $U \approx \text{exo}N + T_1$ (Figure 3.7A). The mathematical derivations underlying this section are performed by S. Masroor, MSc and Dr. A. Zagaris, and are presented in Appendix A.2. Amplified exact adaptation in TMII is driven by unproductive binding of input strand to exonuclease, as a result of large relative decreases in the degradation rates for $U \approx \text{exo}N + T_1$ (Figure 3.7C, bottom graph). To investigate this mechanism in detail, we first outline two operational modes for the circuit specified by the total concentration U relative to the summed concentration $\text{exo}N + T_1$. First, for concentrations $U < \text{exo}N + T_1$, most of the input is bound to exonuclease or template T_1 (Figure 3.7B), while the remainder drives the production of α , inh , and Y . In this mode, the steady state \bar{Y} increases with the concentration U , similar to the linear regime in TMI. When $U > \text{exo}N + T_1$, exonuclease and template T_1 are nearly saturated and $U_{\text{free}} \approx U - \text{exo}N - T_1$. In this operational mode, the production terms are also saturated, and the inhibitory connection controls \bar{Y} effectively (Appendix A.2). In the regime intermediate to those modes, $U \approx \text{exo}N + T_1$, exonuclease and template T_1 approach saturation allowing $U_{\text{free}} \ll U$ to increase sharply with U (Figure 3.7B, bottom panel). This increase in U_{free} increases production rates and decreases degradation rates by unproductive sequestration of exonuclease, thereby elevating $\bar{\alpha}$, $\bar{\text{inh}}$, and \bar{Y} sharply (Figure 3.7B, top two graphs). Precision increases rapidly in this transitory regime (Figure 3.7C, top panel), as $\bar{\text{inh}}$ arrests further production of \bar{Y} . Concurrently, changes in U affect U_{free} commensurately ($\delta U_{\text{free}} \approx \delta U$), but relative changes $\delta U_{\text{free}}/U_{\text{free}} \approx (U/U_{\text{free}})(\delta U/U)$ are starkly amplified by the factor $(U/U_{\text{free}}) \gg 1$ (Figure 3.7C, bottom graph). Such relative changes in U_{free} instantly affect the degradation rate, while the production machinery of Y responds much slower. The net result is a transient imbalance in the production and degradation of output, which drives a substantial overshoot in the response of the circuit signaling high sensitivity values.

3.4 CONCLUSION

Although expanding the presented experiments into a full experimental implementation of the IFFL circuit may seem straightforward at this point, initial scoping experiments have revealed that the use of fluorescent probes that transiently sequester ssDNA output Y leads to strong retroactivity effects that negatively influence adaptation dynamics making characterization of the circuit dynamics extremely challenging³⁶. Currently, we are exploring strategies to attenuate retroactivity effects in PEN-based circuits. Next to the experimental characterization of the catalytic platform, we have

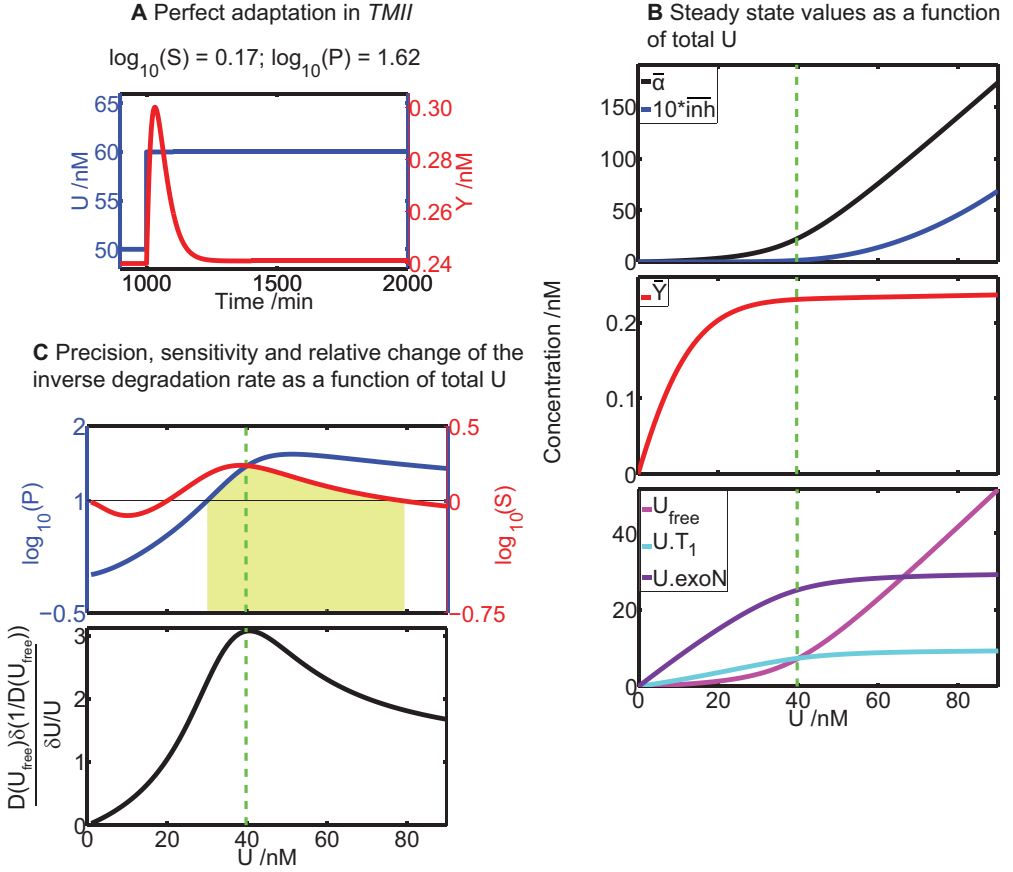


Figure 3.7 Analysis of TMII. (A) Amplified exact adaptation in TMII as a result of increasing the total concentration of input ssDNA by 20%. Parameter values for this simulation are $K_U = 0.2$ nM, $K_\alpha = 1250$ nM, $K_{inh} = 0.001$ nM, $K_{I,U} = 1.4$ nM, $K_{M,exoN} = 440$ nM, $k_\alpha T_{1,tot} = 2.5$ nM min⁻¹, $k_{inh} T_2 = 8$ nM min⁻¹, $k_Y T_3 = 106.5$ nM min⁻¹, and $V_{max,exoN} = 300$ nM min⁻¹. (B) Steady state values of the three states in the model, that is, $\bar{\alpha}$, $\bar{m}h$, and \bar{Y} , as a function of the total concentration of input strand, U (top two panels); the effect on both U_{free} and $U \cdot T_1$ of unproductive binding of input strand U to template T_1 and exonuclease (bottom panel). The dashed green lines represent the summed concentration $exoN + T_1$. (C) The top graph shows the logarithm of precision (blue) and sensitivity (red) as a function of the total concentration of input strand, U , with the amplified exact adaptation regime highlighted in yellow ($\log_{10}(S) > 0$ and $\log_{10}(P) > 1$). The bottom graph displays the relative change in inverse degradation rate (Appendix A.2); degradation slows down when this measure becomes large, therefore correlating very well with sensitivity.

characterized the unproductive sequestration of the degradation machinery of the PEN toolbox. The noncanonical crosscoupling between network components results in strong nonlinearities in their degradation, which can either suppress or promote target dynamic behavior³¹. For the case of the IFFL circuit, we have shown how noncanonical cross-coupling by enzyme-limited competition provides a novel amplification mechanism and broadens the parameter range in which the network displays adaptation. In accordance with previous work, complex behavior is strongly influenced by network topology rather than by cross-coupling. Therefore, the results in Chapters 2 and 3 will aid

in the development of integrated biochemical circuits that can probe their chemical environment, process chemical signals, make decisions, and take action at the molecular level.

3.5 REFERENCES

1. B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts and P. Walter (2014), *Molecular Biology of the Cell* (6th ed.). Garland Science
2. E. H. Davidson (2010), Emerging properties of animal gene regulatory networks. *Nature* **468**, 911–920.
3. O. Atay and J. M. Skotheim (2014), Modularity and predictability in cell signaling and decision making. *Mol. Biol. Cell* **25**, 3445–3450.
4. A. Doncic, O. Atay, E. Valk, A. Grande, A. Bush, G. Vasen, A. Colman-Lerner, M. Loog and J. M. Skotheim (2015), Compartmentalization of a bistable switch enables memory to cross a feedback-driven transition. *Cell* **160**, 1182–1195.
5. M. C. Thomas and C.-M. Chiang (2006), The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–178.
6. A. Smallwood and B. Ren (2013), Genome organization and long-range regulation of gene expression by enhancers. *Curr. Opin. Cell Biol.* **25**, 387–394.
7. M. Kozak (1999), Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208.
8. N. Malys and J. E. G. McCarthy (2010), Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* **68**, 991–1003.
9. D. Vucic, V. M. Dixit and I. E. Wertz (2011), Ubiquitylation in apoptosis: a post-translational modification at the edge of life and death. *Nat. Rev. Mol. Cell Biol.* **12**, 439–452.
10. K. K. Biggar and S. S.-C. Li (2015), Non-histone protein methylation as a regulator of cellular signalling and function. *Nat. Rev. Mol. Cell Biol.* **16**, 5–17.
11. E. Verdin and M. Ott (2015), 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat. Rev. Mol. Cell Biol.* **16**, 258–264.
12. D. Bray (1995), Protein molecules as computational elements in living cells. *Nature* **376**, 307–312.
13. K.-E. Jaeger and T. Eggert (2004), Enantioselective biocatalysis optimized by directed evolution. *Curr. Opin. Biotechnol.* **15**, 305–313.
14. L. Hedstrom (2010), Enzyme Specificity and Selectivity. In *eLS*. John Wiley & Sons, Ltd.
15. C. I. Butré, S. Sforza, H. Gruppen and P. A. Wierenga (2014), Introducing enzyme selectivity: a quantitative parameter to describe enzymatic protein hydrolysis. *Anal. Bioanal. Chem.* **406**, 5827–5841.
16. A. Cornish-Bowden (2013), The origins of enzyme kinetics. *FEBS Lett.* **587**, 2725–2730.
17. J. J. Tyson, K. C. Chen and B. Novak (2003), Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231.
18. W. Ma, A. Trusina, H. El-Samad, W. A. Lim and C. Tang (2009), Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760–773.
19. P. M. Kim and B. Tidor (2003), Limitations of quantitative gene regulation models: a case study. *Genome Res.* **13**, 2391–2395.
20. W. H. Mather, N. A. Cookson, J. Hasty, L. S. Tsimring and R. J. Williams (2010), Correlation resonance generated by coupled enzymatic processing. *Biophys. J.* **99**, 3172–3181.
21. X. Chen, N. Briggs, J. R. McLain and A. D. Ellington (2013), Stacking nonenzymatic circuits for high signal gain. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5386–5391.
22. T. Wakamatsu, Y. Kitamura, Y. Kotera, N. Nakagawa, S. Kuramitsu and R. Masui (2010), Structure of RecJ exonuclease defines its specificity for single-stranded DNA. *J. Biol. Chem.* **285**, 9762–9769.
23. T. Wakamatsu, K. Kim, Y. Uemura, N. Nakagawa, S. Kuramitsu and R. Masui (2011), Role of RecJ-like protein with 5′-3′ exonuclease activity in oligo(deoxy)nucleotide degradation. *J. Biol. Chem.* **286**, 2807–2816.
24. K. Montagne, R. Plasson, Y. Sakai, T. Fujii and Y. Rondelez (2011), Programming an *in vitro* DNA oscillator using a molecular networking strategy. *Mol. Syst. Biol.* **7**, 466.
25. A. Padiac, T. Fujii and Y. Rondelez (2012), Bottom-up construction of *in vitro* switchable memories. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3212–E3220.
26. T. Fujii and Y. Rondelez (2013), Predator–prey molecular ecosystems. *ACS Nano* **7**, 27–34.

27. N. E. Buchler and F. R. Cross (2009), Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* **5**, 272.
28. D. Chen and A. P. Arkin (2012), Sequestration-based bistability enables tuning of the switching boundaries and design of a latch. *Mol. Syst. Biol.* **8**, 620.
29. O. S. Venturelli, H. El-Samad and R. M. Murray (2012), Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3324–E3333.
30. M. J. Rust, J. S. Markson, W. S. Lane, D. S. Fisher and E. K. O’Shea (2007), Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science* **318**, 809–812.
31. Y. Rondelez (2012), Competition for catalytic resources alters biological network dynamics. *Phys. Rev. Lett.* **108**, 018102.
32. T. S. Hatakeyama and K. Kaneko (2012), Generic temperature compensation of biological clocks by autonomous regulation of catalyst concentration. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8109–8114.
33. A. Prindle, J. Selimkhanov, H. Li, I. Razinkov, L. S. Tsimring and J. Hasty (2014), Rapid and tunable post-translational coupling of genetic circuits. *Nature* **508**, 387–391.
34. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (2007), *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press.
35. A. Yamagata, R. Masui, Y. Kakuta, S. Kuramitsu and K. Fukuyama (2001), Overexpression, purification and characterization of RecJ protein from *Thermus thermophilus* HB8 and its core domain. *Nucleic Acids Res.* **29**, 4617–4624.
36. D. Del Vecchio, A. J. Ninfa and E. D. Sontag (2008), Modular cell biology: retroactivity and insulation. *Mol. Syst. Biol.* **4**, 161.

Chapter 4

Design and Analysis of a Tunable Oscillating Enzymatic Network

Parts of this Chapter are described in:

Sergey N. Semenov, Albert S. Y. Wong, R. Martijn van der Made, Sjoerd G. J. Postma, Joost Groen, Hendrik W. H. van Roekel, Tom F. A. de Greef and Wilhelm T. S. Huck (2015), Rational design of functional and tunable oscillating enzymatic networks. *Nature Chemistry* **7**, 160–165.

ABSTRACT

Life is sustained by complex systems operating far from equilibrium and consisting of a multitude of enzymatic reaction networks. The operating principles of biology's regulatory networks are known, but the *in vitro* assembly of out-of-equilibrium enzymatic reaction networks has proved challenging, limiting the development of synthetic systems showing autonomous behavior. Recently, a strategy was presented that enables rational design of programmable functional reaction networks that exhibit dynamic behavior. In collaboration with our experimental partners, we have engineered an enzymatic oscillator based autoactivation and delayed negative feedback of the enzyme trypsin. The system is capable of sustained oscillations for over 65 hours. Here, we analyze the computational model that was used to guide the design of the network. We first show how parameter sampling can provide bounds on the functional ranges of the reaction rates. Optimization of the computational model is then performed to match the experimental data and constraints. Finally, we perform a profile likelihood analysis on the parameters in the model, providing insights into the necessary tradeoffs between chemical observations, parameter identifiability, computational accuracy and network tunability.

4.1 INTRODUCTION

The response of living cells to physicochemical changes in their environment is regulated by complex integrated networks of reactions that control the functioning of biomolecules in space and time. Time-keeping, chemical amplification and signal modulation are examples of essential functions in signal transduction, vision, metabolic regulation and cell division in biological systems^{1,2}. These functions are a property of complex networks of biochemical reactions that operate far from equilibrium^{3,4}. Moreover, cells prevent unwanted crosstalk in their metabolic and biochemical networks by separating them in space and time. Spatial separation is accomplished *e.g.* by compartmentalization of the biochemical networks in organelles, while temporal separation is achieved by timely biosynthesis of proteins and enzymes.

Chemical reaction networks (CRNs) and biochemical reaction networks (BRNs) arise out of the myriad interactions between the components of the cell, and their characteristics transcend the properties of individual molecules and reactions. In recent years, the possibility has arisen to design synthetic gene networks that control some of the fundamental properties of living systems^{5–7}, but it remains a major challenge to create synthetic dynamic molecular systems that capture the extraordinary richness in behavior displayed by living cells. Significant efforts have been made in applying the principal regulatory motifs of biochemical reaction networks to dissipative systems based on DNA replication, transcription and translation^{8–15}. Complex spatiotemporal pattern formation has been observed in the classic example of Belousov-Zhabotinsky (BZ) oscillations¹⁶ and together with related CRNs, these have been harnessed into a rich variety of self-organizing systems^{17–20}.

Oscillations are a well-established hallmark of out-of-equilibrium CRNs^{21,22}. A methodology for constructing tunable, oscillating reaction networks based on enzymatic conversions of small molecules was proposed by our experimental collaborators Dr. S. N. Semenov and A. S. Y. Wong, MSc in the group of Prof. W. T. S. Huck. An oscillating BRN was engineered based on autoactivation and delayed negative feedback of the enzyme trypsin. Employing small molecules as substrates allowed accurate fine-tuning of the reaction rates for each step involved in the network. The BRN showed sustained oscillations in a microfluidic reactor and—combined with the modular nature of the network components—this setup allowed for functional linkage of the core oscillator to downstream processes in separate reactors.

In this Chapter we focus on the computational description of the core oscillator. We first show how parameter sampling can provide bounds on the functional ranges of the reaction rates for the network to show sustained oscillations. We perform optimizations of the computational model with different sets of experimental data and constraints that lead to slightly diverging parameter values and hence, to slightly different responses of the model. Nevertheless, these responses are all qualitatively similar and equally plausible. Therefore, we use a parameter identifiability analysis to assess the similarities and the differences between the optimal sets of parameters. Finally, we analyze the influence of network components on the period and amplitude of the oscillations, providing insight into the tunability of the oscillator.

4.2 MATERIALS AND METHODS

4.2.1 The synthetic network, derivation of the model and parameter sampling

Biochemical networks are able to generate oscillations using several minimal network topologies^{22–25}. The BRN that is studied in this Chapter was designed by Dr. S. N. Semenov and A. S. Y. Wong, MSc in the group of Prof. W. T. S. Huck. The design of the BRN is based on delayed inhibition of an autocatalytic enzymatic process (Figure 4.1A). The autocatalytic species is trypsin (*Tr*), which catalyzes its own production from trypsinogen (*Tg*). The enzymatic activity of *Tr* should result in the ultimate production of an inhibitor that suppresses its enzymatic activity. Delayed inhibition of trypsin is accomplished by two enzymatic deprotection reactions of a masked inhibitor, which is based on the strong and irreversible inhibitor 4-[2-aminoethyl]benzenesulfonyl fluoride (Figure 4.1B). The molecule is protected by two amino acid residues, and is called the proinhibitor (*proInh*) in this protected form. Acetylation of *proInh* at the N-terminus enhances the solubility, and protection prevents inhibition of trypsin as demonstrated experimentally²⁶. The first deprotection is cleavage of the lysine (Lys) residue by *Tr* and results in an intermediate inhibitor (*intInh*) that is also unable to react with trypsin²⁶. The second deprotection is catalyzed by aminopeptidase M (*Ap*), which cleaves the second amino acid residue, releasing the strong inhibitor (*Inh*). The reactions are performed in a continuously-stirred tank reactor (CSTR, Figure 4.1C), and is fed by four separate inlets to supply for *Tg*, *Tr*, *Ap*, and *proInh*. The deprotection steps of the masked inhibitor are amenable to rate tuning, which we will analyze using parameter sampling of the model (*vide infra*). The first deprotection step is examined numerically in terms of the enzymatic efficiency of *Tr* acting on *proInh*. This enzymatic efficiency can be modulated by methylation of the Lys residue, hence the numerical results are compared to experimentally measured cleavage rates of *proInh* by *Tr* with an unmodified Lys residue (proinhibitor **1**) and a methylated ϵ -NH₂ of the Lys residue (proinhibitor **2**). The second deprotection step is investigated by varying the inflow concentration of *Ap*.

In close collaboration with Dr. S. N. Semenov, A. S. Y. Wong, MSc and R. M. van der Made, BSc, we derived an ordinary differential equation (ODE) model of the synthetic chemical network in order to quantitatively analyze the oscillatory behavior. The model takes into account all species that are fed into the CSTR *via* the three inlets (*i.e.* *Tg*, *Tr* and proinhibitor (*proInh*), Figure 4.1C), the products that are formed by molecular conversions (intermediate inhibitor (*intInh*), active inhibitor (*Inh*)), and the reactions that are associated with these species. Figure 4.2 displays the reaction equations associated with the network. The ODE model is based on the following assumptions:

- 1) The enzymatic reactions carried out by trypsin (autocatalysis of *Tg* to *Tr* and conversion of *proInh* to *intInh*) are modeled using mass-action kinetics. Reverse reactions of these conversions are considered negligible.
- 2) In accordance with literature, conversion of trypsinogen to trypsin occurs *via* enzymatic conversion *and* bimolecular autoactivation^{27,28}.

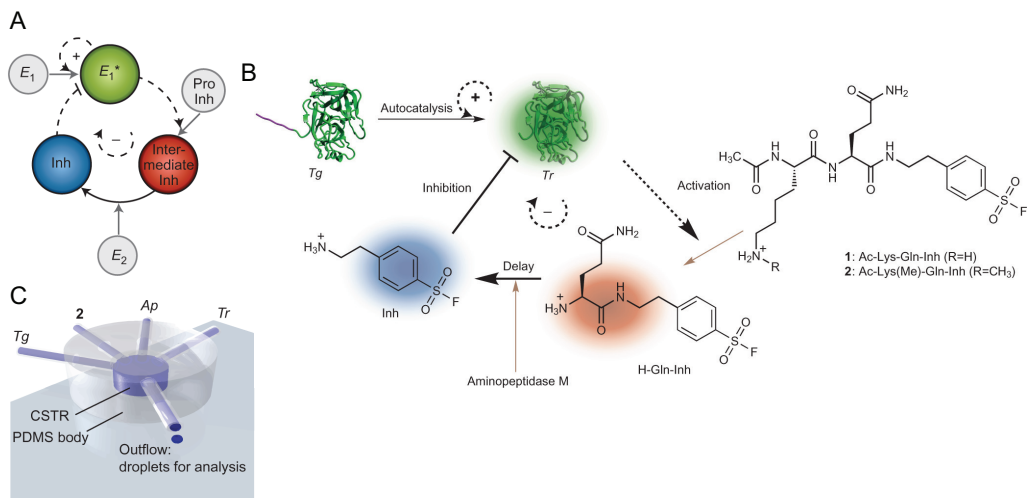


Figure 4.1 Network design and experimental assembly of a flow-based enzymatic oscillator. (A) Schematic network layout of the enzymatic oscillator based on autocatalytic production and delayed inhibition of an enzyme. Enzymatically inactive E_1 is converted to active E_1^* . In a positive feedback loop, E_1^* catalyzes its own formation. In addition, E_1^* catalyzes the first of a two-step sequence that unmasks an inhibitor of itself, with the second step being catalysed by a second enzyme E_2 . This two-step process constitutes a negative feedback loop. The combination of positive and negative feedbacks results in an oscillating system. (B) Detailed reaction diagram of the BRN. Proinhibitor 1 has a regular lysine residue, while proinhibitor 2 contains a methylated ϵ -NH₂ in its lysine residue (see section 3.3). (C) Schematic representation of the flow reactor. The concentration of Tr in the outflow is measured using a standard Tr activity assay³⁰.

3) The enzymatic conversion of $intInh$ to Inh by Ap is assumed to follow Michaelis-Menten kinetics. This assumption is supported both by a relatively high Michaelis-Menten constant ($K_M \approx 3$ mM, see section 4.3)²⁹ and a constant inflow of Ap , rendering the total enzyme concentration for this enzymatic conversion approximately constant.

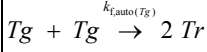
4) Inhibition of trypsin by active inhibitor (Inh) is considered irreversible.

5) Both $proInh$ and $intInh$, could in theory also irreversibly inhibit trypsin.

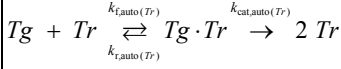
6) Aqueous hydrolysis of $proInh$, $intInh$ and Inh is assumed to follow first-order kinetics, and its reverse reaction is negligible.

We are now in a position to derive the system of ODEs that describes the dynamics of the network. Appendix section A.3 displays the full derivation that ultimately leads to the full model in equation (4.1), guided by Figure 4.2. Because the reactions are carried out in a flow reactor, the inflow of new material should also be incorporated. The inflow contains concentrations of trypsinogen $[Tg]_0$, trypsin $[Tg]_0$ and proinhibitor $[PFL]_0$ (reaction 7, Figure 4.2). The concentration of aminopeptidase is implicitly incorporated in the parameter $V_{max, delay}$. The reaction volume is $V = 250$ μ L, the flow rate is k_{flow} and the total system of differential equations for the synthetic chemical network is given in equation (4.1). A description of all the variables and parameters in this ODE model is given in Table 4.1.

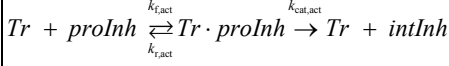
1) Bimolecular autoactivation of trypsinogen:



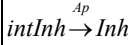
2) Autocatalytic activation of trypsinogen:



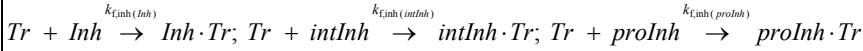
3) Enzymatic conversion of pro-inhibitor to intermediate inhibitor:



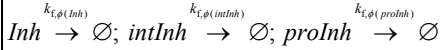
4) Enzymatic conversion of intermediate inhibitor to active inhibitor:



5) Inhibition of trypsin by active inhibitor and background activity of pro- and intermediate inhibitors:



6) Aqueous hydrolysis of inhibitor species:



7) The inflow contains the following species:

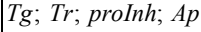


Figure 4.2 Reactions categorized to their effect, corresponding to the chemical network shown in Figure 4.1.

$$\begin{aligned}
 \frac{d[Tg]}{dt} &= -k_{f,auto}(Tg)[Tg][Tg] - k_{f,auto}(Tr)[Tg][Tr] + k_{r,auto}(Tg)[Tg \cdot Tr] + k_{flow}([Tg]_0 - [Tg]) \\
 \frac{d[Tr]}{dt} &= k_{f,auto}(Tg)[Tg][Tg] \\
 &\quad - k_{f,auto}(Tr)[Tg][Tr] + k_{r,auto}(Tr)[Tg \cdot Tr] + 2k_{cat,auto}(Tr)[Tg \cdot Tr] \\
 &\quad - k_{f,act}[Tr][proInh] + k_{r,act}[Tr \cdot proInh] + k_{cat,act}[Tr \cdot proInh] \\
 &\quad - k_{f,inh}(Inh)[Tr][Inh] - k_{f,inh}(intInh)[Tr][intInh] - k_{f,inh}(proInh)[Tr][proInh] + k_{flow}([Tr]_0 - [Tr]) \\
 \frac{d[Tg \cdot Tr]}{dt} &= k_{f,auto}(Tr)[Tg][Tr] - k_{r,auto}(Tr)[Tg \cdot Tr] - k_{cat,auto}(Tr)[Tg \cdot Tr] - k_{flow}[Tg \cdot Tr] \\
 \frac{d[proInh]}{dt} &= -k_{f,act}[Tr][proInh] + k_{r,act}[Tr \cdot proInh] \\
 &\quad - k_{f,inh}(proInh)[Tr][proInh] - k_{f,\phi}(proInh)[proInh] + k_{flow}([proInh]_0 - [proInh]) \\
 \frac{d[Tr \cdot proInh]}{dt} &= k_{f,act}[Tr][proInh] - k_{r,act}[Tr \cdot proInh] \\
 &\quad - k_{cat,act}[Tr \cdot proInh] - k_{flow}[Tr \cdot proInh] \\
 \frac{d[intInh]}{dt} &= k_{cat,act}[Tr \cdot proInh] \\
 &\quad - k_{f,inh}(intInh)[Tr][intInh] - k_{f,\phi}(intInh)[intInh] - k_{flow}[intInh] \\
 \frac{d[Inh]}{dt} &= \frac{k_{cat,delay}[Ap]_0[intInh]}{K_{M,delay} + [intInh]} - k_{f,inh}(Inh)[Tr][Inh] - k_{f,\phi}(Inh)[Inh] - k_{flow}[Inh]
 \end{aligned} \tag{4.1}$$

Table 4.1 Description of the variables and parameters in the ODE model in equation (4.1).

| Variable or parameter | Description |
|--------------------------------------|---|
| $[Tg]_0$ | Inflow concentration of trypsinogen |
| $[Tg]$ | Concentration of trypsinogen |
| $[Tr]_0$ | Inflow concentration of trypsin |
| $[Tr]$ | Concentration of trypsin |
| $[Tg \cdot Tr]$ | Concentration of the trypsinogen-trypsin complex |
| $[proInh]_0$ | Inflow concentration of proinhibitor |
| $[proInh]$ | Concentration of proinhibitor |
| $[intInh]$ | Concentration of intermediate inhibitor |
| $[Inh]$ | Concentration of active inhibitor |
| $[Tr \cdot proInh]$ | Concentration of trypsin-proinhibitor complex |
| $k_{f,auto(Tg)} [mM^{-1} h^{-1}]$ | Rate constant of autoactivation of trypsinogen to trypsin |
| $k_{f,auto(Tr)} [mM^{-1} h^{-1}]$ | Forward rate constant of association between trypsinogen and trypsin |
| $k_{r,auto(Tr)} [h^{-1}]$ | Reverse rate constant of association between trypsinogen and trypsin |
| $k_{cat,auto(Tr)} [h^{-1}]$ | Turnover number of trypsin acting on trypsinogen |
| $k_{f,act} [mM^{-1} h^{-1}]$ | Forward rate constant of association between trypsin and proinhibitor |
| $k_{r,act} [h^{-1}]$ | Reverse rate constant of association between trypsin and proinhibitor |
| $k_{cat,act} [h^{-1}]$ | Turnover number of trypsin acting on proinhibitor |
| $[Ap]_0$ | Inflow concentration of aminopeptidase M |
| $K_{M,delay} [mM]$ | Michaelis-Menten constant of conversion of intermediate inhibitor to active inhibitor catalyzed by aminopeptidase M |
| $k_{f,inh(Inh)} [mM^{-1} h^{-1}]$ | Rate constant of inhibition of trypsin by active inhibitor |
| $k_{f,inh(proInh)} [mM^{-1} h^{-1}]$ | Rate constant of inhibition of trypsin by proinhibitor |
| $k_{f,inh(intInh)} [mM^{-1} h^{-1}]$ | Rate constant of inhibition of trypsin by intermediate inhibitor |
| $k_{i,\phi(Inh)} [h^{-1}]$ | Rate constant of aqueous hydrolysis of active inhibitor |
| $k_{i,\phi(proInh)} [h^{-1}]$ | Rate constant of aqueous hydrolysis of proinhibitor |
| $k_{i,\phi(intInh)} [h^{-1}]$ | Rate constant of aqueous hydrolysis of intermediate inhibitor |
| $k_{flow} [h^{-1}]$ | Rate of the inflow per unit volume PDMS |

Parameter sampling of the model

Based on the rate constants listed in Table 4.1, the ODE model (4.1) is used to determine the range in kinetic parameter values of the proinhibitor activation step and the range in aminopeptidase concentration under both batch ($k_{flow} = 0$) and flow conditions ($k_{flow} > 0$). Under batch conditions, the network should show a single peak and under flow conditions it should show sustained oscillations. During these two sampling procedures, the other parameters were kept constant at their experimentally determined values (see section 3.3). Sampling of the model under batch conditions allows finding of the appropriate rate of the trypsin-catalyzed cleavage of proinhibitor (Figure 4.1B, 1 and 2), while sampling of the model under flow conditions finds appropriate conditions for the CSTR experiments. The simulated temporal $[Tr]$ -profile under both conditions is classified according to specific criteria which are explained in detail (*vide infra*). Sampling was performed in

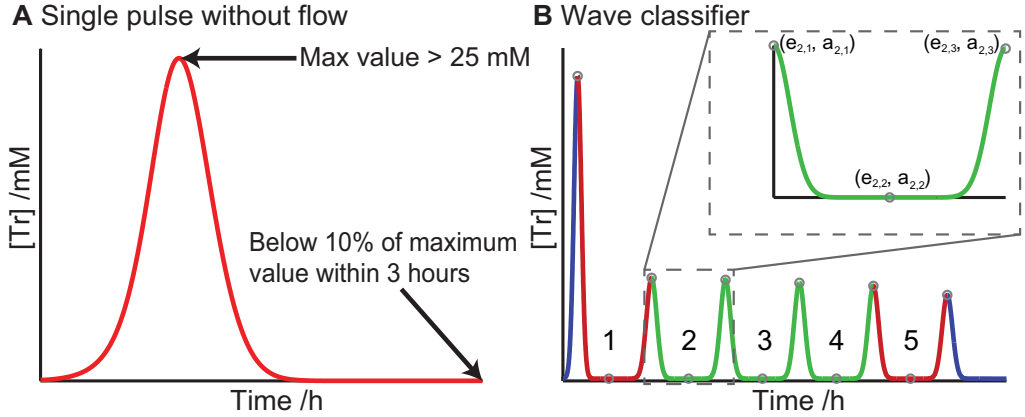


Figure 4.3 Classification criteria for parameter sampling under batch and flow conditions. (A) Classification of the trypsin response of the ODE model reflecting batch conditions, *i.e.* $k_{\text{flow}} = 0$. Simulations of 3 hours are performed and the trypsin concentration should exceed a threshold concentration of 25 μM , and return to below 10% of the maximum value in order for the response to be classified as a pulse. (B) A typical trypsin response under flow conditions. Its max-min-max patterns are numbered. The first and fifth max-min-max patterns are not accepted since $a_{i,1} - a_{i,3}$ is larger than the tolerance. The second, third and fourth max-min-max patterns are accepted since $a_{i,1} - a_{i,3}$ is smaller than the tolerance in all these cases. This particular trypsin response is considered a sustained oscillation overall as there are at least three consecutive accepted patterns required. The output of the model under batch conditions is classified as a single peak of $[Tr]$ if it exceeds a certain threshold concentration (25 μM) and returns within 10% of its maximum value within three hours; otherwise the model is termed unresponsive (see Figure 4.3A). Scoping experiments indicated that these thresholds are suitable for classifying a response as potentially oscillatory when the same experimental conditions are used when $k_{\text{flow}} > 0$.

logarithmic space using Latin Hypercube Sampling ensuring an optimal coverage of the parameter space. The aminopeptidase concentration was sampled between 10^{-3} and 1 U mL^{-1} . The enzymatic efficiency of trypsin in the proinhibitor activation step $\frac{k_{\text{cat,act}}}{K_{\text{M,act}}} = \frac{k_{\text{f,act}}k_{\text{cat,act}}}{k_{\text{r,act}} + k_{\text{cat,act}}}$ was sampled by varying the forward rate constant between 10^{-2} and $10^4 \text{ mM}^{-1} \text{ h}^{-1}$. The reverse rate $k_{\text{r,act}}$ and the turnover number $k_{\text{cat,act}}$ were kept constant at 10^4 h^{-1} and 10^3 h^{-1} , respectively, at their expected order of magnitude (see section 3.3). The sampling is performed in this manner as it ensures a log-uniform coverage of the enzymatic efficiency. The model is integrated in MATLAB® R2010a with compiled MEX files^{31,32} using numerical integrators from the SUNDIALS CCode package³³ for each sampled parameter set.

In order to discriminate between sustained and non-sustained responses we search the responses for so-called 'max-min-max' patterns. The i -th max-min-max pattern is defined as a sequence of a primary local maximum ($t = e_{i,1}$, $[Tr] = a_{i,1}$), a local minimum ($t = e_{i,2}$, $[Tr] = a_{i,2}$) and secondary local maximum ($t = e_{i,3}$, $[Tr] = a_{i,3}$). If, however, the difference between the primary local maximum and the secondary local maximum ($a_{i,1} - a_{i,3}$) is larger than a certain tolerance, then the max-min-max pattern is not accepted. If a trypsin response contains a sufficient number of consecutive accepted patterns it is considered a sustained oscillation, in which case its amplitude and period are defined. A typical trypsin response and its max-min-max patterns are shown in Figure 4.3B. Let k denote the number of identified max-min-max patterns. The amplitude (A), and the period (p_t) of a sustained trypsin response are defined in equations (4.2) and (4.3), respectively.

The tolerance is a fraction of the third peak in the max-min-max pattern ($f_{\text{tol}} \cdot a_{i,3}$). We have written a MATLAB® routine that classifies a response as sustained or non-sustained. The routine reports the amplitude and period of the trypsin response if it is sustained. Throughout this work, we set the tolerance as $f_{\text{tol}} = 0.035$ (*i.e.* a deviation of less than 96.5% is allowed, as this proved to be in best accordance with our judgment of sustained oscillations) and we require that a sustained response shows at least three consecutive peaks. This procedure is also employed to construct the phase planes.

$$A = \frac{1}{k} \sum_{i=1}^k \frac{(a_{i,1} - a_{i,2}) - (a_{i,3} - a_{i,2})}{2} \quad (4.2)$$

$$p_t = \frac{1}{k} \sum_{i=1}^k (e_{i,3} - e_{i,1}) \quad (4.3)$$

4.2.2 Experimental data and model optimization

We here briefly describe the experiments performed in the CSTR and the trypsin activity assay. These experiments were performed by Dr. S. N. Semenov and A. S. Y. Wong, MSc, and for full details of syntheses and characterizations we refer to the original paper²⁶.

CSTR experiments

Four glass syringes were loaded with trypsinogen (8 mg mL⁻¹, 338 μM in 4 mM HCl, 36 mM CaCl₂), trypsin (27 μg mL⁻¹, 1.16 μM in 500 mM Tris-HCl, 20.5 mM CaCl₂, pH 7.7), proinhibitor **2** (4.72 mg mL⁻¹, 7.68 mM in 2 mM HCl) and aminopeptidase (ten times the desired final concentration in the CSTR, which varies, in 10 mM Tris-HCl, 10 mM MgCl₂, pH 7.7) and connected with tubing to the four inlets of a 250 μL PDMS reactor. Fractions of the total flow rate were 0.5 for trypsinogen (that is, 27.5 μL h⁻¹ at a total flow of 55 μL h⁻¹), 0.2 for both trypsin and functional proinhibitor **2** and 0.1 for aminopeptidase. Aliquots of the reaction mixture were diluted with 150 μL of an aqueous 0.1 M KHSO₄ solution to quench all reactions.

Trypsin activity assay

Trypsin activity was measured by mixing 100 μL of the quenched reaction mixture with 3 mL of 5 μg mL⁻¹ bis-(Cbz-L-Arg)-rhodamine fluorogenic substrate in 50 mM Tris-HCl, pH 7.7. The increase in fluorescence intensity ($\lambda_{\text{ex}} = 450$ nm, $\lambda_{\text{em}} = 520$ nm) was monitored for 40 s and the initial, linear slope was compared to a calibration curve to find the concentration of active trypsin.

Model optimization and profile likelihood analysis

Several experiments have been conducted to study the behavior of trypsin in the oscillatory BRN. Based on these experiments we compose a parameter set that, in consolidation with the model, optimally resembles the experimental data. A good choice in optimization would be the maximization of the likelihood^{34,35}, which makes use of the residual sum of squares (RSS). However, we are particularly interested in the oscillatory behavior of trypsin. Therefore, we extend the RSS cost function to account for deviations in the peak height H , the frequency F and the

damping coefficient D . A similar empirical cost function was proposed by Kim *et al.*⁹ and by Simmel *et al.*³⁶ Consider a trypsin response with n peaks (*i.e.* the total number of minima and maxima) and let t_i^{peak} denote the time at which the i -th peak occurs and t_i^{valley} the time at which the subsequent valley occurs. The damping coefficient D is estimated by fitting the absolute consecutive peak differences $|y(t_i^{\text{peak}}) - y(t_{i-1}^{\text{peak}})|$ over time as an exponential decay process ($B = B_0 e^{-\lambda t}$). Specifically, we compute the linear fit of $\ln|y(t_i^{\text{peak}}) - y(t_{i-1}^{\text{peak}})|$ and extract its slope $-\lambda$. H , F and D are denoted by equations (4.4), (4.5) and (4.6), respectively.

$$H = \max_i |y(t_i^{\text{peak}}) - y(t_i^{\text{valley}})| \quad (4.4)$$

$$F = \frac{n-1}{2(t_n^{\text{peak}} - t_1^{\text{peak}})} \quad (4.5)$$

$$D = -\lambda \quad (4.6)$$

The final component of the cost function is related to deviation from experimental measurements of the Michealis-Menten constant $K_{M,\text{auto}}(Tr)$ and the enzymatic efficiency $\frac{k_{\text{cat,act}}}{K_{M,\text{act}}}$. The intervals have been determined as $0.2 < \frac{k_{r,\text{auto}}(Tr) + k_{\text{cat,auto}}(Tr)}{k_{f,\text{auto}}(Tr)} < 0.6$ and $30 < \frac{k_{\text{cat,act}} k_{f,\text{act}}}{k_{r,\text{act}} + k_{\text{cat,act}}} < 90$. We define quadratic errors C_{auto} and C_{act} , respectively, when parameter sets deviate from these two intervals, as in equations (4.7), (4.8). The sum of these C_{exp} in equation (4.9) is incorporated into the cost function.

$$C_{\text{auto}} = \begin{cases} \left(\frac{k_{r,\text{auto}}(Tr) + k_{\text{cat,auto}}(Tr)}{k_{f,\text{auto}}(Tr)} - 0.2 \right)^2 & \text{if } \frac{k_{r,\text{auto}}(Tr) + k_{\text{cat,auto}}(Tr)}{k_{f,\text{auto}}(Tr)} \leq 0.2 \\ 0 & \text{if } 0.2 < \frac{k_{r,\text{auto}}(Tr) + k_{\text{cat,auto}}(Tr)}{k_{f,\text{auto}}(Tr)} < 0.6 \\ \left(\frac{k_{r,\text{auto}}(Tr) + k_{\text{cat,auto}}(Tr)}{k_{f,\text{auto}}(Tr)} - 0.6 \right)^2 & \text{if } \frac{k_{r,\text{auto}}(Tr) + k_{\text{cat,auto}}(Tr)}{k_{f,\text{auto}}(Tr)} \geq 0.6 \end{cases} \quad (4.7)$$

$$C_{\text{act}} = \begin{cases} \left(\frac{k_{\text{cat,act}} k_{f,\text{act}}}{k_{r,\text{act}} + k_{\text{cat,act}}} - 30 \right)^2 & \text{if } \frac{k_{\text{cat,act}} k_{f,\text{act}}}{k_{r,\text{act}} + k_{\text{cat,act}}} \leq 30 \\ 0 & \text{if } 30 < \frac{k_{\text{cat,act}} k_{f,\text{act}}}{k_{r,\text{act}} + k_{\text{cat,act}}} < 90 \\ \left(\frac{k_{\text{cat,act}} k_{f,\text{act}}}{k_{r,\text{act}} + k_{\text{cat,act}}} - 90 \right)^2 & \text{if } \frac{k_{\text{cat,act}} k_{f,\text{act}}}{k_{r,\text{act}} + k_{\text{cat,act}}} \geq 90 \end{cases} \quad (4.8)$$

$$C_{\text{exp}} = C_{\text{auto}} + C_{\text{act}} \quad (4.9)$$

Additionally, we regulate the weight of the five components of the cost function using the factors f_1, f_2, f_3, f_4 and f_5 . The weight factors consist of a product of two contributions, that is $f_j = w_j c_j$, where w_j denotes the adjustable weight factor ($w_j \in [0,1]$, $\sum_{j=1}^5 w_j = 1$, $w_1 = 0.4$, $w_2 = w_3 = w_4 = w_5 = 0.15$) and c_j denotes the normalization constant that when multiplied with the value of the j -th component of the cost function, on average equals one. Indeed, the normalization constants are determined beforehand by running the optimization procedure a number of times while storing the values of each individual component. The normalization constant c_j is defined as one over the mean of the observed errors of the j -th component $c_j = \frac{1}{\text{mean}(\vec{e}_j)}$ where \vec{e}_j denotes the observed errors resulting from all datasets of the j -th component of the cost function, $j \in \{1, \dots, 5\}$. This cost function is defined in equation (4.10).

$$J(\vec{\theta}) = \sum_{i=1}^M \left(\frac{f_1}{N_i} \sum_{j=1}^{N_i} \left(\frac{y_i^{\text{exp}}(t_j) - y_i(t_j, \vec{\theta})}{\sigma_i(t_j)} \right)^2 + f_2 (\Delta H_i)^2 + f_3 (\Delta F_i)^2 + f_4 (\Delta D_i)^2 + f_5 C_{\text{exp}} \right) \quad (4.10)$$

$$\Delta H_i = H_i^{\text{exp}} - H_i(\vec{\theta}), \Delta F_i = F_i^{\text{exp}} - F_i(\vec{\theta}), \Delta D_i = D_i^{\text{exp}} - D_i(\vec{\theta})$$

Here, $\vec{\theta}$ is the vector containing the parameters in the model, M is the number of datasets to be used for optimization of the model, N_i is the number of datapoints in the i -th dataset and $y_i^{\text{exp}}(t_j)$ is the experimental measurement of the trypsin concentration on timepoint t_j of the i -th experiment. Furthermore, $y_i(t_j, \vec{\theta})$ is the corresponding model prediction of this experimental data point employing parameter set $\vec{\theta}$, and $\sigma_i(t_j)$ is the measurement error of this data point (equals one when unavailable). For each experiment, we also incorporate the quadratic difference between the peak height, the frequency, the damping coefficient and the deviation from experimental intervals of parameter ratios. Minimization of the cost function J provides the optimal parameter set, that is $\vec{\theta}_{\text{opt}} = \arg \min(J(\vec{\theta}))$. Optimization is performed using the Matlab routine `lsqnonlin` with a subspace trust-region method based on the interior-reflective Newton method. Since the parameters are all nonnegative, the parameter space is considered in the logarithmic domain³⁴. This consideration also eases the optimization procedure, as parameter values can vary several orders of magnitude. Latin Hypercube sampling is performed (10^5 sets) within preset parameter bounds, and for each set, optimization is performed, preventing entrapment in local minima of the cost function. Finally, the optimization is constrained within the same bounds as the sampling, depicted in Table 4.2 (section 4.3). Most of these bounds are based on a span of two orders of magnitude, except for the bimolecular autoactivation of trypsinogen²⁸, which is set to four as is it less well known.

From the optimal set, we can assess parameter identifiability with a profile likelihood analysis³⁷. The profile likelihood is calculated individually for each parameter and is defined by (4.11).

$$J_{\text{PL}}(\theta_i) = \min_{\forall j \neq i} (J(\vec{\theta} \mid \theta_i)) \quad (4.11)$$

In this particular case, we will optimize 15 parameters (Table 4.1, with the exception of the concentrations and k_{flow}), and thus obtain 15 profile likelihoods for each performed optimization, that is $i, j \in \{1, \dots, 15\}$. Equation (4.11) implies re-optimization of the model for each preset, fixed

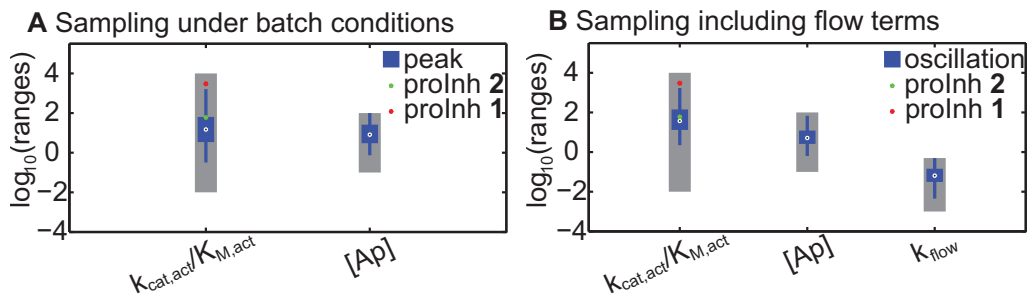


Figure 4.4 Sampling of the BRN, along with the experimentally measured enzymatic efficiencies for nonmethylated (proinhibitor 1) and methylated (proinhibitor 2) ϵ -NH₂ in the lysine residue of the proinhibitor. These are indicated with a red and a green dot, respectively, in both graphs. (A) Box plots of parameter values obtained by response classification in Figure 4.3 A, representing the ranges leading to a single peak. The gray bars indicate the sampling ranges for each parameter. The enzymatic efficiency is expressed in $\text{mM}^{-1} \text{h}^{-1}$ and $[Ap]$ is expressed in 10^{-2} U/mL . (B) Box plots of parameter values obtained by response classification in Figure 4.3B, representing the ranges leading to sustained oscillation. The gray bars indicate the sampling ranges for each parameter. The enzymatic efficiency is expressed in $\text{mM}^{-1} \text{h}^{-1}$, $[Ap]$ is expressed in 10^{-2} U/mL and the flow rate is expressed in h^{-1} .

value of θ_i with respect to all other parameters. In other words, θ_i does not participate in the optimization, so J is kept minimal along θ_i , resulting in J_{PL} as a function of θ_i , i.e. $\theta_i \mapsto J_{PL}(\theta_i)$.

4.3 RESULTS AND DISCUSSION

4.3.1 Sampling of the network

Table 4.2 displays the values of the experimentally determined values of the parameters of the BRN. These values are kept constant and sampling procedures are performed as described in section 4.2.1. The box plots depicted in Figure 4.4A show the ranges of the enzymatic efficiency of trypsin ($\frac{k_{cat,act}}{K_{M,act}}$) and aminopeptidase concentration $[Ap]$ that match the criterion for the trypsin concentration to show a single peak under batch conditions. The enzymatic efficiency consists of the turnover number $k_{cat,act}$ and the Michaelis-Menten constant $K_{M,act}$ of trypsin acting on *proInh*. The gray bars indicate the sampling ranges for each parameter within which a million parameter sets are screened. Therefore, the sampling suggests that the rate of the *Ap*-catalyzed step is in a range accessible by choosing a suitable $[Ap]$, but the enzymatic efficiency of *Tr* acting on **1** is too high (experimental value of $\frac{k_{cat,act}}{K_{M,act}} > 3300 \text{ mM}^{-1} \text{h}^{-1}$, red dot), as a single peak only occurs if the enzymatic efficiency of this enzymatic step is significantly reduced ($\frac{k_{cat,act}}{K_{M,act}} < 1500 \text{ mM}^{-1} \text{h}^{-1}$, green dot). It is known that methylation of the ϵ -NH₂ in Lys lowers the affinity between the substrate and *Tr*, providing a route to molecularly engineer a decreased sensitivity of the negative feedback loop to *Tr*³⁸. This observation is supported by the sampling including flow terms, which is set to screen for sustained oscillations. Figure 4.4B displays the box plots that show the ranges of the enzymatic efficiency of trypsin activating proinhibitor ($\frac{k_{cat,act}}{K_{M,act}}$), aminopeptidase concentration $[Ap]$ and flow rate (k_{flow}) based on the assessment described in section 4.2.1. For sustained oscillations under flow conditions, the enzymatic efficiency of *Tr* acting on **1** is also too high. Again, the gray bars indicate the sampling

ranges for each parameter which are drawn from a sampling of a million parameter sets. Finally, the range of flow rates for the system to show sustained oscillations is accessible.

Table 4.2 Experimentally determined parameter values that are kept constant during sampling (second column), and the lower and upper bounds set for optimization (third and fourth column, respectively).

| Parameter | Experimental value | Lower bound | Upper bound |
|--|--------------------|-------------------|-------------------|
| $k_{f,auto(Tg)}$ [$\text{mM}^{-1} \text{h}^{-1}$] | - | 10^{-6} | 10^{-2} |
| $k_{f,auto(Tr)}$ [$\text{mM}^{-1} \text{h}^{-1}$] | - | 8.2 | 820 |
| $k_{r,auto(Tr)}$ [h^{-1}] | - | 0.204 | 20.4 |
| $k_{cat,auto(Tr)}$ [h^{-1}] | 34 | 3.48 | 348 |
| $K_{M,auto(Tr)}$ [mM] | 0.5* | - | - |
| $k_{f,act}$ [$\text{mM}^{-1} \text{h}^{-1}$] | - | 44.8 | $4.48 \cdot 10^3$ |
| $k_{r,act}$ [h^{-1}] | - | $1.47 \cdot 10^3$ | $1.47 \cdot 10^5$ |
| $k_{cat,act}$ [h^{-1}] | - | 112 | $1.12 \cdot 10^4$ |
| $k_{cat,act}/K_{M,act}$ [$\text{mM}^{-1} \text{h}^{-1}$] | 60** | - | - |
| $k_{cat,delay}$ [$\text{h}^{-1} (\text{U/mL})^{-1}$] | 11.2 | 0.77 | 77 |
| $K_{M,delay}$ [mM] | 2.71 | 0.27 | 27 |
| $k_{f,inh(intInh)}$ [$\text{mM}^{-1} \text{h}^{-1}$] | 1.17 | 0.195 | 19.5 |
| $k_{f,inh(Inh)}$ [$\text{mM}^{-1} \text{h}^{-1}$] | 52 | 4.3 | 430 |
| $k_{f,inh(proInh)}$ [$\text{mM}^{-1} \text{h}^{-1}$] | 1.34 | 0.0234 | 2.34 |
| $k_{f,\phi(intInh)}$ [h^{-1}] | 0.036 | 0.0025 | 0.25 |
| $k_{f,\phi(Inh)}$ [h^{-1}] | 0.034 | 0.0027 | 0.27 |
| $k_{f,\phi(proInh)}$ [h^{-1}] | 0.18 | 0.002 | 0.20 |

$$*K_{M,auto(Tr)} = \frac{k_{r,auto(Tr)} + k_{cat,auto(Tr)}}{k_{f,auto(Tr)}} \text{ with } k_{f,auto(Tr)} = 80 \text{ mM}^{-1} \text{h}^{-1} \text{ and } k_{r,auto(Tr)} = 6 \text{ h}^{-1}.$$

$$**\frac{k_{cat,act}}{K_{M,act}} = \frac{k_{f,act}k_{cat,act}}{k_{r,act} + k_{cat,act}} \text{ with } k_{r,act} = 10^4 \text{ h}^{-1} \text{ and } k_{cat,act} = 10^3 \text{ h}^{-1}.$$

4.3.2 Optimization of the model with time-course data

In this section, we first perform two model optimizations by making use of different sets of data. For the first optimization we only utilize sustained oscillatory datasets that adhere to the criteria discussed in section 4.2.1. For the second optimization we employ all ten available datasets. The datasets were obtained using proinhibitor **2**. The experimental conditions under which these datasets were obtained are depicted in Table 4.3, and those that are considered oscillatory are indicated with a star. Optimization of the parameters in the model is performed as described in section 4.2.2. The two optimal parameter sets are indicated in Table 4.4. The results of these two optimizations are depicted in Figures 4.5 and 4.6, respectively, along with the experimental data. Figure 4.5 displays the time-course simulations after fitting of the model to datasets 1, 2, 3, 8 and 10, along with the predictions of the model with these parameters for the non-oscillating datasets. Figure 4.6 shows the time-course simulations after optimization of the model to all datasets. These time-course simulations show slight deviations from each other, as do the optimal parameter sets, shown in Table 4.4. For the optimization making use of oscillatory sets only, $K_{M,auto(Tr)} = 0.6 \text{ mM}$ and $\frac{k_{cat,act}}{K_{M,act}} = 30.7 \text{ mM}^{-1} \text{h}^{-1}$. This optimization thus satisfies the experiment constraints. For the optimization employing all the datasets, these values are 0.67 mM and $32.1 \text{ mM}^{-1} \text{h}^{-1}$,

respectively, indicating that it is just barely outside the experimental regime. In conclusion, adding the factor defined in equations (4.7)-(4.9) to the cost function takes experimental constraints sufficiently into account.

Table 4.3 Experimental conditions of the acquired datasets. Datasets indicated with a star (*) show sustained oscillations, and are used in a separate optimization procedure. The fixed experimental conditions are $[Tg]_0 = 0.1667$ mM, $[Tr]_0 = 0.2$ μ M and $[proInh]_0 = 1.5$ mM.

| Dataset | Conditions |
|---------|---|
| 1* | $[Ap]_0 = 0.33$ U/mL; $k_{flow} = 0.234$ h ⁻¹ |
| 2* | $[Ap]_0 = 0.22$ U/mL; $k_{flow} = 0.234$ h ⁻¹ |
| 3* | $[Ap]_0 = 0.50$ U/mL; $k_{flow} = 0.234$ h ⁻¹ |
| 4 | $[Ap]_0 = 0.07$ U/mL; $k_{flow} = 0.234$ h ⁻¹ |
| 5 | $[Ap]_0 = 0.33$ U/mL; $k_{flow} = 0.156$ h ⁻¹ |
| 6 | $[Ap]_0 = 0.33$ U/mL; $k_{flow} = 0.375$ h ⁻¹ |
| 7 | $[Ap]_0 = 0.33$ U/mL; $k_{flow} = 0.0585$ h ⁻¹ |
| 8* | $[Ap]_0 = 0.28$ U/mL; $k_{flow} = 0.195$ h ⁻¹ |
| 9 | $[Ap]_0 = 0.67$ U/mL; $k_{flow} = 0.234$ h ⁻¹ |
| 10* | $[Ap]_0 = 0.21$ U/mL; $k_{flow} = 0.12$ h ⁻¹ |

Table 4.4 Optimization results. Optimization 1 was obtained using the five datasets indicated with a star in Table 4.3, optimization 2 resulted from using all ten datasets.

| Parameter | Optimization 1 | Optimization 2 |
|---|---------------------|---------------------|
| $k_{f,auto(Tg)}$ [mM ⁻¹ h ⁻¹] | $1.6 \cdot 10^{-5}$ | $1.5 \cdot 10^{-4}$ |
| $k_{f,auto(Tr)}$ [mM ⁻¹ h ⁻¹] | 80.2 | 47.5 |
| $k_{r,auto(Tr)}$ [h ⁻¹] | 11.1 | 0.93 |
| $k_{cat,auto(Tr)}$ [h ⁻¹] | 38.6 | 31.1 |
| $k_{f,act}$ [mM ⁻¹ h ⁻¹] | 743 | 415 |
| $k_{r,act}$ [h ⁻¹] | $2.18 \cdot 10^4$ | $3.13 \cdot 10^4$ |
| $k_{cat,act}$ [h ⁻¹] | 939 | $2.62 \cdot 10^3$ |
| $k_{cat,delay}$ [h ⁻¹ (U/mL) ⁻¹] | 2.8 | 6.8 |
| $K_{M,delay}$ [mM] | 0.42 | 1.31 |
| $k_{f,inh(intInh)}$ [mM ⁻¹ h ⁻¹] | 12.0 | 12.1 |
| $k_{f,inh(Inh)}$ [mM ⁻¹ h ⁻¹] | 53.2 | 59.1 |
| $k_{f,inh(proInh)}$ [mM ⁻¹ h ⁻¹] | 0.31 | 0.19 |
| $k_{f,\phi(intInh)}$ [h ⁻¹] | 0.0057 | 0.012 |
| $k_{f,\phi(Inh)}$ [h ⁻¹] | 0.0066 | 0.098 |
| $k_{f,\phi(proInh)}$ [h ⁻¹] | 0.15 | 0.16 |

Generally, when comparing the two optimizations, the optimization employing only a part of the data tends to describe all of the data less well than the optimization employing all datasets, even though the contribution of the experimental data to the cost function was normalized in both cases, and contributes in expectation 40%, *i.e.* $w_1 = 0.4$. This suggests that either more data is needed for the model to increase in predictive capabilities (*i.e.* the problem is underdetermined), or that predictive power is limited in this model – meaning that the model does not fully describe the

system. The first cannot be tested in the second optimization, because there is no more data available for validation. Nevertheless, in both cases, the model shows very good qualitative agreement with experimental data, and is able to fairly accurately predict the period times of the oscillations.

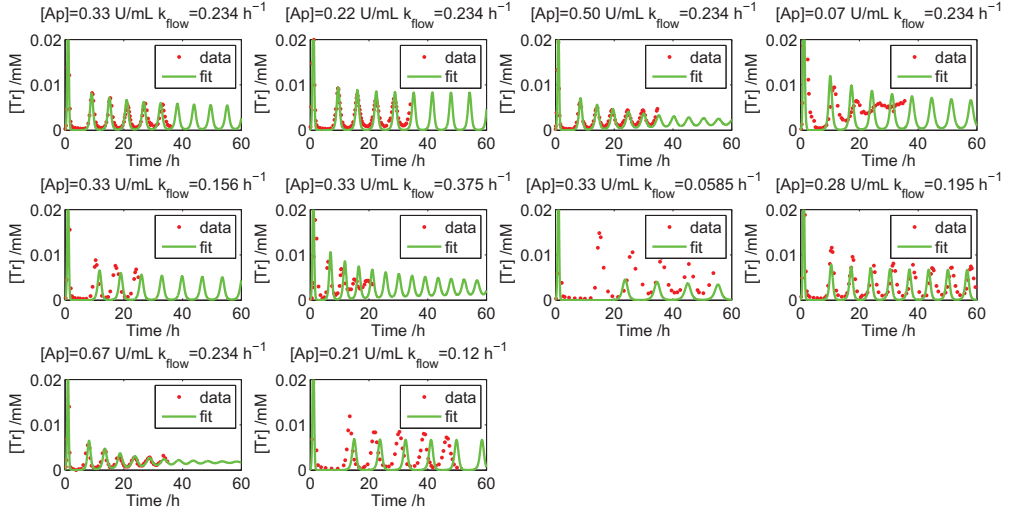


Figure 4.5 Experimental time-course data (red dots) along with the model simulations after optimization using datasets 1, 2 and 3 (first three panels in the top row), 8 (last panel in the second row) and 10 (last panel in the bottom row). Thus, the other panels are model predictions based on this optimization. Noteworthy, the period of oscillations is correctly reproduced and predicted, but predictions for non-oscillatory trajectories do not follow the data very well.

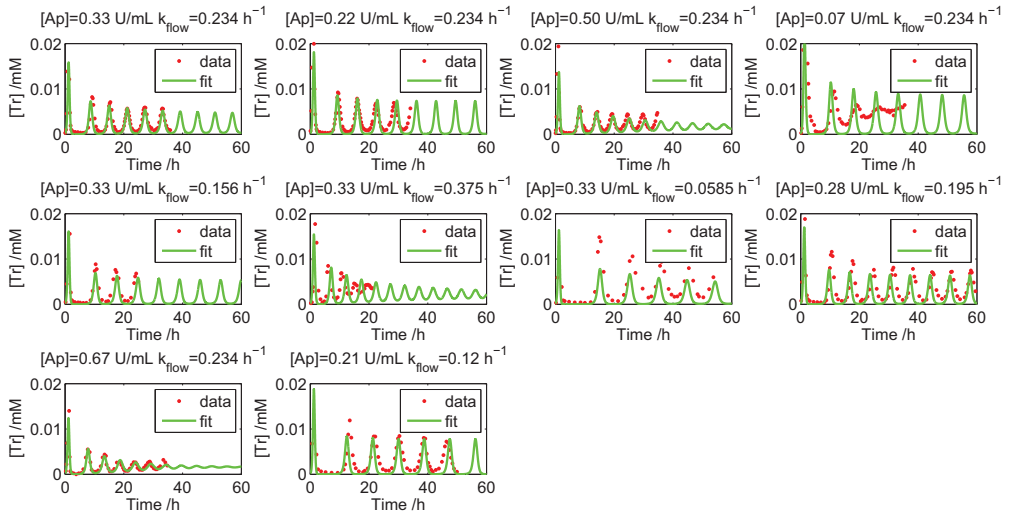


Figure 4.6 Experimental time-course data (red dots) along with the model simulations after optimization using all datasets. Generally, all data is better described than the case where only oscillatory datasets are used to perform the optimization.

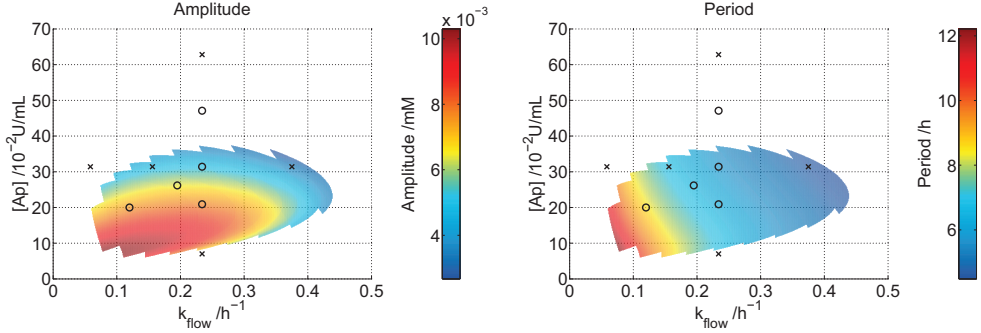


Figure 4.7 Phase diagrams corresponding to sustained oscillatory behavior when varying $[Ap]_0$ and k_{flow} based on optimization using sustained oscillatory datasets only. The oscillatory regimes are colored; all other responses are in the white area. The left diagram is an amplitude heatmap and the right diagram is a period heatmap. For each grid point in the diagrams, simulations of 60 hours were performed and the model output $[Tr]$ was classified as described in section 4.2.1 Figure 4.3. Experimental points are shown as \circ when oscillatory and as \times when non-oscillatory. Note the falsely classified experiments 3, 5 and 6.

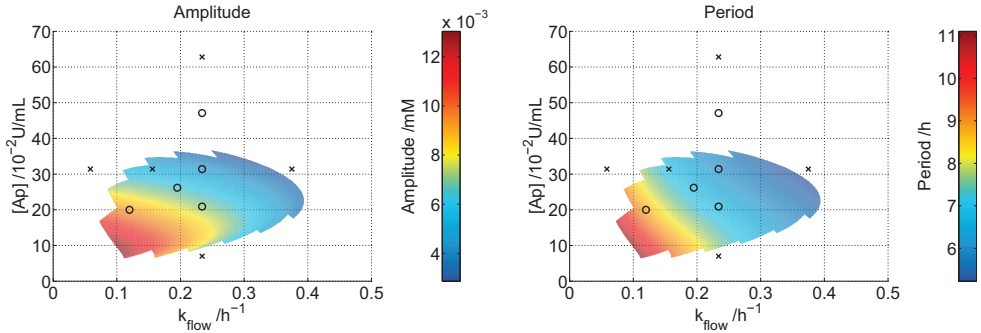


Figure 4.8 Phase diagrams corresponding to sustained oscillatory behavior when varying $[Ap]_0$ and k_{flow} based on optimization using all datasets. The oscillatory regimes are colored; all other responses are in the white area. The left diagram is an amplitude heatmap and the right diagram is a period heatmap. For each grid point in the diagrams, simulations of 60 hours were performed and the model output $[Tr]$ was classified as described in section 4.2.1 Figure 4.3. The oscillatory regime appears to be smaller compared to utilizing oscillatory datasets only. Experimental points are shown as \circ when oscillatory and as \times when non-oscillatory. While both optimizations do not correctly classify experiments 3 and 5, this optimization correctly excludes dataset 6 from the oscillatory regime.

The external conditions that were varied in the experiments, that is $[Ap]_0$ and k_{flow} , are mapped in onto several $[Ap]_0$ versus k_{flow} diagrams. These diagrams indicate the computed regime of sustained oscillations based on the two optimizations, and also show the computed amplitude and period times in Figures 4.7 and 4.8. Figure 4.7 is generated with the optimal parameter set resulting from the optimization using sustained oscillatory datasets only, and Figure 4.8 is constructed with the parameter set obtained from optimization utilizing all datasets. Both optimizations do not contain the $[Ap]_0 = 0.50$ U/mL; $k_{\text{flow}} = 0.234$ h⁻¹ dataset in the oscillatory regime (supported by the third panels in Figures 4.5 and 4.6) and both incorrectly include the $[Ap]_0 = 0.33$ U/mL; $k_{\text{flow}} = 0.156$ h⁻¹ dataset as oscillatory. The optimization employing all the datasets better reproduces the classification of the experiments on sustained oscillatory behavior. While the

oscillatory regimes are not uniquely classified by both optimizations, they strongly overlap in shape, amplitude and period time.

4.3.3 Profile likelihood

The two optimal parameter sets obtained in the previous section show numerical differences, yet qualitatively, they show similar results. The responses and the heatmaps only slightly deviate from each other. Here, we perform a profile likelihood analysis on both optimal parameter sets to assess parameter identifiability. This analysis shows the cost function value with one less degree of freedom, *i.e.* as a function of a preset parameter value not participating in the re-optimization. This is performed first of all to investigate if the optimization utilizing all datasets is indeed superior to the one that employs oscillatory sets only. Second, assessment of the identifiability of the parameters can reveal functional relationships between parameters in the model (*i.e.* structurally unidentifiable) or indicate that more data is needed to confine parameters in a defined range (*i.e.* practically unidentifiable).

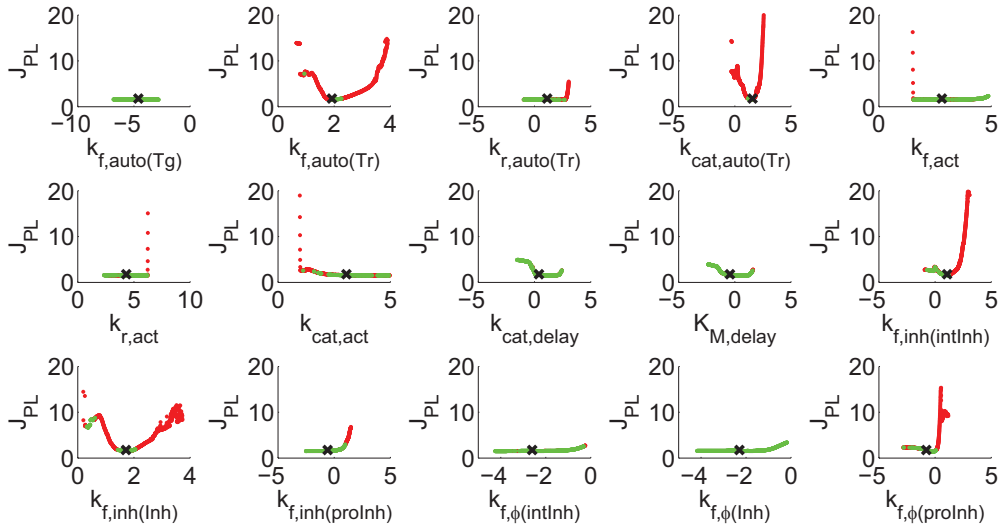


Figure 4.9 Profile likelihoods *versus* the order of magnitude of each parameter (*i.e.* $\log_{10}(\theta_i)$) making use of oscillatory datasets only. The cross indicates the optimal value, and each parameter is varied four orders of magnitude. Green dots indicate points along the profile that adhere to the criteria set for $K_{M,auto(Tr)}$ and $\frac{k_{cat,act}}{K_{M,act}}$, red dots are points that fail to adhere to these criteria during re-optimization. The parameters $k_{f,auto(Tr)}$, $k_{cat,auto(Tr)}$ and $k_{f,inh(Inh)}$ are identifiable, as these contain a unique minimum of the cost function. $k_{cat,delay}$ and $K_{M,delay}$ also appear to be identifiable.

Figures 4.9 and 4.10 show the profile likelihood plots against the order of magnitude of each parameter value. Parameters are varied 4 orders of magnitude from their optimized value, even if this means they go beyond the bounds set during optimization in Table 4.2. The remaining 14 parameters during re-optimization still are strictly bound by these values. The profiles look similar in both profile likelihood trajectories, but the one that is performed utilizing all datasets, generally has a higher cost function value as this function includes more terms. Concluding from both

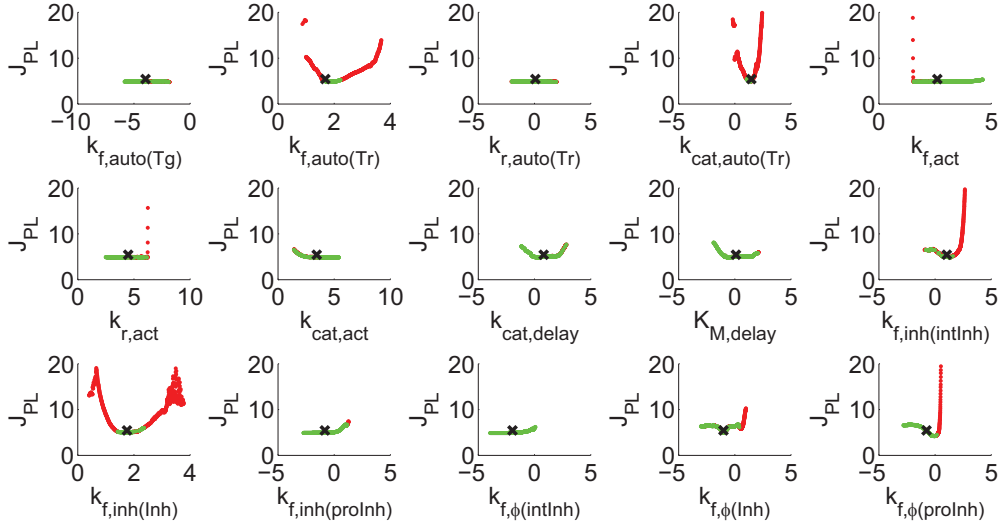


Figure 4.10 Profile likelihoods *versus* the order of magnitude of each parameter (*i.e.* $\log_{10}(\theta_i)$) making use of all datasets. Generally, the cost function value is higher than of the previous optimization, as the cost function contains more terms. The cross indicates the optimal value, and each parameter is varied four orders of magnitude. Green dots indicate points along the profile that adhere to the criteria set for $K_{M,auto}(Tr)$ and $\frac{k_{cat,act}}{K_{M,act}}$, red dots are points that fail to adhere to these criteria during re-optimization. The parameters $k_{f,auto}(Tr)$, $k_{cat,auto}(Tr)$ and $k_{f,inh}(Inh)$ are identifiable, as these contain a unique minimum of the cost function. $k_{cat,delay}$ and $K_{M,delay}$ also appear to be identifiable. In this optimization, $k_{f,phi}(Inh)$ has obtained a tighter upper bound compared to the previous optimization.

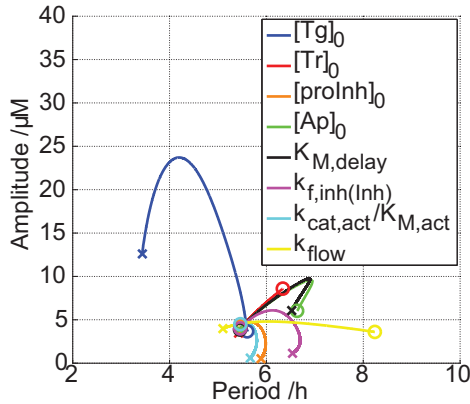
analyses, the parameters $k_{f,auto}(Tr)$, $k_{cat,auto}(Tr)$ and $k_{f,inh}(Inh)$ are identifiable, as these contain a unique minimum of the cost function. The same is true for $k_{cat,delay}$ and $K_{M,delay}$. Moreover, when making use of the cost function in equation (4.10), the inclusion of more data does not provide identifiability of any the remaining parameters (except for $k_{f,phi}(Inh)$, which has obtained a tighter upper bound compared to the optimization utilizing oscillatory sets only), suggesting that these parameters are structurally unidentifiable³⁷. Structural unidentifiability is obvious for $k_{f,auto}(Tg)$, considering this reaction has the same effect in the network as $k_{cat,auto}(Tr)$, which is much larger and leads to the dominant production term of Tr . However, the profiles do suggest that $k_{f,inh}(intInh)$ and $k_{f,phi}(proInh)$ are practically unidentifiable as they still contain obvious minima. This would suggest that the inclusion of yet more data could identify these parameters. The same is true for $k_{f,inh}(proInh)$ and $k_{f,phi}(intInh)$, considering that the inclusion of more data has already provided a defined upper bound for $k_{f,phi}(Inh)$. Therefore, we expect that addition of more data eventually provides identifiability of these parameters, especially when it concerns temporal concentration profiles of other species in the network. Finally, the parameters describing the activity of trypsin on proinhibitor, seem to be structurally unidentifiable, and their sudden increase in cost function value is mainly determined by the constraints defined by (4.8). Indeed, when $k_{f,act}$ rapidly decreases, its contribution to the cost function will rapidly increase as this eventually cannot be compensated for

due to the bounds set on $k_{r,act}$ and $k_{cat,act}$. Similarly, when $k_{r,act}$ increases, so does its contribution to the cost function, and eventually this increase cannot be compensated for by $k_{f,act}$ and $k_{cat,act}$.

4.3.4 Tunability

A final asset we would like to look at in this Chapter is the tunability of the model resulting from both optimizations in terms of amplitude and phase of the network, by varying conditions that can be varied relatively easily experimentally. While it is certainly interesting to include all parameters for this analysis that are discussed in the previous section, we limit ourselves to tunability that can be established experimentally, rather than a theoretical one based on parameters that cannot be affected. This concerns the inflow concentrations $[Tr]_0$, $[Tg]_0$, $[proInh]_0$ and $[Ap]_0$. The parameters $K_{M,delay}$, $k_{f,inh(Inh)}$ and $\frac{k_{cat,act}}{K_{M,act}}$ can be indirectly tuned by chemically modifying or even choosing different substrate molecules. The flow rate k_{flow} can also be affected directly. These parameters are varied between 0 and 10 times their nominal value, while keeping the others constant at their optimal values. Figure 4.11 shows the tunabilities for both optimized models, as long as variation still allows the model to show sustained oscillations. These tunabilities do not differ very much from each other, indicating that the model embedded with both optimal parameter sets similarly predict deviations from experimental points. The central point in both cases corresponds to dataset 1, that is $[Ap]_0 = 0.33$ U/mL; $k_{flow} = 0.234$ h⁻¹. So, as far as tunability is concerned, either using oscillatory datasets only or all ten datasets makes little difference. Generally, the optimization utilizing oscillatory datasets only appears to be more robust in the sense that deviations from the nominal value are less than for the model calibrated with all sets in terms of $[Tg]_0$, while other

A Optimization using oscillatory datasets



B Optimization using all datasets

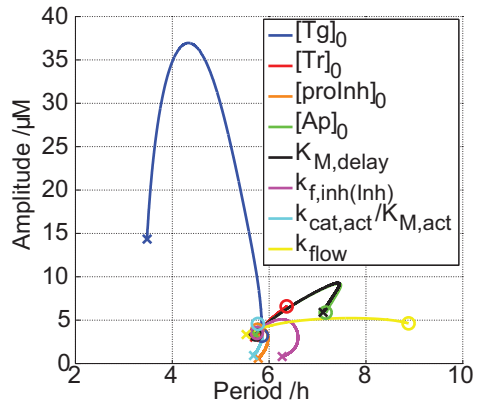


Figure 4.11 Tunability plots of the BRN resulting from both optimizations. Trajectories of parameter tuning are shown for reference experiment (at the intersection of trajectories simulation condition corresponds to dataset 1, that is $[Ap]_0 = 0.33$ U/mL; $k_{flow} = 0.234$ h⁻¹). The trajectories within the maximal (10x) and minimal (0x) variation of the nominal values are only shown for regimes showing sustained oscillating behavior (shown by bounds). Starting points of the oscillatory regime are \circ , ending points are \times . (A) Tunability of the model calibrated with oscillatory datasets only. (B) Tunability of the model calibrated with all data. Generally, the tunabilities obtained by employing both optimized parameter sets in the model show similar behavior. The largest deviation between the two in amplitude is observed for increasing $[Tg]_0$.

trends are similar. Because of the large variation in the amplitude of $[Tr]$ as a result of changing $[Tg]_0$, it may be informative to conduct experiments with increased $[Tg]_0$.

4.4 CONCLUSION

In this Chapter, we have analyzed a computational model describing an enzymatic network capable of showing sustained oscillations in a CSTR. We have shown the usefulness of exhaustive parameter sampling in search for certain model behaviors, and here it was employed to aid in the choice of proinhibitor **2** in the BRN.

We defined a cost function incorporating every necessary aspect of calibrating the model as close as possible to realistic values. We performed two local optimizations from many initial points to prevent entrapment in local minima. One optimization was performed only with oscillatory datasets and one with all available data. The optimization utilizing all data proved to reconstitute a more accurate phase diagram than the other. By comparing the optimization in terms of profile likelihood, on first sight, the addition of more data did not improve the identifiability of most parameters. While some parameters are indeed structurally unidentifiable, others seem practically unidentifiable in which case using more data will eventually provide feasible values for these parameters. If $J(\vec{\theta})$ is statistically well-defined, *i.e.* based on a probability distribution, then this procedure allows assessment of the true confidence intervals of the parameter estimates^{34,37}, but as $J(\vec{\theta})$ is purely empirically determined, unfortunately no statistically useful information can be extracted from this analysis. Nevertheless, the analysis provided a qualitative yet insightful notion of the identifiability of each parameter with the use of the current cost function, and suggests that both optimization problems in this Chapter are underdetermined.

For both optimizations, the amplitude and period of the oscillator can be tuned over a relatively broad range. Furthermore, since the output of this oscillating BRNs is a periodically changing concentration of a catalyst, by coupling multiple reactors, this chemical signal can be modulated and processed further³⁹, without retroactivity effects that negatively influence the oscillatory behavior. Indeed, our experimental collaborators in the group of Prof. W. T. S. Huck demonstrated that spatial separation minimized unwanted crosstalk of the core oscillator with reactions that take place in downstream coupled reactors. One experiment showed the amplification of the signal, another was conducted to achieve a filtering of the oscillating signal into block-shaped pulses, while a third utilized the oscillation to drive an equilibrium process.

4.5 REFERENCES

1. D. Bray (1995), Protein molecules as computational elements in living cells. *Nature* **376**, 307–312.
2. D. E. Koshland, A. Goldbeter and J. B. Stock (1982), Amplification and adaptation in regulatory and sensory systems. *Science* **217**, 220–225.
3. B. N. Kholodenko (2006), Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **7**, 165–176.
4. J. E. Ferrell Jr., T. Y.-C. Tsai and Q. Yang (2011), Modeling the cell cycle: why do certain circuits oscillate? *Cell* **144**, 874–885.
5. M. B. Elowitz and S. Leibler (2000), A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338.
6. T. S. Gardner, C. R. Cantor and J. J. Collins (2000), Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342.
7. J. Hasty, D. McMillen and J. J. Collins (2002), Engineered gene circuits. *Nature* **420**, 224–230.

8. J. Kim, K. S. White and E. Winfree (2006), Construction of an *in vitro* bistable circuit from synthetic transcriptional switches. *Mol. Syst. Biol.* **2**, 68.
9. J. Kim and E. Winfree (2011), Synthetic *in vitro* transcriptional oscillators. *Mol. Syst. Biol.* **7**, 465.
10. K. Montagne, R. Plasson, Y. Sakai, T. Fujii and Y. Rondelez (2011), Programming an *in vitro* DNA oscillator using a molecular networking strategy. *Mol. Syst. Biol.* **7**, 466.
11. E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree and F. C. Simmel (2011), Timing molecular motion and production with a synthetic transcriptional clock. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E784–E793.
12. S. M. Chirieleison, P. B. Allen, Z. B. Simpson, A. D. Ellington and X. Chen (2013), Pattern transformation with DNA circuits. *Nat. Chem.* **5**, 1000–1005.
13. T. Fujii and Y. Rondelez (2013), Predator–prey molecular ecosystems. *ACS Nano* **7**, 27–34.
14. H. Niederholtmeyer, V. Stepanova and S. J. Maerkl (2013), Implementation of cell-free biological networks at steady state. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15985–15990.
15. J. Kim, I. Khetarpal, S. Sen and R. M. Murray (2014), Synthetic circuit for exact adaptation and fold-change detection. *Nucleic Acids Res.* **42**, 6078–6089.
16. I. R. Epstein and K. Showalter (1996), Nonlinear chemical dynamics: oscillations, patterns, and chaos. *J. Phys. Chem.* **100**, 13132–13147.
17. P. De Kepper, I. R. Epstein and K. Kustin (1981), A systematically designed homogeneous oscillating reaction: the arsenite-iodate-chlorite system. *J. Am. Chem. Soc.* **103**, 2133–2134.
18. R. Yoshida, T. Takahashi, T. Yamaguchi and H. Ichijo (1996), Self-oscillating gel. *J. Am. Chem. Soc.* **118**, 5134–5135.
19. A. F. Taylor, M. R. Tinsley, F. Wang, Z. Huang and K. Showalter (2009), Dynamical quorum sensing and synchronization in large populations of chemical oscillators. *Science* **323**, 614–617.
20. J. Horváth, I. Szalai and P. De Kepper (2009), An experimental design method leading to chemical Turing patterns. *Science* **324**, 772–775.
21. B. C. Goodwin (1965), Oscillatory behavior in enzymatic control processes. *Adv. Enzyme Regul.* **3**, 425–437.
22. B. Novák and J. J. Tyson (2008), Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9**, 981–991.
23. J. Boissonade and P. De Kepper (1980), Transitions from bistability to limit cycle oscillations. theoretical analysis and experimental evidence in an open chemical system. *J. Phys. Chem.* **84**, 501–506.
24. A. Goldbeter (2002), Computational approaches to cellular rhythms. *Nature* **420**, 238–245.
25. D. Aubel and M. Fussenegger (2010), Watch the clock—engineering biological systems to be on time. *Curr. Opin. Genet. Dev.* **20**, 634–643.
26. S. N. Semenov, A. S. Y. Wong, R. M. van der Made, S. G. J. Postma, J. Groen, H. W. H. van Roekel, T. F. A. de Greef and W. T. S. Huck (2015), Rational design of functional and tunable oscillating enzymatic networks. *Nat. Chem.* **7**, 160–165.
27. M. García-Moreno, B. H. Havsteen, R. Varón and H. Rix-Matzen (1991), Evaluation of the kinetic parameters of the activation of trypsinogen by trypsin. *Biochim. Biophys. Acta* **1080**, 143–147.
28. J. Kay and B. Kassell (1971), The autoactivation of trypsinogen. *J. Biol. Chem.* **246**, 6661–6665.
29. L. A. Segel and M. Slemrod (1989), The quasi-steady-state assumption: a case study in perturbation. *SIAM Rev.* **31**, 446–477.
30. J. P. Abita, M. Delaage, M. Lazdunski and J. Savrda (1969), The mechanism of activation of trypsinogen. *Eur. J. Biochem.* **8**, 314–324.
31. J. Vanlier, C. A. Tiemann, P. A. J. Hilbers and N. A. W. van Riel (2012), An integrated strategy for prediction uncertainty analysis. *Bioinformatics* **28**, 1130–1135.
32. J. Vanlier, C. A. Tiemann, P. A. J. Hilbers and N. A. W. van Riel (2012), A Bayesian approach to targeted experiment design. *Bioinformatics* **28**, 1136–1142.
33. A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker and C. S. Woodward (2005), SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Software* **31**, 363–396.
34. A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelke, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, U. Klingmüller and J. Timmer (2013), Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE* **8**, e74335.
35. J. Vanlier, C. A. Tiemann, P. A. J. Hilbers and N. A. W. van Riel (2013), Parameter uncertainty in biochemical models described by ordinary differential equations. *Math. Biosci.* **246**, 305–314.
36. M. Weitz, J. Kim, K. Kapsner, E. Winfree, E. Franco and F. C. Simmel (2014), Diversity in the dynamical behavior of a compartmentalized programmable biochemical oscillator. *Nat. Chem.* **6**, 295–302.
37. A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller and J. Timmer (2009), Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929.
38. J. H. Seely and N. L. Benoit (1970), Effect of N-methylation and chain length on kinetic constants of trypsin substrates. Epsilon-N-methyllysine and homolysine derivatives as substrates. *Can. J. Biochem.* **48**, 1122–1131.

39. K. Kurin-Csörgei, I. R. Epstein and M. Orbán (2005), Systematic design of chemical oscillators using complexation and precipitation equilibria. *Nature* **433**, 139–142.

Chapter 5

Evaporative Self-Assembly of Single-Chain, Polymeric Nanoparticles

This work has been published:

Hendrik W. H. van Roekel, Patrick J. M. Stals, Martijn A. J. Gillissen, Peter A. J. Hilbers, Albert J. Markvoort and Tom F. A. de Greef (2013), Evaporative self-assembly of single-chain, polymeric nanoparticles. *Chemical Communications* **49**, 3122–3124.

ABSTRACT

Evaporative self-assembly of dilute solutions containing single-chain polymeric nanoparticles results in characteristic morphologies imaged using atomic force microscopy. In this study, we employed a lattice-gas model that evolves its state by Monte Carlo sampling. The system is set to operate under conditions that simulate small amounts of unstable droplets of liquid solvent and wetting layer remaining on the surface following evaporation of the majority of the solvent, reminiscent of the nonequilibrium conditions in a closed chamber. We perform quantitative comparison of experimental data to morphologies obtained by these lattice-gas simulations. This comparison shows that the nonequilibrium patterns emerge from a complex interplay between dewetting, solvent evaporation and nanoparticle diffusion, and paves the way to ultimately *in silico* predict the morphology of such patterns.

5.1 INTRODUCTION

Single-chain polymeric nanoparticles (SCPNs) have received increasing interest as advances in controlling their structure and stability yield novel stimuli-responsive materials. Such materials can adapt to environmental conditions and as a result they can show self-healing properties, regulate ionic and molecular transport, change their adhesiveness to different species and process or convert (bio)chemical signals^{1–3}. The impressive features of these dynamic materials are system-level properties that arise from a consideration of intramolecular and intermolecular interactions within ordered ensembles. These interactions are in part determined by the structure of polymeric nanoparticles, which in turn can be controlled at the primary level by regulating the precise sequence⁴ as well as at the secondary level *via* covalent^{4–9} and non-covalent^{10–14} interactions. This structural control ultimately determines the size, shape and physicochemical properties of the nanoparticles, and therefore their functionality in solvents or on surfaces. Thus, this information is essential in view of their recent applications as catalysts^{15,16}, sensors¹⁷ and target-specific drug-carriers¹⁸.

Nanoparticles on solid surfaces are commonly characterized by means of atomic force microscopy^{1,3,5,7,10–14,19–29} (AFM)³⁰. Preparation of samples for AFM proceeds by fixation of nanoparticles on a solid surface by complete evaporation of a thin film of a dilute solution containing dissolved polymers. Previous quantitative studies on evaporative self-assembly of inorganic nanoparticles have shown that the resulting morphologies arise from a complex interplay involving evaporation, dewetting and diffusion of nanoparticles^{19–29}. However, for single-chain polymeric nanoparticles such quantitative studies are lacking. Understanding the evaporative self-assembly process is a critical step in ultimately engineering spatial organization of SCPNs on surfaces. This spatial engineering can for example lead to diverse hierarchically constructed, catalytically active materials that can probe and adapt to their environment, perform reaction cascades, and show complex out-of-equilibrium behavior^{1,15–18}. Furthermore, quantitative mapping of experimental AFM images with coarse-grained self-assembly models gives access to microscopic interaction parameters that describe nanoparticle-nanoparticle interactions and nanoparticle-solvent interactions. The attraction between nanoparticles and the degree of solvation of the nanoparticles in certain volatile solvents are important design parameters for constructing such self-assembled materials.

In this Chapter, experimental morphologies of dried solutions containing SCPNs are rationalized by comparing them to 2D lattice-gas simulations. This computational approach has previously been shown to successfully describe evaporative self-assembly of inorganic nanoparticles^{21–29}. In the first part of the Chapter, we describe the model and the evolution of its states by Monte Carlo sampling, and the way in which this procedure leads to a simulated morphology. Then, the experimental and computed morphologies are compared using image processing algorithms. We also briefly discuss these algorithms. The morphological assessments allow for quantitative comparison between experimental observations and theoretical simulations. Finally, we present evidence that evaporative self-assembly of SCPNs proceeds via a spinodal dewetting instability, triggered by decrease of the thickness of the thin liquid film.

5.2 MATERIALS AND METHODS

5.2.1 Experimental

In this study, we have used poly(methyl methacrylate) (PMMA) based polymers functionalized with pendant *o*-nitrobenzyl protected 2-ureidopyrimidinone (UPy, Figure 5.1) moieties, as synthesized by Dr. P. J. M. Stals. Previous work has shown that UV irradiation of these polymers results in cleavage of the *o*-nitrobenzyl protecting groups causing chain collapse^{11–13} driven by strong intramolecular hydrogen bonding between pendant UPy groups. The PMMA polymer has a number averaged molecular weight $M_n = 67900 \text{ g mol}^{-1}$ and a polydispersity index of 1.43 as measured using gel permeation chromatography (GPC, CHCl_3 /polystyrene standards). Irradiation of the polymers with 350 nm light for 2 hours was followed by a series of dilution steps to obtain three solutions containing SCPNs in dioxane at $0.0025 \text{ mg mL}^{-1}$, 0.01 mg mL^{-1} and 0.05 mg mL^{-1} , resulting in low, intermediate and high surface coverage, respectively (*vide infra*). On average, the polymer has 47 UPy groups per chain. Details of the synthesis and physical characterization of the folding of this polymer using GPC and dynamic light scattering (DLS) are discussed elsewhere³¹. In short, detailed DLS measurements confirm that the polymer has a globular shape after deprotection of the protected UPy groups.

Solvents used during measurements were of analytical grade quality or better and obtained from Biosolve. Atomic force micrographs were recorded by Dr. M. A. J. Gillissen under ambient conditions with silicon cantilever tips (PPP-NCH, 300–330 kHz, 42 N m^{-1} from Nanosensors) using an Asylum Research MFP-3D-Bio in non-contact mode. Samples were prepared in two different ways. For the sample with intermediate areal coverage ($\phi \approx 60\%$, Figure 5.2B), a drop of a 0.01 mg mL^{-1} solution of polymer in dioxane was spincoated on a freshly cleaved mica surface (grade V1, Ted Pella Inc.) at 2500 rpm for 60 seconds. The samples with low coverage ($\phi \approx 0.17\%$, Figure 5.2A) and high coverage ($\phi \approx 90\%$, Figure 5.2C) were obtained by dropcasting $5 \text{ }\mu\text{L}$ of a $0.0025 \text{ mg mL}^{-1}$ solution of polymer in dioxane (low coverage) and $10 \text{ }\mu\text{L}$ of 0.05 mg mL^{-1} solution of polymer in dioxane (high coverage) on a freshly cleaved mica surface (grade V1, Ted Pella Inc.) after which the sample was allowed to equilibrate for three days in a solvent chamber filled with dioxane. All samples were measured at different positions on the surface of each sample and in all cases comparable topologies were observed on each position. Images were processed using Scanning Probe Image Processor software (Image Metrology A/S). The areal coverages of SCPNs have been estimated using a threshold determined by Otsu's method (*vide infra*, section 5.2.3).

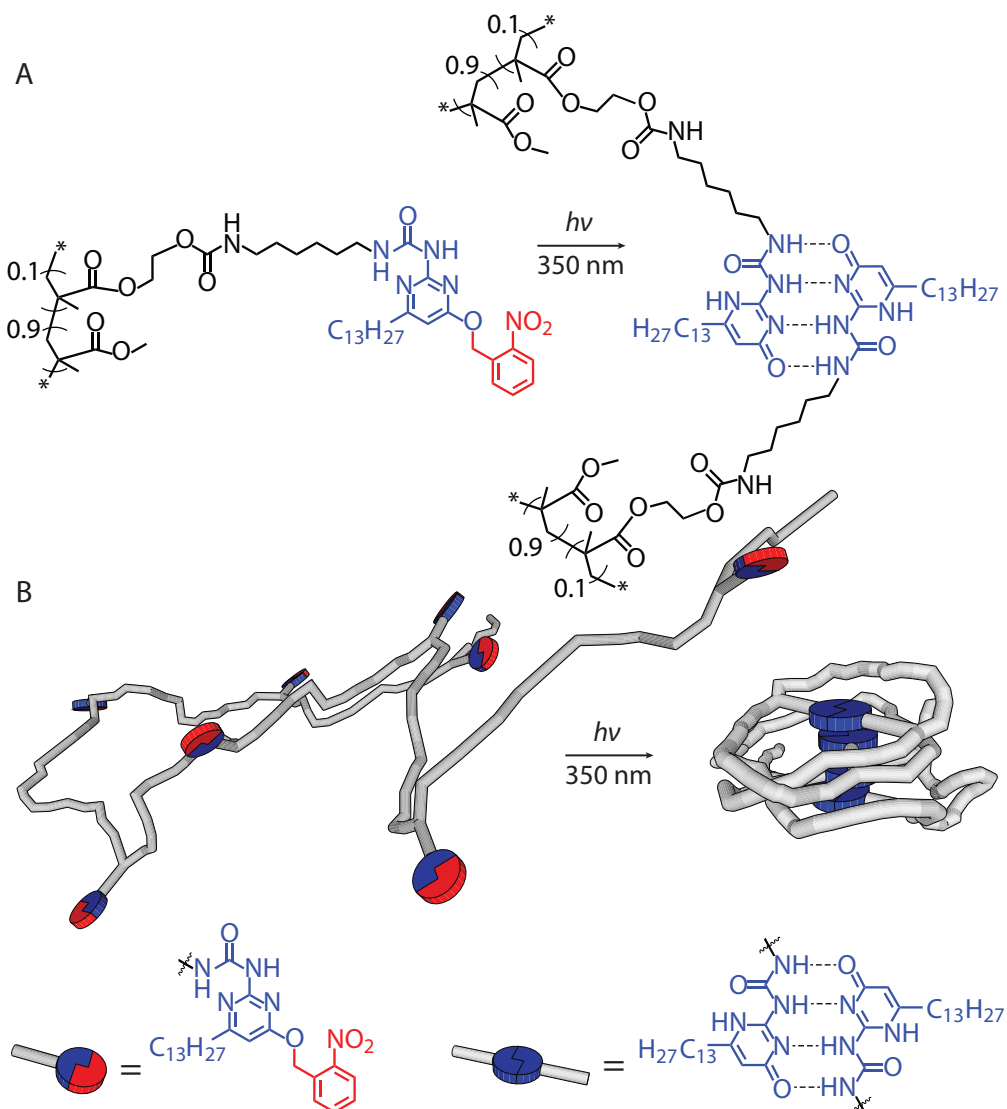


Figure 5.1 Intramolecular collapse of a UPy functionalized^{11–13} PMMA polymer resulting in SCPNs. (A) Reaction scheme depicting cleavage of the protecting groups (red) by UV irradiation and the formation of intramolecular hydrogen bonds between two UPy moieties (blue); (B) graphical representation of the collapse of a single polymeric chain by intramolecular hydrogen-bonding between pendant groups.

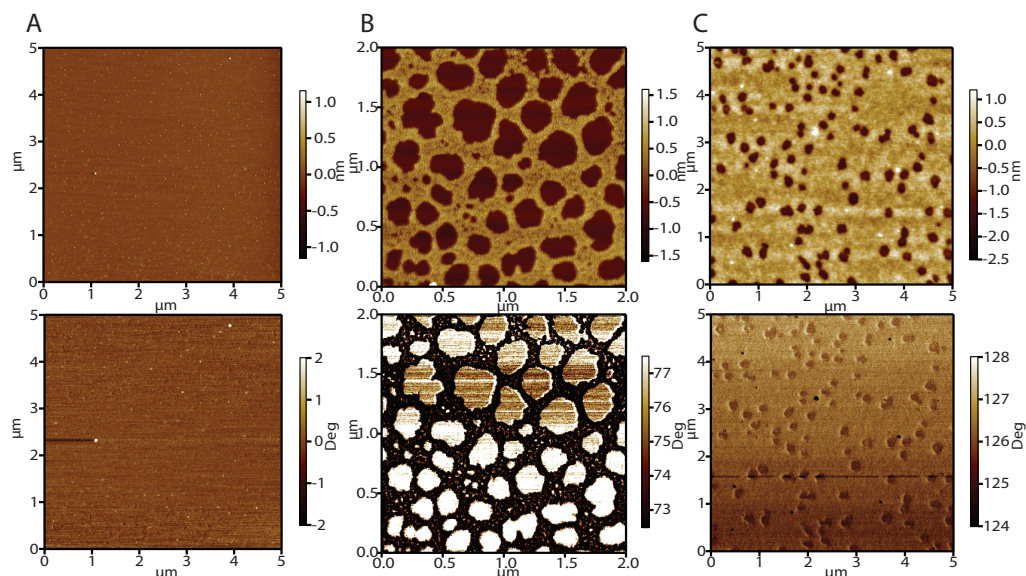


Figure 5.2 Atomic force micrographs of self-assembled morphologies. (A) Single particles dispersed on the surface at a coverage of $\phi \approx 0.17\%$: (top) height image; (bottom) phase image. Images have been measured after dropcasting 5 μL of a $0.0025 \text{ mg mL}^{-1}$ solution of polymer in dioxane on mica. (B) Aggregation of SCPNs at a coverage of $\phi \approx 60\%$: (top) height image; (bottom) phase image. Images have been measured after spin-coating a 0.01 mg mL^{-1} solution of polymer in dioxane on mica. (C) A continuous film with holes formed through close aggregation of SCPNs at a coverage of $\phi \approx 90\%$: (top) height image; (bottom) phase image. Images have been measured after dropcasting 10 μL of a 0.05 mg mL^{-1} solution of polymer in dioxane on mica. Areal coverages have been estimated using a threshold determined by Otsu's method (*vide infra*).

5.2.2 Computational

The lattice-gas model developed by Rabani²¹ is used here to rationalize the observed morphologies as measured by atomic force microscopy (AFM). First, it is assumed that the relevant processes governing evaporative self-assembly of polymeric nanoparticles can be captured in a two-dimensional model, thus neglecting any changes in film-thickness. Second, it is assumed that all relevant dynamics arise from the diffusion of nanoparticles and evaporation of the solvent. These assumptions have proven to be adequate in understanding the formation of polygonal networks and spinodal structures of inorganic nanoparticles resulting from the dewetting of an ultrathin liquid film that remains behind the mesoscopic dewetting front, where the nanoparticles are regarded as a two-dimensional subsystem^{25,27,28}. At an initial diameter of $\sim 10 \text{ nm}$, globular polymeric nanoparticles are comparable in size to inorganic nanoparticles^{12,25,27,28}, so their diffusive properties ought to be similar. For a film-thickness equal to or below 10 nm and about equal to nanoparticle thickness, convective transport of the solution can be neglected²⁷. The height of the polymeric nanoparticle after evaporation¹² is smaller than 10 nm . Furthermore, intermolecular interactions between polymeric nanoparticles are not 'sticky', allowing association and dissociation of diffusing particles. And finally, through AFM, single layers of polymeric nanoparticles are measured. Therefore, the aforementioned assumptions underlying the Monte Carlo model seem very reasonable for our systems with polymeric nanoparticles as well.

Hamiltonian

The Monte Carlo model utilizes a Hamiltonian which defines the interactions between adjacent lattice cells and takes into account the chemical potential. We define a square lattice, where a single lattice cell i can be occupied by solvent, either in the liquid state of matter ($l_i = 1$, $n_i = 0$) or in the gaseous state of matter ($l_i = 0$, $n_i = 0$), or by a part of a nanoparticle ($l_i = 0$, $n_i = 1$). A polymeric nanoparticle is represented by 3×3 lattice cells, and will keep this shape throughout the simulation. An illustration of the computational grid is depicted in Figure 5.3A. Assuming that the interaction between two liquid particles, two nanoparticles or between a liquid particle and a nanoparticle is much higher than any interaction with gas, three interaction terms are introduced. The liquid-liquid interaction term ε_{ll} reflects interactions between liquid lattice cells, the nanoparticle-nanoparticle interaction term ε_{nl} reflects adhesive interactions between nanoparticles and liquid particles, and the nanoparticle-nanoparticle interaction term ε_{nn} reflects the adhesive interactions between the nanoparticles. Furthermore, the solvent in the liquid phase has a chemical potential, μ_1 , which is higher than the chemical potential of the solvent in the gas phase, μ_2 . Solvent molecules tend to move from areas of higher chemical potential to lower chemical potential³². Polymeric nanoparticles follow random walk statistics, meaning that a nanoparticle can move in all directions with equal probability. The Hamiltonian of the system can now be written as in equation (5.1).

$$H = -\varepsilon_{ll} \sum_{\langle ij \rangle} l_i l_j - \varepsilon_{nl} \sum_{\langle ij \rangle} n_i l_j - \varepsilon_{nn} \sum_{\langle ij \rangle} n_i n_j - \mu \sum_i l_i \quad (5.1)$$

Here, $\langle ij \rangle$ indicates summation over nearest neighbour interactions and μ is the effective chemical potential comprising both the liquid-gas chemical potential $\mu_2 - \mu_1$ and the homogeneous interaction of liquid with the substrate²⁶.

In the computational implementation, we first define the entire lattice as a matrix called L . If a certain lattice cell i represents solvent in the gas phase, $L_i = 0$, if it represents solvent in the liquid phase, $L_i = 1$, and if it represents a single lattice cell of a 3×3 nanoparticle, $L_i = 2$. With the lattice matrix defined, the Hamiltonian can now be written in terms of L as depicted in equation (5.2).

$$H = -\varepsilon_{ll} \sum_{\langle ij \rangle} \delta_{L_j,1} \delta_{L_i,1} - \varepsilon_{nl} \left(\sum_{\langle ij \rangle} \delta_{L_i,1} \delta_{L_j,2} + \sum_{\langle ij \rangle} \delta_{L_i,2} \delta_{L_j,1} \right) - \varepsilon_{nn} \sum_{\langle ij \rangle} \delta_{L_i,2} \delta_{L_j,2} - \mu \sum_i \delta_{L_i,1} \quad (5.2)$$

In equation (5.2), $\delta_{a,b}$ represents the Kronecker delta, and the summations run over two adjacent lattice cells: the one on the right and the one below the current lattice cell i . This manner of summation avoids counting interactions twice. Our simulations employ periodic boundary conditions, that is, cells on the boundary of the lattice have interactions with cells on the opposite side of the lattice. Lastly, L_i^Ω indicates that a lattice cell on the edge of a 3×3 nanoparticle is considered.

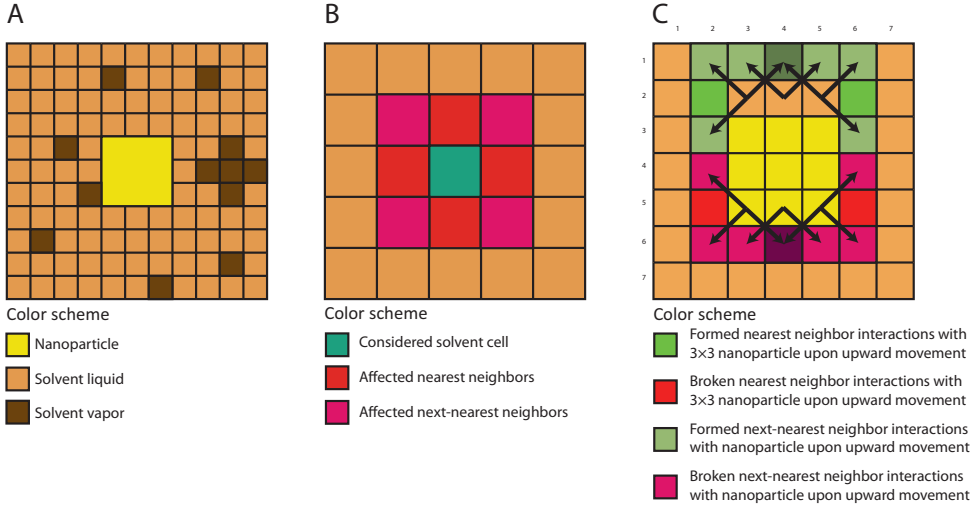


Figure 5.3 Simulation lattice and local considerations. (A) A color scheme representation of a lattice (11×11) containing solvent, gas and nanoparticle cells. In this configuration, the nanoparticle is unable to move to the left, as one of the neighboring cells is in the gaseous state. (B) A solvent cell under consideration (turquoise) with its nearest neighboring lattice cells (red) and next-nearest neighboring lattice cells (purple) relevant for the evaluation of the Hamiltonian difference upon solvent evaporation/condensation. (C) Diffusion of a nanoparticle in the upward direction on a sublattice S : diffusion results in the loss of the interactions with the next-nearest neighbor lattice sites indicated in purple, and new interactions are formed with the next-nearest neighbor lattice sites indicated in turquoise. Conversely, the liquid lattice cells lose next-nearest neighbor interactions with the lattice cells indicated in turquoise and form new interactions with the next-nearest neighboring lattice cells in purple. The diagonal arrows indicate the next-nearest neighbor interactions, and show why some lattice cells appear twice in equation (5.5) (dark purple and dark turquoise).

Next-nearest neighbor Hamiltonian

In order to obtain more realistic simulation results, Moriarty and coworkers²² introduced the use of next-nearest neighbor interactions when calculating the energy change upon a solvent transition or diffusion step. Particularly, it was shown that the inclusion of these next-nearest neighbor interactions, scaled with $1/\sqrt{2}$ to reflect a linear decay in interaction strength with distance, resulted in an increased isotropy in the evaporation of the solvent^{22–29}. Hence, the Hamiltonian becomes (5.3)

$$\begin{aligned}
 H = & \frac{\sqrt{2}}{1+\sqrt{2}} \left[-\varepsilon_{ll} \sum_{\langle ij \rangle} \delta_{L_i,1} \delta_{L_j,1} - \varepsilon_{nl} \left(\sum_{\langle ij \rangle} \delta_{L_i,1} \delta_{L_j^{\Omega},2} + \sum_{\langle ij \rangle} \delta_{L_i^{\Omega},2} \delta_{L_j,1} \right) - \varepsilon_{nn} \sum_{\langle ij \rangle} \delta_{L_i^{\Omega},2} \delta_{L_j^{\Omega},2} + \right. \\
 & \left. - \frac{1}{\sqrt{2}} \left(\varepsilon_{ll} \sum_{\langle ik \rangle} \delta_{L_k,1} \delta_{L_i,1} + \varepsilon_{nl} \left(\sum_{\langle ik \rangle} \delta_{L_i,1} \delta_{L_k^{\Omega},2} + \sum_{\langle ik \rangle} \delta_{L_i^{\Omega},2} \delta_{L_k,1} \right) + \varepsilon_{nn} \sum_{\langle ik \rangle} \delta_{L_i^{\Omega},2} \delta_{L_k^{\Omega},2} \right) \right] - \mu \sum_i \delta_{L_i,1}
 \end{aligned} \quad (5.3)$$

Renormalization of $\sqrt{2}/(1+\sqrt{2})$ is introduced to provide a similar global energy for the next-nearest neighbour model as for the original, nearest neighbor model in equation (5.2), where $\langle ij \rangle$ indicates summation over the adjacent cells of current cell i (to the right and below) and $\langle ik \rangle$ indicates summation over the bottom right and bottom left cells of current cell i .

Local Hamiltonian differences

Simulations start with a lattice that is initially covered with randomly placed nanoparticles and further filled with liquid solvent. Updates are performed using the Metropolis algorithm³³, where a single Monte Carlo step is defined such that in expectation, each solvent cell is considered once for evaporation or condensation, and each nanoparticle is considered a certain number of times (MR, the mobility ratio) for diffusion in one of the directions left, right, up or down^{21,22,24–29}. Upon a local move not the entire Hamiltonian needs to be recalculated, instead ΔH can be calculated locally. Figure 5.3B depicts a solvent cell $L_{i,j}$ under consideration for evaporation or condensation and the lattice cells which are used in determining the change in the Hamiltonian, ΔH . Note the two-dimensional notation $L_{i,j}$, for ease of evaluating local Hamiltonian differences. If the considered solvent cell is in the liquid phase, the local change of the Hamiltonian is defined in equation (5.4).

$$\Delta H_{\text{evap}} = \frac{\sqrt{2}}{1+\sqrt{2}} \left[\epsilon_{ll} (\delta_{L_{i-1,j},1} + \delta_{L_{i,j-1},1} + \delta_{L_{i+1,j},1} + \delta_{L_{i,j+1},1}) + \epsilon_{nl} (\delta_{L_{i-1,j},2} + \delta_{L_{i,j-1},2} + \delta_{L_{i+1,j},2} + \delta_{L_{i,j+1},2}) + \right. \\ \left. + \frac{1}{\sqrt{2}} \{ \epsilon_{ll} (\delta_{L_{i-1,j-1},1} + \delta_{L_{i-1,j+1},1} + \delta_{L_{i+1,j-1},1} + \delta_{L_{i+1,j+1},1}) + \epsilon_{nl} (\delta_{L_{i-1,j-1},2} + \delta_{L_{i-1,j+1},2} + \delta_{L_{i+1,j-1},2} + \delta_{L_{i+1,j+1},2}) \} \right] + \mu \quad (5.4)$$

If this solvent cell is in the gas phase, the change is $\Delta H_{\text{cond}} = -\Delta H_{\text{evap}}$. Figure 5.3C displays a nanoparticle in a graphical 7×7 submatrix S and shows the affected nearest neighbor and next-nearest neighbor lattice sites when the particle would move upward. The associated change in Hamiltonian upon movement of the nanoparticle in the upward direction using nearest and next-nearest neighbor interactions can now be expressed as in equation (5.5). The derivation is similar for the movement of a nanoparticle into the other directions, except that other lattice cells in S should be considered.

$$\Delta H = \frac{\sqrt{2}}{1+\sqrt{2}} \left[(\epsilon_{ll} - \epsilon_{nl}) (\delta_{S_{1,3},1} + \delta_{S_{1,4},1} + \delta_{S_{1,5},1} + \delta_{S_{2,2},1} + \delta_{S_{2,6},1}) + \right. \\ + (\epsilon_{nl} - \epsilon_{nn}) (\delta_{S_{1,3},2} + \delta_{S_{1,4},2} + \delta_{S_{1,5},2} + \delta_{S_{2,2},2} + \delta_{S_{2,6},2}) + \\ + (\epsilon_{nl} - \epsilon_{ll}) (\delta_{S_{6,3},1} + \delta_{S_{6,4},1} + \delta_{S_{6,5},1} + \delta_{S_{5,2},1} + \delta_{S_{5,6},1}) + \\ + (\epsilon_{nn} - \epsilon_{nl}) (\delta_{S_{6,3},2} + \delta_{S_{6,4},2} + \delta_{S_{6,5},2} + \delta_{S_{5,2},2} + \delta_{S_{5,6},2}) + \\ \left. + \frac{1}{\sqrt{2}} \{ (\epsilon_{ll} - \epsilon_{nl}) (\delta_{S_{1,2},1} + \delta_{S_{1,3},1} + 2\delta_{S_{1,4},1} + \delta_{S_{1,5},1} + \delta_{S_{1,6},1} + \delta_{S_{2,2},1} + \delta_{S_{2,6},1}) + \right. \\ + (\epsilon_{nl} - \epsilon_{nn}) (\delta_{S_{1,2},2} + \delta_{S_{1,3},2} + 2\delta_{S_{1,4},2} + \delta_{S_{1,5},2} + \delta_{S_{1,6},2} + \delta_{S_{2,2},2} + \delta_{S_{2,6},2}) + \\ + (\epsilon_{nl} - \epsilon_{ll}) (\delta_{S_{6,2},1} + \delta_{S_{6,3},1} + 2\delta_{S_{6,4},1} + \delta_{S_{6,5},1} + \delta_{S_{6,6},1} + \delta_{S_{4,2},1} + \delta_{S_{4,6},1}) + \\ \left. + (\epsilon_{nn} - \epsilon_{nl}) (\delta_{S_{6,2},2} + \delta_{S_{6,3},2} + 2\delta_{S_{6,4},2} + \delta_{S_{6,5},2} + \delta_{S_{6,6},2} + \delta_{S_{4,2},2} + \delta_{S_{4,6},2}) \} \right] \quad (5.5)$$

A move (*i.e.* diffusion, condensation or evaporation) is accepted with the Metropolis acceptance probability P_{acc} as given in equation (5.6).

$$P_{\text{acc}} = \min \left(1, e^{-\frac{\Delta H}{kT}} \right) \quad (5.6)$$

Here, T is the temperature and k the Boltzmann constant. If the lattice contains N^2 cells and n nanoparticles ($9n < N^2$), the Metropolis algorithm for a single Monte Carlo (MC) step in the simulations is now summarized as follows:

- 1) Diffusion of nanoparticles is examined first, the following steps are repeated $MR \times n$ times:
 1. Randomly select a nanoparticle in the lattice.
 2. Choose a direction into which it will move with probability 0.25 (up, down, left or right).
 3. If the chosen direction has three liquid cells filled adjacent to the nanoparticle, determine ΔH using equation (5.5), if not, this trial move is rejected.
 4. If step three is successful, accept the move using the Metropolis acceptance probability in equation (5.6).
- 2) Then, phase change of solvent is considered, the following steps are repeated $N^2 - 9n$ times:
 1. Randomly select a solvent cell in the lattice.
 2. If the solvent cell is liquid, determine ΔH_{evap} as in equation (5.4), if the solvent cell is gas, determine $\Delta H_{\text{cond}} = -\Delta H_{\text{evap}}$.
 3. Change the state of matter of the solvent lattice cell using the Metropolis acceptance probability in equation (5.6).

5.2.3 Image analysis

The AFM height images were subjected to image preprocessing steps. The images were converted to binary images using a threshold determined by Otsu's method^{34,35}, so that the percentage of coverage of nanoparticles could be estimated (Figure 5.2). When necessary, noise was reduced in the binary images using median filtering³⁵. This noise reduction was sometimes necessary in order for germs used in the Minkowski functional grain growth analysis^{22,36} (*vide infra*) to be placed automatically on particles or in holes, *i.e.* at their centroids.

For simulated images, binary images were obtained by multiplying all components of the lattice matrix L by 0.5 and thresholding the resulting matrix at 0.5. All components in the matrix $0.5L$ that are larger than 0.5 represent nanoparticles in the simulation, and hence the thresholded matrix represents the final simulation morphology. Median filtering was required for the images corresponding to 60% and 90% coverage, in order to remove artifacts (*i.e.* small holes) from the simulations.

Minkowski functional grain growth analysis

The Minkowski functionals—*i.e.* the total area of white pixels A , the total perimeter of white pixels U and the Euler characteristic χ —are calculated for all images as a function of r , where $2r + 1$ is the square grain side length. The principle is outlined in Figure 5.4. As the number of pixels in the images is likely to differ, these functionals are made independent of the image size by normalization and plotting them against the normalized distance. The normalized Minkowski functionals are defined as (5.7)³⁶

$$A^* = \rho \frac{A}{M}, \quad U^* = \sqrt{\rho} \frac{U}{M}, \quad \chi^* = \frac{\chi}{M} \quad (5.7)$$

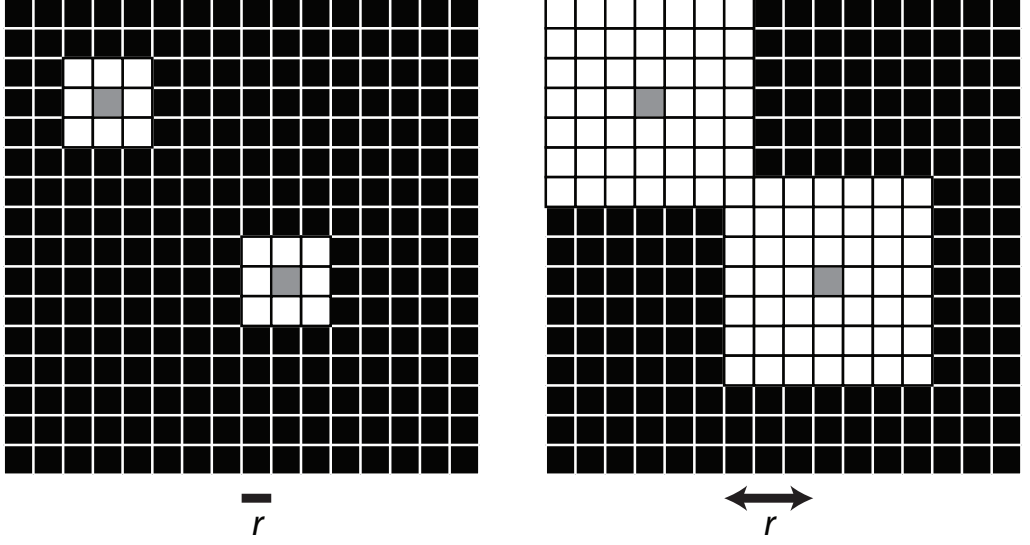


Figure 5.4 Principle of Minkowski grain growth analysis. The Minkowski functionals are determined for each increment in r , where the square grain side length is $a = 2r + 1$. In the left image, $r = 1$ and $a = 3$; in the right image, $r = 3$ and $a = 7$.

where M represents the number of germs in the image and $\rho = \frac{M}{N^2}$ is the germ density in the image having N^2 pixels. The deviation from the expected normalized Minkowski functionals for a random point set gives an indication whether the germ distribution on the image is the result of a physical mechanism, or the result of chance alone³⁶⁻³⁹. The expected normalized Minkowski functionals for growing square grains from random germs are defined as (5.8)³⁶

$$\langle A^* \rangle = 1 - e^{-\rho a^2}, \quad \langle U^* \rangle = 4a\sqrt{\rho}e^{-\rho a^2}, \quad \langle \chi^* \rangle = (1 - \rho a^2)e^{-\rho a^2} \quad (5.8)$$

with $a = 2r + 1$. To compare the calculated normalized Minkowski measures with their matching expectation for a random point set, we need to express the functionals as a function of the normalized distance x between germs, again to account difference in image sizes. The normalized distance is defined as $x = \frac{r}{s_{\text{mean}}}$ with s_{mean} the mean separation distance, given in (5.9)

$$s_{\text{mean}} = \frac{1}{M} \sum_{i=1}^M \min_{i \neq j} (D_{ij}) \quad (5.9)$$

where D is an $M \times M$ matrix containing all distances between all germs and M is the number of germs in the image. For every germ i taken as a reference, the distance to its closest germ is found. This is the reason that we find the minimum of all distances to all other germs j . The mean separation distance is then calculated by taking the average of these closest distances.

5.3 RESULTS AND DISCUSSION

Estimates of the experimental coverage of SCPNs (ϕ) are obtained by image processing the atomic force micrographs images and are $\phi \approx 0.17\%$ for the $0.0025 \text{ mg mL}^{-1}$ solution, $\phi \approx 60\%$ for the 0.01 mg mL^{-1} solution and $\phi \approx 90\%$ for the 0.05 mg mL^{-1} solution of polymer. These estimates of the areal coverages are used as input parameters for simulations with the lattice-gas model. In order to rationalize the mechanism by which nonequilibrium patterns of single-chain polymeric nanoparticles arise during drying, a quantitative comparison between the experimental AFM images and computed morphologies was performed.

Exploration of the lattice-gas model in μ, T space shows that it is capable of producing a wealth of different morphologies. Depending on the type of simulation—which is a direct result of the choice of the interaction parameters, the effective chemical potential and the temperature—500 to 20,000 MC steps are required for most solvent cells to turn into vapor. These morphologies originate either from a nucleation growth or a spinodal dominated dewetting process. The lattice-gas model is capable of simulating these types of self-assembly mechanisms^{21,22,24–29}. Furthermore, the two-dimensional Monte Carlo model can be transformed in a pseudo three-dimensional model by expressing the effective chemical potential as a function of the vapor fraction^{25,26,32}. We found that simulations performed in the early spinodal regime most closely resemble the morphologies as observed on the atomic force micrographs. In the early spinodal regime, the effective chemical potential is chosen such that when the system is mapped onto an Ising lattice, there is approximately no external magnetic field, that is $\mu \approx -2\epsilon_{ll} - 2\phi(\epsilon_{nl} - \epsilon_{ll})$, and the temperature is close to the critical temperature, that is $kT \approx kT_c^\infty \approx 0.57\epsilon_{ll}$. Indeed, as the areal coverage $\phi \in [0,1]$ of SCPNs increases, the effective chemical potential becomes more negative. Figure 5.5 shows the experimental atomic force micrographs along with the corresponding computed morphologies using simulation parameters corresponding to the early spinodal regime, *i.e.* with a sufficiently low chemical potential at a temperature near the critical temperature. The computed morphologies showing the best correspondence to the experimental images are presented in Figure 5.5.

In order to rationalize the mechanism by which nonequilibrium patterns of single-chain polymeric nanoparticles arise during drying, a quantitative comparison between the experimental AFM images and computed morphologies was performed. The distribution of points, *i.e.* particles or hole centers, is compared to the expectation for randomly (Poisson) distributed points using Minkowski functional grain growth analysis³⁶. Such analyses of the Minkowski measures, *i.e.* the normalized total area of white pixels A^* , the normalized total perimeter of white pixels U^* and the normalized Euler characteristic χ^* , have previously been used to rationalize the origin of self-organized morphologies resulting from the evaporation of inorganic nanoparticles²². Strong deviations of the calculated Minkowski measures from the expected Minkowski measures as a function of the normalized grain size x are indicative of a spinodal dewetting instability^{37–39}. We observe for all coverages of SCPNs that the normalized Minkowski functionals as a function of the normalized grain size deviate significantly from the expectation for a random point set, which is a strong indication that the experimental morphologies arise via a spinodal dewetting instability. The normalized Minkowski functionals are depicted in Figures 5.6A, 5.7A and 5.8A for experimental data, for 0.17% coverage, 60% coverage and 90% coverage, respectively, along with the graphs that indicate the deviation from the expectation for a random point set.

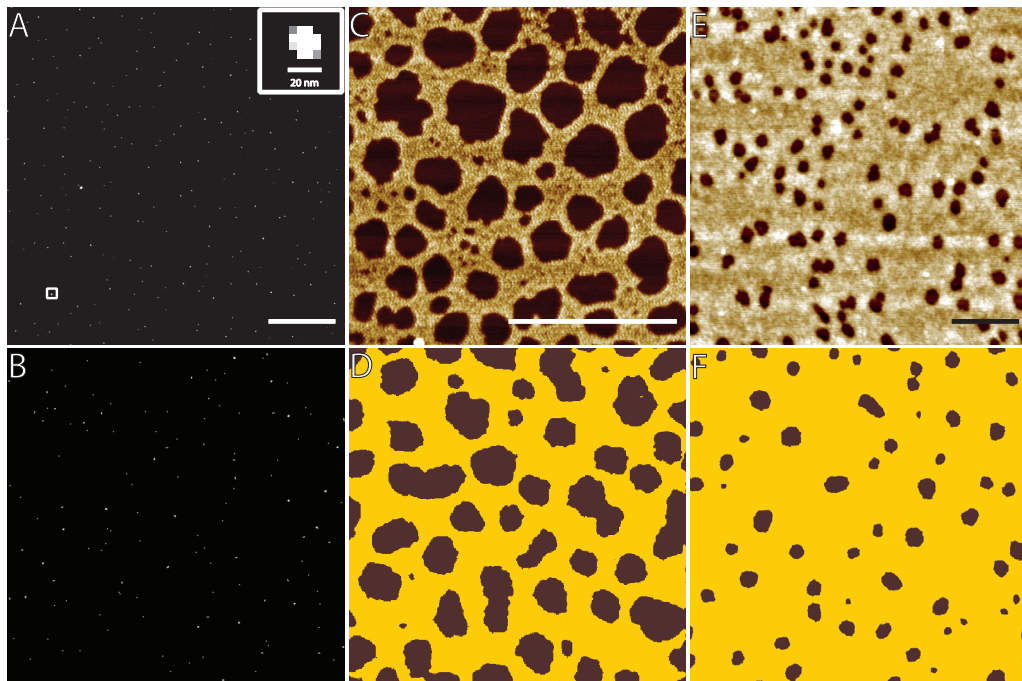


Figure 5.5 Experimental and computed morphologies of SCPNs obtained by evaporative self-assembly. The top row depicts experimental atomic force micrographs, with the scale bar at the bottom of each image indicating a length of 1 μm . The bottom row shows the corresponding 2D Monte Carlo simulations, with parameters $\varepsilon_{nl} = 1.5\varepsilon_{ll}$, $\varepsilon_{nn} = 2\varepsilon_{ll}$, $kT = 0.6\varepsilon_{ll}$ and $MR = 30$. (A) Single particles dispersed on the substrate. The inset shows a single nanoparticle with a dimension of approximately 20 nm. A simulated morphology is shown in (B) corresponding to a nanoparticle coverage ϕ of 0.17% on a 1024×1024 lattice and chemical potential $\mu = -2.25\varepsilon_{ll}$; (C) AFM image of an aggregated nanoparticle network with intermediate coverage and the corresponding simulated morphology (D) with $\phi = 60\%$ on a 410×410 lattice and $\mu = -2.5\varepsilon_{ll}$; (E) AFM image of a continuous film of polymeric nanoparticles containing small holes, and the corresponding simulated morphology (F) with $\phi = 90\%$ on a 1024×1024 lattice and $\mu = -2.75\varepsilon_{ll}$.

However, the fact that the centroids of nanoparticles or holes in experimental AFM images deviate from random statistics does not—in itself—prove that a spinodal mechanism is the driving force behind the formation of such patterns. Previous work has demonstrated that coalescence of holes resulting from heterogeneous evaporation could also account for similar deviations^{22,37–39}. To strengthen the conclusions, we compared the normalized Minkowski functionals as a function of the normalized grain size of the experimental and computed morphologies obtained by simulations in the spinodal regime, and we observed that they deviate similarly (Figures 5.6B, 5.7B and 5.8B).

Here we reveal that evaporative self-assembly of SCPNs proceeds *via* a spinodal dewetting instability resulting in a wealth of different morphologies that can be rationalized by coarse-grained lattice-gas simulations, showing the possibility to predict nanoscale patterns obtained from thinning nanoparticle-solvent films. This insight will facilitate the production of hierarchically constructed, catalytically active surfaces based on SCPNs, capable of high-order spatiotemporal dynamics. Moreover, since we observe self-assembly mechanisms for SCPNs similar to those for inorganic nanoparticles, it is interesting to extend the scope of evaporative self-assembly to biomimetic and biological nanoparticles such as catalytically active SCPNs in water⁴⁰ and proteins.

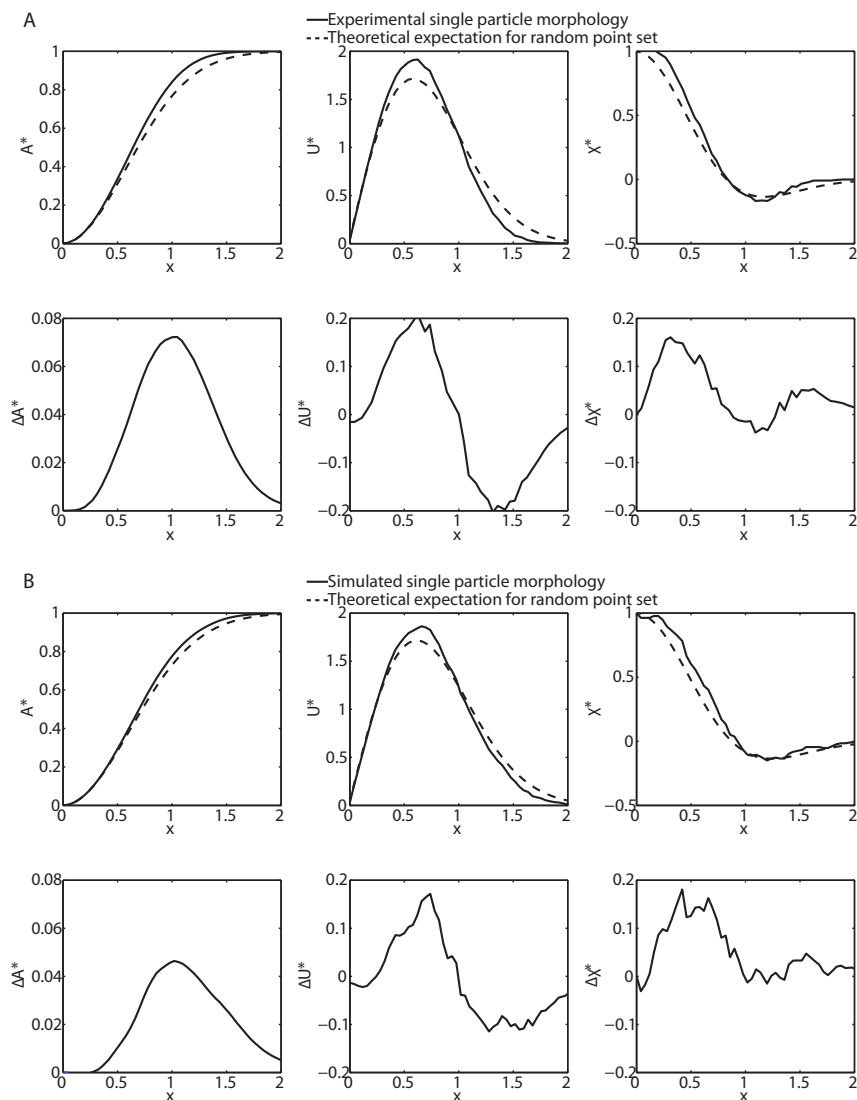


Figure 5.6 Minkowski grain growth analysis for the single particle morphologies as a function of the normalized distance x . (A) (top row) Minkowski grain growth analysis performed on AFM image of single particle morphology along with the theoretical expectation for a random point set with the same germ density; (bottom row) Difference of the analysis in the top row of A; (B) (top row) Minkowski grain growth analysis performed on simulated image of the single particle morphology along with the theoretical expectation for a random point set with the same germ density; (bottom row) Difference of the analysis in the top row of B.

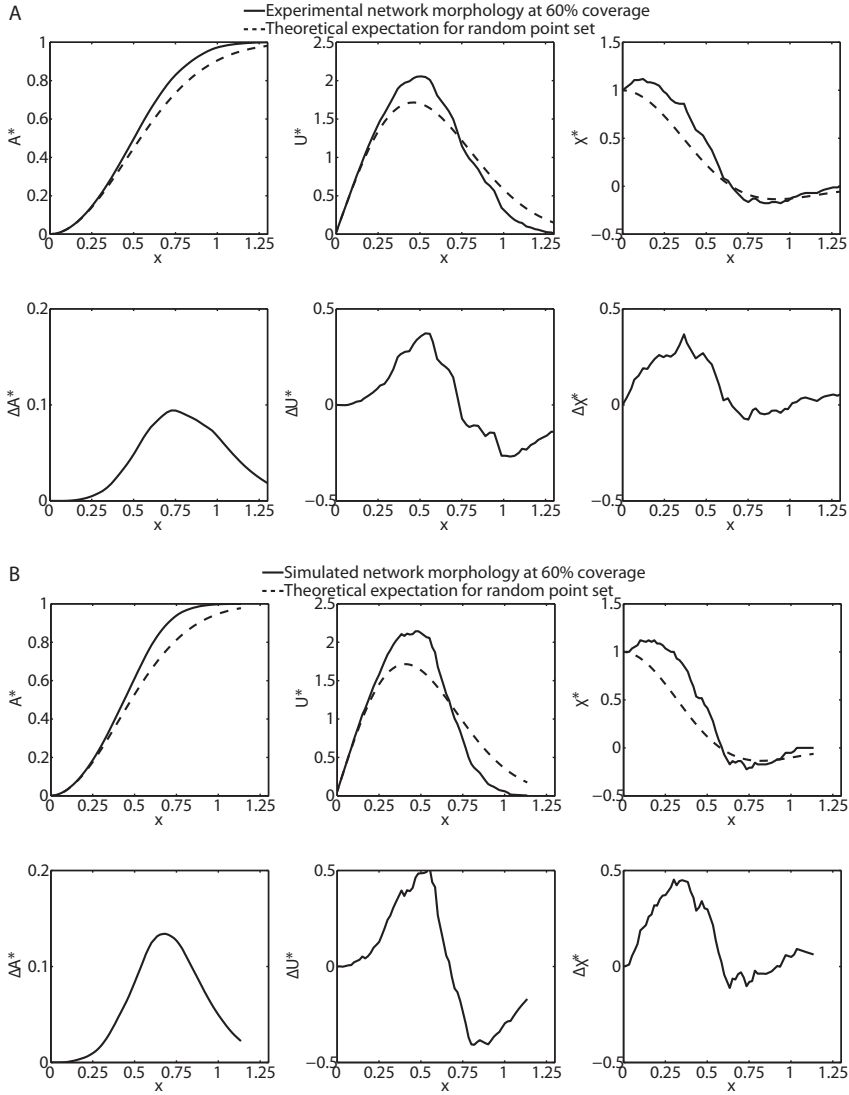


Figure 5.7 Minkowski grain growth analysis for the network morphologies at 60% coverage as a function of the normalized distance x . (A) (top row) Minkowski grain growth analysis performed on AFM image of the network morphology along with the theoretical expectation for a random point set with the same germ density; (bottom row) Difference of the analysis in the top row of A; (B) (top row) Minkowski grain growth analysis performed on simulated image of the network morphology along with the theoretical expectation for a random point set with the same germ density; (bottom row) Difference of the analysis in the top row of B.

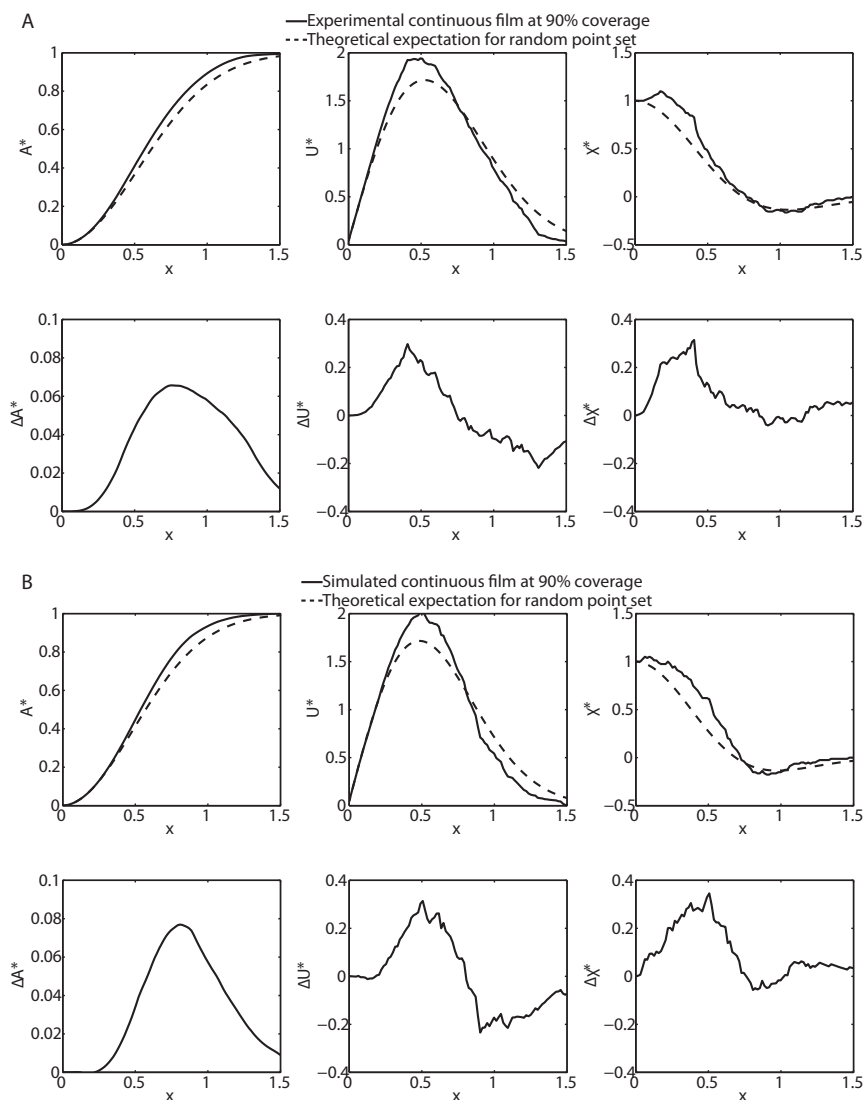


Figure 5.8 Minkowski grain growth analysis for the network morphologies at 90% coverage as a function of the normalized distance x . (A) (top row) Minkowski grain growth analysis performed on AFM image of the film morphology along with the theoretical expectation for a random point set with the same germ density; (bottom row) Difference of the analysis in the top row of A; (B) (top row) Minkowski grain growth analysis performed on simulated image of the film morphology along with the theoretical expectation for a random point set with the same germ density; (bottom row) Difference of the analysis in the top row of B.

5.4 REFERENCES

1. M. A. C. Stuart, W. T. S. Huck, J. Genzer, M. Müller, C. Ober, M. Stamm, G. B. Sukhorukov, I. Szleifer, V. V. Tsukruk, M. Urban, F. Winnik, S. Zauscher, I. Luzinov and S. Minko (2010), Emerging applications of stimuli-responsive polymer materials. *Nat. Mater.* **9**, 101–113.
2. J. M. Spruell and C. J. Hawker (2010), Triggered structural and property changes in polymeric nanomaterials. *Chem. Sci.* **2**, 18–26.
3. E. A. Appel, J. Dyson, J. del Barrio, Z. Walsh and O. A. Scherman (2012), Formation of single-chain polymer nanoparticles in water through host–guest interactions. *Angew. Chem., Int. Ed.* **51**, 4185–4189.
4. B. V. K. J. Schmidt, N. Fechner, J. Falkenhagen and J.-F. Lutz (2011), Controlled folding of synthetic polymer chains through the formation of positionable covalent bridges. *Nat. Chem.* **3**, 234–238.
5. M. Ouchi, N. Badi, J.-F. Lutz and M. Sawamoto (2011), Single-chain technology using discrete synthetic macromolecules. *Nat. Chem.* **3**, 917–924.
6. E. Harth, B. V. Horn, V. Y. Lee, D. S. Germack, C. P. Gonzales, R. D. Miller and C. J. Hawker (2002), A facile approach to architecturally defined nanoparticles via intramolecular chain collapse. *J. Am. Chem. Soc.* **124**, 8653–8660.
7. A. E. Cherian, F. C. Sun, S. S. Sheiko and G. W. Coates (2007), Formation of nanoparticles by intramolecular cross-linking: following the reaction progress of single polymer chains by atomic force microscopy. *J. Am. Chem. Soc.* **129**, 11350–11351.
8. J. B. Beck, K. L. Killops, T. Kang, K. Sivanandan, A. Bayles, M. E. Mackay, K. L. Wooley and C. J. Hawker (2009), Facile preparation of nanoparticles by intramolecular cross-linking of isocyanate functionalized copolymers. *Macromolecules* **42**, 5629–5635.
9. B. S. Murray and D. A. Fulton (2011), Dynamic covalent single-chain polymer nanoparticles. *Macromolecules* **44**, 7242–7252.
10. M. Seo, B. J. Beck, J. M. J. Paulusse, C. J. Hawker and S. Y. Kim (2008), Polymeric nanoparticles via noncovalent cross-linking of linear chains. *Macromolecules* **41**, 6413–6418.
11. E. J. Foster, E. B. Berda and E. W. Meijer (2009), Metastable supramolecular polymer nanoparticles via intramolecular collapse of single polymer chains. *J. Am. Chem. Soc.* **131**, 6964–6966.
12. E. B. Berda, E. J. Foster and E. W. Meijer (2010), Toward controlling folding in synthetic polymers: fabricating and characterizing supramolecular single-chain nanoparticles. *Macromolecules* **43**, 1430–1437.
13. E. J. Foster, E. B. Berda and E. W. Meijer (2011), Tuning the size of supramolecular single-chain polymer nanoparticles. *J. Polym. Sci., Part A: Polym. Chem.* **49**, 118–126.
14. P. Englebienne, P. A. J. Hilbers, E. W. Meijer, T. F. A. de Greef and A. J. Markvoort (2012), Directional interactions in semiflexible single-chain polymer folding. *Soft Matter* **8**, 7610–7616.
15. T. Terashima, T. Mes, T. F. A. de Greef, M. A. J. Gillissen, P. Besenius, A. R. A. Palmans and E. W. Meijer (2011), Single-chain folding of polymers for catalytic systems in water. *J. Am. Chem. Soc.* **133**, 4742–4745.
16. E. Huerta, P. J. M. Stals, E. W. Meijer and A. R. A. Palmans (2013), Consequences of folding a water-soluble polymer around an organocatalyst. *Angew. Chem., Int. Ed.* **52**, 2906–2910.
17. M. A. J. Gillissen, I. K. Voets, E. W. Meijer and A. R. A. Palmans (2012), Single chain polymeric nanoparticles as compartmentalized sensors for metal ions. *Polym. Chem.* **3**, 3166–3174.
18. P. A. Bertin, J. M. Gibbs, C. K.-F. Shen, C. S. Thaxton, W. A. Russin, C. A. Mirkin and S. T. Nguyen (2006), Multifunctional polymeric nanoparticles from diverse bioactive agents. *J. Am. Chem. Soc.* **128**, 4168–4169.
19. R. Seemann, S. Herminghaus and K. Jacobs (2001), Dewetting patterns and molecular forces: a reconciliation. *Phys. Rev. Lett.* **86**, 5534–5537.
20. R. Seemann, S. Herminghaus and K. Jacobs (2001), Gaining control of pattern formation of dewetting liquid films. *J. Phys.: Condens. Matter* **13**, 4925.
21. E. Rabani, D. R. Reichman, P. L. Geissler and L. E. Brus (2003), Drying-mediated self-assembly of nanoparticles. *Nature* **426**, 271–274.
22. C. P. Martin, M. O. Blunt and P. Moriarty (2004), Nanoparticle networks on silicon: self-organized or disorganized? *Nano Lett.* **4**, 2389–2392.
23. G. Yosef and E. Rabani (2006), Self-assembly of nanoparticles into rings: a lattice-gas model. *J. Phys. Chem. B* **110**, 20965–20972.
24. C. P. Martin, M. O. Blunt, E. Pauliac-Vaujour, A. Stannard, P. Moriarty, I. Vancea and U. Thiele (2007), Controlling pattern formation in nanoparticle assemblies via directed solvent dewetting. *Phys. Rev. Lett.* **99**, 116103.
25. E. Pauliac-Vaujour, A. Stannard, C. P. Martin, M. O. Blunt, I. Nottingher, P. J. Moriarty, I. Vancea and U. Thiele (2008), Fingering instabilities in dewetting nanofluids. *Phys. Rev. Lett.* **100**, 176102.
26. A. Stannard, C. P. Martin, E. Pauliac-Vaujour, P. Moriarty and U. Thiele (2008), Dual-scale pattern formation in nanoparticle assemblies. *J. Phys. Chem. C* **112**, 15195–15203.
27. I. Vancea, U. Thiele, E. Pauliac-Vaujour, A. Stannard, C. P. Martin, M. O. Blunt and P. J. Moriarty (2008), Front instabilities in evaporatively dewetting nanofluids. *Phys. Rev. E* **78**, 041601.

28. U. Thiele, I. Vancea, A. J. Archer, M. J. Robbins, L. Frastia, A. Stannard, E. Pauliac-Vaujour, C. P. Martin, M. O. Blunt and P. J. Moriarty (2009), Modelling approaches to the dewetting of evaporating thin films of nanoparticle suspensions. *J. Phys.: Condens. Matter* **21**, 264016.
29. A. Stannard (2011), Dewetting-mediated pattern formation in nanoparticle assemblies. *J. Phys.: Condens. Matter* **23**, 083001.
30. F. J. Giessibl (2003), Advances in atomic force microscopy. *Rev. Mod. Phys.* **75**, 949–983.
31. P. J. M. Stals, M. A. J. Gillissen, R. Nicolaÿ, A. R. A. Palmans and E. W. Meijer (2013), The balance between intramolecular hydrogen bonding, polymer solubility and rigidity in single-chain polymeric nanoparticles. *Polym. Chem.* **4**, 2584–2597.
32. R. Baierlein (2001), The elusive chemical potential. *Am. J. Phys.* **69**, 423–434.
33. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller (1953), Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21**, 1087–1092.
34. N. Otsu (1979), A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber.* **9**, 62–66.
35. R. C. Gonzalez, R. E. Woods and S. L. Eddins (2009), *Digital Image Processing Using MATLAB* (2nd ed.). Gatesmark Publishing.
36. K. Michielsen and H. De Raedt (2001), Integral-geometry morphological image analysis. *Phys. Rep.* **347**, 461–538.
37. K. Jacobs, S. Herminghaus and K. R. Mecke (1998), Thin liquid polymer films rupture via defects. *Langmuir* **14**, 965–969.
38. P. Moriarty, M. D. R. Taylor and M. Brust (2002), Nanostructured cellular networks. *Phys. Rev. Lett.* **89**, 248303.
39. F. Frehill, K. H. G. Schulte, C. P. Martin, L. Wang, S. Patel, J. A. Purton, J. G. Vos and P. Moriarty (2004), Iron wheels on silicon: wetting behavior and electronic structure of adsorbed organostannoxane clusters. *Langmuir* **20**, 6421–6429.
40. M. Artar, T. Terashima, M. Sawamoto, E. W. Meijer and A. R. A. Palmans (2014), Understanding the catalytic activity of single-chain polymeric nanoparticles in water. *J. Polym. Sci., Part A: Polym. Chem.* **52**, 12–20.

Epilogue

General Discussion, Ethics and Outlook

6.1 GENERAL DISCUSSION

In this Thesis, we have focused on the bottom-up engineering of synthetic biochemical networks using *in silico* strategies. We have shown several studies where *a priori* design and *a posteriori* analysis have proven very useful in describing the behavior of such networks *in vitro*. These *in silico* strategies will continue to play an important role in synthetic biological research, but as networks increase in complexity, the need arises to automatically construct network descriptions rather than deriving them by hand. We have made an initial step into this direction by automated generation of sets of ODEs for the PEN toolbox. Such efforts should be expanded to networks based on other platforms, *e.g.* those based on genelets, on enzymes and on transcription-translation machinery. Such models of complex biochemical systems usually contain many parameters and limited data sets. This implies that underlying parameters can only be estimated with very limited accuracy and therefore model verification is very limited. One way to tackle this downside is to apply an appropriate form of optimal experiment design, *i.e.* a method that selects experiments or measurements that optimally reduce the variance or a related property of the model parameters¹. This optimal experiment design is fairly to achieve when frequently consulting with experimental collaborators. A final asset to keep in mind in this model-driven approach is the nature of the descriptions. When the concentrations of species become very small ($\sim 1\text{-}10^3$ molecules, *e.g.* in experiments involving compartmentalization into small droplets/reactors), stochastic descriptions of biochemical networks will become more appropriate than deterministic approaches. The behavior of such systems, however, will be more difficult to predict due to their stochastic nature.

The studies in this Thesis present fundamental research with the main aim of trying to elucidate the design principles and functionalities of living systems employing predefined biochemical platforms. Synthetic biology in general comes with a number of aspects that we wish to address in this Epilogue, *i.e.* additional applications and ethics. Firstly, we distinguish between *in vitro* synthetic biology, aimed at emulation, reconstitution and *de novo* engineering of existing or new biological functions in a cell-free environment, and *in vivo* synthetic biology, aimed at engineering genetic circuitry in cells to perform novel functions and beneficial tasks. Even though the definition of synthetic biology is quite clear, it is important to note that synthetic biology in itself is a translational field that is based on several foundational sciences, *e.g.* life sciences, biochemistry and molecular biology². It is an engineering discipline that is based on the principles of standardization, modularization and characterization, coupled to systematic design³.

In Chapter 1, we mentioned biosensors as one of the applications of synthetic biology. Biosensors are extremely versatile in terms of use and are possibly an area where applications of synthetic biology are realized on a relatively short time-scale⁴. Clinically, biosensors are important for the detection of pathogenic biofilms⁵ and parasites⁶. In society, biosensors can be applied to detect arsenic in drinking water⁷. In bio manufacture, biosensing is applicable to monitor bioprocess development *e.g.* in terms of temperature, pH, dissolved oxygen, and the availability of key nutrients, but also to detect the concentration of desired products. Next to monitoring manufactured products directly, tracking metabolites that correlate with production efficiency are more general due to applicability to multiple systems, and therefore are also important candidates for biosensing.

The production of high value chemicals, synthetic materials and pharmaceuticals by organisms provide a sustainable alternative to traditional manufacture. Metabolic reprogramming of the periwinkle plant enabled the resulting transgenic plant cell culture to produce several unnatural alkaloid compounds from precursors⁸. The production of synthetic versions of naturally occurring materials (e.g. spider silk⁹) and the synthesis of polymeric materials from renewable resources such as starch and sugar cane (polylactic acid¹⁰) are very notable examples of applying synthetic biology in materials science. In pharmaceuticals, an exemplary application is the rewiring of the genetic circuitry of yeast to increase microbial production of the antimalarial drug artemisinin to an industrial scale^{11,12}. The development of new drugs and personalized vaccines¹³ is also still ongoing. Such efforts are preceded with deduction of disease mechanisms and identification of drug-targets⁴.

Perhaps one of the most promising applications of synthetic biology is in the production of biofuels. For this purpose, the genetic circuitry in microorganisms and plants is optimized such that they inherit the ability to efficiently convert feedstock into biofuels. These include cellulosic bio-ethanol, biodiesel, algal biofuel and bio-based succinic acid. Numerous studies have been conducted and are still being performed and this suggests that the full potential of this application has not been exploited to this day¹⁴.

6.2 ETHICAL ASPECTS

For many people, the terms ‘genetically modified’ and ‘synthetic’ induce an association with something negative and unhealthy. Also, the idea of ‘playing God’ can evoke moral resistance in portions of the general public. Therefore, it is of utmost importance that the debate concerning these issues is carried out, both to inform the public and also to confine ethical boundaries for the conduction of research. In this section, we discuss benefits and potential harms of synthetic biology in an ethical context. We take a paper¹⁵ from the Hastings Center, an institution focused on bioethics in the US, as fundament for this section and indicate several points of ongoing discussion about the field of synthetic biology between ‘enthusiasts’ and ‘critics’ of the field. We complement this discussion with arguments and recommendations from a book¹⁶ published by the Dutch Rathenau Institute which is an independent organization that promotes the formation of political and public opinions on science and technology.

The benefits of conducting research in synthetic biology are twofold: advancing biological knowledge and producing new products. Richard Feynman supposedly wrote on a blackboard “What I cannot create, I do not understand”. By employing an engineering approach to manipulating life in the lab, scientists will gain insights into the answers to questions about the origin of life and how a collection of organic chemicals can become functional life. The second benefit of synthetic biology is practical; employ novel energy sources, synthesize biodegradable plastics and provide alternative routes for the manufacture of medicinal drugs.

Besides benefits, synthetic biology also comes with potential harms. The authors argue¹⁵ that potential harms can be divided into two categories: physical harms and non-physical harms. The former can be exemplified by the effect on public health and safety or the environment when a genetically modified organism would escape the confinements of the research setting, or by employing synthetic biology for the production of new types of biological weapons¹⁷. These harms can in turn be known *a priori*, or be unknown even to the scientist. Needless to say, an *a posteriori*

assessment of physical harms is undesirable. The latter category is more difficult to articulate and is associated with harm to moral standards (what is right, what is fair). For instance, when commercializing and patenting inventions, questions arise about who controls these and has access to them. Concerns about the prospect of using synthetic biological technologies to enhance human traits and the creation of novel organisms are ethical concerns as well as related to our safety.

The aim of preventing physical harms is good and is generally a common goal of researchers, enthusiasts and critics. This goal is also nicely defined as there is general consensus on what exactly physical harms embody. However, views on how this prevention should be implemented vary, one view being ‘self-regulation’. Namely, the community working on synthetic biology projects supposedly best knows how to prevent physical harm, and scientists have already proven capable of regulating their research in 1975 at the Asilomar conference on recombinant DNA molecules¹⁸. Critics of self-regulation point out that although members of the synthetic biology community seem to be aware of potential harms, they fail to pay much attention to the lessons learnt from previous and ongoing social science research. A solution to this problem would be public engagement, *i.e.* allow citizens to provide an upstream governance of science and technology. Public engagement could influence topics of research based on critical feedback on hypothetical future applications and establish a timely response to issues as they arise.

There is fewer consensus even about the nature of non-physical harms, hence it is difficult to give a good overview of ethical standpoints regarding prevention of non-physical harms. The authors¹⁵ state that there is more work to be done in identifying and addressing these harms, and formulating them from within both the enthusiast and critic point-of-view is imperative. Through empirical and conceptual social research, we need to gain a better understanding of what enthusiasts celebrate in synthetic biology and what causes concern for critics.

‘Playing God’ is a term used by critics to object to the reductionist approach employed in synthetic biology, as it may threaten the view that life is special and that all living things are connected in one way or another. After all, the argument is, life can be reduced to functionalized and compartmentalized DNA and is not worth more than organized matter¹⁶. The fear that follows from this argument is the loss of respect for living things as such. In other words, should we continue to manipulate organisms and construct life-like systems from nonliving matter, we are at risk that our respect for living creatures is undermined, ultimately including lack of respect for animals or even humans. Additionally, when we adopt the role of creators, the argument is that we overestimate ourselves in the type of beings that we are. Enthusiasts on the other hand, take the stand that the natural world is ours to mold, shape and change, including ourselves, our offspring and our environment. Without this interference, we would not have vaccines and antibiotics, clean drinking water and there would be no progress.

In conclusion, the ethical debate on synthetic biology (and any technology for that matter) can be very fruitful when people listen carefully to the concerns of others in a respectful manner, and when the concerns are clearly articulated. In such a setting, discussions can facilitate the selection of experiments, the communication of results and the synthesis of eventual products. This can be achieved *e.g.* by monitoring scientific and technological developments, by studying scientific and social effects and the risks and by involving social organizations, politicians and ecologists. Moreover, an international network where people, knowledge and skill are brought together as well

as by openly communicating results and pitfalls aids in keeping this discussion going and shape the path of where research in synthetic biology takes us as a society.

6.3 FUTURE PERSPECTIVES

In this section, we mention potential outcomes and benefits as a result of conducting research in synthetic biology and applying the field in industry. An obvious ultimate goal of *in vitro* synthetic biology is the synthesis of artificial life, or an autonomous functioning artificial cell. The construction of biochemical networks which are able to operate out-of-equilibrium in a sustained manner combined with minimization of retroactivity effects paves the way for the possibility of constructing ever larger biochemical synthetic networks. This includes the combination of synthetic networks based on different platforms that simultaneously perform their intended functionalities. Several efforts have been conducted in the functionalization of so-called protocells. Functionalization includes protocells being able to synthesize proteins¹⁹, replicate DNA²⁰, produce sugars and communicate with living cells²¹. Parallel research on the deduction of a minimal genome²² and other continued efforts in this branch will shed light on abiogenesis²³ and evolution. Such bottom-up approaches are best accompanied in close collaboration with physical chemists and biologists and require integration of mathematical modeling of complex reaction networks into the design process. The bottom-up construction of CRNs is a unique approach to understand, predict, and possibly control the behavior of complex molecular systems. Accompanied by parallel progress in top-down approaches driven by ‘omics’ data which attempt to cover genome-wide descriptions, together these approaches will lead to an increased understanding of the full functionality of living systems.

Synthetic biology can potentially facilitate or even shape the transition from an oil-based feedstock to a greener, more sustainable bio-based feedstock. As fossil sources will at some point in time be depleted, the necessity arises to look for alternative and potentially cheaper feedstocks. The focus of such efforts is currently best focused on the optimization of biosynthetic pathways involving compounds that are currently used as chemical feedstock, such as isobutyric acid, alkaloids and longer-chain alcohols, alkenes and fatty acids. This shift towards cheaper and perhaps more easily available raw materials, combined with ultimately cheaper operating costs, presents a continuous need for furtherance of synthetic biology. Analogously to information technology, airplanes and cars, which are readily available to us but in a sense never ‘finished’ as developments are still ongoing, synthetic bio-manufactured products will become readily available to the public but always under scrutiny for continued development and optimization.

6.4 REFERENCES

1. K. Chaloner and I. Verdinelli (1995), Bayesian experimental design: a review. *Stat. Sci.* **10**, 273–304.
2. R. Kitney and P. Freemont (2012), Synthetic biology – the state of play. *FEBS Lett.* **586**, 2029–2036.
3. C. A. Voigt (2006), Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **17**, 548–557.
4. A. S. Khalil and J. J. Collins (2010), Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367–379.
5. N. Saeidi, C. K. Wong, T.-M. Lo, H. X. Nguyen, H. Ling, S. S. J. Leong, C. L. Poh and M. W. Chang (2014), Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. *Mol. Syst. Biol.* **7**, 521–521.
6. K. E. Mach, R. Mohan, S. Patel, P. K. Wong, M. Hsieh and J. C. Liao (2015), Development of a biosensor-based rapid urine test for detection of urogenital schistosomiasis. *PLoS Negl. Trop. Dis.* **9**, e0003845.
7. K. de Mora, N. Joshi, B. L. Balint, F. B. Ward, A. Elfick and C. E. French (2011), A pH-based biosensor for detection of arsenic in drinking water. *Anal. Bioanal. Chem.* **400**, 1031–1039.

8. W. Runguphan and S. E. O'Connor (2009), Metabolic reprogramming of periwinkle plant culture. *Nat. Chem. Biol.* **5**, 151–153.
9. D. M. Widmaier, D. Tullman-Ercek, E. A. Mirsky, R. Hill, S. Govindarajan, J. Minshull and C. A. Voigt (2009), Engineering the salmonella type III secretion system to export spider silk monomers. *Mol. Syst. Biol.* **5**, 309.
10. Y. K. Jung, T. Y. Kim, S. J. Park and S. Y. Lee (2010), Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. *Biotechnol. Bioeng.* **105**, 161–171.
11. V. J. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman and J. D. Keasling (2003), Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802.
12. 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 (2006), Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943.
13. G. Kanter, J. Yang, A. Voloshin, S. Levy, J. R. Swartz and R. Levy (2007), Cell-free production of scFv fusion proteins: an efficient approach for personalized lymphoma vaccines. *Blood* **109**, 3393–3399.
14. C. A. Rabinovitch-Deere, J. W. K. Oliver, G. M. Rodriguez and S. Atsumi (2013), Synthetic biology and metabolic engineering approaches to produce biofuels. *Chem. Rev.* **113**, 4611–4632.
15. E. Parens, J. Johnston and J. Moses (2009), *Ethical Issues in Synthetic Biology*. The Hastings Center.
16. H. de Vriend (2006), *Constructing life. Early social reflections on the emerging field of synthetic biology*. Rathenau Institute.
17. M. Schmidt (2006), Public will fear biological accidents, not just attacks. *Nature* **441**, 1048–1048.
18. P. Berg, D. Baltimore, S. Brenner, R. O. Roblin and M. F. Singer (1975), Summary statement of the Asilomar conference on recombinant DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 1981–1984.
19. V. Noireaux and A. Libchaber (2004), A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17669–17674.
20. S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco and J. W. Szostak (2008), Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**, 122–125.
21. P. M. Gardner, K. Winzer and B. G. Davis (2009), Sugar synthesis in a protocellular model leads to a cell signalling response in bacteria. *Nat. Chem.* **1**, 377–383.
22. J. I. Glass, N. Assad-Garcia, N. Alperovich, S. Yooseph, M. R. Lewis, M. Maruf, C. A. Hutchison, H. O. Smith and J. C. Venter (2006), Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 425–430.
23. P. L. Luisi (2006), *The emergence of life*. Cambridge University Press.

Appendix

Supplement to the Chapters

A.1 MODEL EQUATIONS, PARAMETER VALUES, SAMPLING RANGES AND NUPACK SETTINGS CORRESPONDING TO CHAPTER 2

Table A.1 Parameters that are kept constant (symbols, meanings and values) throughout the sampling procedures of the Oligator and the IFFL circuit.

| Oligator | | |
|---|--|---|
| Parameter | Description | Value |
| k_a^α | Forward rate constant of binding of oligomer α | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| k_a^β | Forward rate constant of binding of oligomer β | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| k_a^{inh} | Forward rate constant of binding of inhibitor inh | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| $k_{aTH}^{inh} = k_a^{inh}$ | Forward rate constant of binding of inhibitor inh with a toehold of 6 or 7 bases | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| $k_{aTH}^\alpha = k_a^{inh} K_d^{inhT1} / K_d^\alpha$ | Forward rate constant of binding of oligomer α with a toehold of 3 bases | - |
| $V_{\max, \text{pol}}$ | Maximum rate of polymerization | 1200 nM min^{-1} |
| $K_{M, \text{pol}}$ | Michaelis-Menten constant of DNA polymerase | 44 nM |
| $V_{\max, \text{polSD}}$ | Maximum rate of strand displacing polymerization (only for production of inhibitor) | 40 nM min^{-1} |
| $K_{M, \text{polSD}}$ | Michaelis-Menten constant of DNA polymerase exhibiting strand displacing activity (only for production of inhibitor) | 3.5 nM |
| $V_{\max, \text{exoN}}$ | Maximum rate of exonuclease activity | 300 nM min^{-1} |
| $K_{M, \text{exoNlong}}$ | Michaelis-Menten constant of exonuclease activity on long oligomers (inhibitors) | 150 nM |
| $K_{M, \text{exoNshort}}$ | Michaelis-Menten constant of exonuclease activity on short oligomers (activators/signals) | 440 nM |
| $V_{\max, \text{nick}}$ | Maximum rate of nickase activity | 80 nM min^{-1} |
| $K_{M, \text{nick}}$ | Michaelis-Menten constant of nickase | 30 nM |
| IFFL circuit | | |
| Parameter | Description | Value |
| k_a^U | Forward rate constant of binding of oligomer U | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| k_a^α | Forward rate constant of binding of oligomer α | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| k_a^Y | Forward rate constant of binding of oligomer Y | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| k_a^{inh} | Forward rate constant of binding of inhibitor inh | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| $k_{aTH}^{inh} = k_a^{inh}$ | Forward rate constant of binding of inhibitor inh with a toehold of 6 or 7 bases | $0.06 \text{ nM}^{-1} \text{ min}^{-1}$ |
| $k_{aTH}^\alpha = k_a^{inh} K_d^{inhT3} / K_d^\alpha$ | Forward rate constant of binding of oligomer α with a toehold of 3 bases | - |
| $k_{aTH}^Y = k_a^{inh} K_d^{inhT3} / K_d^Y$ | Forward rate constant of binding of oligomer Y with a toehold of 3 bases | - |
| $V_{\max, \text{pol}}$ | Maximum rate of polymerization | 1200 nM min^{-1} |
| $K_{M, \text{pol}}$ | Michaelis-Menten constant of DNA polymerase | 44 nM |
| $V_{\max, \text{polU}}$ | Maximum rate of polymerization of protected primer U | 144 nM min^{-1} |
| $K_{M, \text{polU}}$ | Michaelis-Menten constant of DNA polymerase activity on substrate $U \cdot T1$ | 54.4 nM |
| $V_{\max, \text{polSD}}$ | Maximum rate of strand displacing polymerization | 40 nM min^{-1} |

| | | |
|-------------------------------|--|--------------------------|
| | (only for production of inhibitor) | |
| $K_{M, \text{polSD}}$ | Michaelis-Menten constant of DNA polymerase exhibiting strand displacing activity (only for production of inhibitor) | 3.5 nM |
| $V_{\text{max}, \text{exoN}}$ | Maximum rate of exonuclease activity | 300 nM min ⁻¹ |
| $K_{M, \text{exoNlong}}$ | Michaelis-Menten constant of exonuclease activity on long oligomers (inhibitors) | 150 nM |
| $K_{M, \text{exoNshort}}$ | Michaelis-Menten constant of exonuclease activity on short oligomers (activators/signals) | 440 nM |
| $V_{\text{max}, \text{nick}}$ | Maximum rate of nickase activity | 80 nM min ⁻¹ |
| $K_{M, \text{nick}}$ | Michaelis-Menten constant of nickase | 30 nM |

Table A.2 Sampling ranges of the initial concentrations and the dissociation constants in the Oligator and the IFFL circuit.

| Oligator | | | IFFL circuit | | |
|--|----------------------------------|----------------------|--|----------------------------------|----------------------|
| Parameter | Minimum | Maximum | Parameter | Minimum | Maximum |
| $[\alpha]$ /nM | 10^{-1} | 10^1 | $[U]_0$ /nM | 10^{-1} | 10^2 |
| $[T_1]$ /nM | 10^0 | 10^2 | $[T_1]$ /nM | 10^{-1} | 10^2 |
| $[T_2]$ /nM | 10^0 | 10^2 | $[T_2]$ /nM | 10^{-1} | 10^2 |
| $[T_3]$ /nM | 10^0 | 10^2 | $[T_3]$ /nM | 10^{-1} | 10^2 |
| K_d^α /nM | 10^{-3} | 10^4 | K_d^U /nM | 10^{-3} | 10^4 |
| K_d^β /nM | 10^{-3} | 10^4 | K_d^α /nM | 10^{-3} | 10^4 |
| $K_d^{\text{inh}T1}$ /nM | 10^{-3} | 10^4 | K_d^β /nM | 10^{-3} | 10^4 |
| $K_d^{\text{inh}T3}$ /nM | $\frac{1}{5} K_d^{\text{inh}T1}$ | $K_d^{\text{inh}T1}$ | $K_d^{\text{inh}T3}$ /nM | 10^{-3} | 10^4 |
| k_{aTH}^α | - | - | $K_d^{\text{inh}T2}$ /nM | $\frac{1}{5} K_d^{\text{inh}T3}$ | $K_d^{\text{inh}T3}$ |
| $= k_a^{\text{inh}} \frac{K_d^{\text{inh}T3}}{K_d^\alpha}$ | | | | | |
| /nM ⁻¹ min ⁻¹ | | | k_{aTH}^α | - | - |
| k_{aTH}^β | - | - | $= k_a^{\text{inh}} \frac{K_d^{\text{inh}T3}}{K_d^\alpha}$ | | |
| $= k_a^{\text{inh}} \frac{K_d^{\text{inh}T3}}{K_d^\beta}$ | | | /nM ⁻¹ min ⁻¹ | | |
| /nM ⁻¹ min ⁻¹ | | | k_{aTH}^Y | - | - |
| | | | $= k_a^{\text{inh}} \frac{K_d^{\text{inh}T3}}{K_d^Y}$ | | |
| | | | /nM ⁻¹ min ⁻¹ | | |

Table A.3 Settings for NUPACK binary complexes. Calling the file from a terminal is done as follows: complexes [-T temperature] [-material parameters] [-ordered] [-pairs] [-mfe] [-degenerate] [-dangles treatment] [-timeonly] [-quiet]. If the value is False in the Table, the input is not provided to complexes.

| Additional input | Value |
|------------------|-------|
| -T | 42.0 |
| -material | dna |
| -ordered | True |
| -pairs | False |
| -mfe | True |
| -degenerate | False |
| -dangles | some |
| -timeonly | False |
| -quiet | False |
| -sodium | 0.051 |
| -magnesium | 0.006 |

```
IFFL_MMLiterature
```

```
Reactions
```

```
U -> U + a, T1, MMpol, U; a; g; Y, MMnick, T1; T2; T3, ...
```

```
a -> a + g, T2, MMpol, U; a; g; Y, MMnick, T1; T2; T3, ...
```

```
a -> a + Y, T3, MMpol, U; a; g; Y, MMnick, T1; T2; T3, ...
```

```
g -| a + Y, T3, ...
```

```
a -X, MM, a; g; Y, ...
```

```
g -X, MM, a; g; Y, ...
```

```
Y -X, MM, a; g; Y, ...
```

```
Initials
```

```
U = 4
```

```
a = 0
```

```
g = 0
```

```
Y = 0
```

```
T1 = 10
```

```
T2 = 60
```

```
T3 = 180
```

Figure A.1 An input file for *modtoml*. This is the input file that was used to generate the IFFL from Chapter 2. The input file makes use of association constants rather than dissociation constants, *i.e.* $K_a = \frac{1}{K_d}$. The resulting ODE m-file has to be invoked by the sampling script and all scripts necessary for subsequent analyses.

```

Constants
ka_U = 0.06
Ka_U_T1 = 0.5
ka_a = 0.06
Ka_a_T1 = 5e-4
Vm_pol_a = 144
Km_pol_a = 54.4
Vm_polSD_a = 144
Km_polSD_a = 54.4
Vm_nick_T1 = 80
Km_nick_T1 = 30
Vm_exoN_a = 300
Km_exoN_a = 440
ka_a = 0.06
Ka_a_T2 = 5e-4
ka_g = 0.06
Ka_g_T2 = 500
Vm_pol_g = 1200
Km_pol_g = 44
Vm_polSD_g = 40
Km_polSD_g = 3.5
Vm_nick_T2 = 80
Km_nick_T2 = 30
Vm_exoN_g = 300
Km_exoN_g = 150
ka_a = 0.06
Ka_a_T3 = 5e-4
ka_Y = 0.06
Ka_Y_T3 = 0.6
Vm_pol_Y = 1200
Km_pol_Y = 44
Vm_polSD_Y = 1200
Km_polSD_Y = 44
Vm_nick_T3 = 80
Km_nick_T3 = 30
Vm_exoN_Y = 300
Km_exoN_Y = 440
ka_g = 0.06
Ka_g_T3 = 17
kaTH_g = 0.06

```

Figure A.1 (continued) An input file for *modtoml*. This is the input file that was used to generate the IFFL from Chapter 2. The input file makes use of association constants rather than dissociation constants, *i.e.* $K_a = \frac{1}{K_d}$. The resulting ODE m-file has to be invoked by the sampling script and all scripts necessary for subsequent analyses.

Model Equations of the Oligator

$$\begin{aligned}
\frac{d[T_1]}{dt} &= -k_a^\alpha[\alpha][T_1] + k_a^\alpha K_d^\alpha[\alpha \cdot T_1] - k_a^\alpha[T_1][\alpha] + k_a^\alpha K_d^\alpha[T_1 \cdot \alpha] \\
\frac{d[\alpha]}{dt} &= -k_a^\alpha[\alpha][T_1] + k_a^\alpha K_d^\alpha[\alpha \cdot T_1] - k_a^\alpha[\alpha \cdot T_1][\alpha] + k_a^\alpha K_d^\alpha[\alpha \cdot T_1 \cdot \alpha] \\
&\quad - k_a^\alpha[T_1][\alpha] + k_a^\alpha K_d^\alpha[T_1 \cdot \alpha] - k_a^\alpha[\alpha][T_1 \cdot \alpha] + k_a^\alpha K_d^\alpha[\alpha \cdot T_1 \cdot \alpha] \\
&\quad - k_a^\alpha[\alpha][T_2] + k_a^\alpha K_d^\alpha[\alpha \cdot T_2] - k_a^\alpha[\alpha][T_2 \cdot \beta] + k_a^\alpha K_d^\alpha[\alpha \cdot T_2 \cdot \beta] \\
&\quad - k_a^{inh} \frac{K_d^{inhT1}}{K_d^\alpha}[\alpha][inh \cdot T_1] + k_a^{inh}[\alpha \cdot T_1][inh] - k_a^{inh} \frac{K_d^{inhT1}}{K_d^\alpha}[inh \cdot T_1][\alpha] + k_a^{inh}[inh][T_1 \cdot \alpha] \\
&\quad - \frac{V_{\max, \text{exoNshort}}[\alpha]}{K_{M, \text{exoN}} \left(1 + \frac{[\alpha]}{K_{M, \text{exoN}}} + \frac{[\beta]}{K_{M, \text{exoN}}} + \frac{[inh]}{K_{M, \text{exoN}}} \right)} \\
&\quad + \frac{V_{\max, \text{pol}}[\alpha \cdot T_1 \cdot \alpha]}{K_{M, \text{pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{M, \text{pol}}} + \frac{[\beta \cdot T_3]}{K_{M, \text{pol}}} + \frac{[\beta \cdot T_3 \cdot inh]}{K_{M, \text{polSD}}} \right)} \\
\frac{d[\alpha \cdot T_1]}{dt} &= k_a^\alpha[\alpha][T_1] - k_a^\alpha K_d^\alpha[\alpha \cdot T_1] - k_a^\alpha[\alpha \cdot T_1][\alpha] + k_a^\alpha K_d^\alpha[\alpha \cdot T_1 \cdot \alpha] \\
&\quad - \frac{V_{\max, \text{pol}}[\alpha \cdot T_1]}{K_{M, \text{pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{M, \text{pol}}} + \frac{[\beta \cdot T_3]}{K_{M, \text{pol}}} + \frac{[\beta \cdot T_3 \cdot inh]}{K_{M, \text{polSD}}} \right)} \\
\frac{d[T_1 \cdot \alpha]}{dt} &= k_a^\alpha[T_1][\alpha] - k_a^\alpha K_d^\alpha[T_1 \cdot \alpha] - k_a^\alpha[\alpha][T_1 \cdot \alpha] + k_a^\alpha K_d^\alpha[\alpha \cdot T_1 \cdot \alpha] \\
\frac{d[\alpha \cdot T_1 \cdot \alpha]}{dt} &= k_a^\alpha[\alpha][T_1 \cdot \alpha] - k_a^\alpha K_d^\alpha[\alpha \cdot T_1 \cdot \alpha] + k_a^\alpha[\alpha \cdot T_1][\alpha] - k_a^\alpha K_d^\alpha[\alpha \cdot T_1 \cdot \alpha] + \\
&\quad - \frac{V_{\max, \text{pol}}[\alpha \cdot T_1 \cdot \alpha]}{K_{M, \text{pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{M, \text{pol}}} + \frac{[\beta \cdot T_3]}{K_{M, \text{pol}}} + \frac{[\beta \cdot T_3 \cdot inh]}{K_{M, \text{polSD}}} \right)} \\
&\quad + \frac{V_{\max, \text{nick}}[\alpha \alpha \cdot T_1]}{K_{M, \text{nick}} \left(1 + \frac{[\alpha \alpha \cdot T_1]}{K_{M, \text{nick}}} + \frac{[\alpha \beta \cdot T_2]}{K_{M, \text{nick}}} + \frac{[\beta inh \cdot T_3]}{K_{M, \text{nick}}} \right)}
\end{aligned} \tag{A.1}$$

$$\begin{aligned}
\frac{d[\alpha\alpha \cdot T_1]}{dt} &= \frac{V_{\max, \text{pol}}[\alpha \cdot T_1]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
&+ \frac{V_{\max, \text{pol}}[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
&- \frac{V_{\max, \text{nick}}[\alpha\alpha \cdot T_1]}{K_{\text{M, nick}} \left(1 + \frac{[\alpha\alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha\beta \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\beta \text{inh} \cdot T_3]}{K_{\text{M, nick}}} \right)} \\
\frac{d[T_2]}{dt} &= -k_a^\alpha[\alpha][T_2] + k_a^\alpha K_d^\alpha[\alpha \cdot T_2] - k_a^\beta[T_2][\beta] + k_a^\beta K_d^\beta[T_2 \cdot \beta] \\
\frac{d[\beta]}{dt} &= -k_a^\beta[T_2][\beta] + k_a^\beta K_d^\beta[T_2 \cdot \beta] - k_a^\beta[\alpha \cdot T_2][\beta] + k_a^\beta K_d^\beta[\alpha \cdot T_2 \cdot \beta] \\
&- k_a^\beta[\beta][T_3] + k_a^\beta K_d^\beta[\beta \cdot T_3] - k_a^\beta[\beta][T_3 \cdot \text{inh}] + k_a^\beta K_d^\beta[\beta \cdot T_3 \cdot \text{inh}] \\
&- \frac{V_{\max, \text{exoNshort}}[\beta]}{K_{\text{M, exoN}} \left(1 + \frac{[\alpha]}{K_{\text{M, exoN}}} + \frac{[\beta]}{K_{\text{M, exoN}}} + \frac{[\text{inh}]}{K_{\text{M, exoN}}} \right)} \\
&+ \frac{V_{\max, \text{pol}}[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
\frac{d[\alpha \cdot T_2]}{dt} &= k_a^\alpha[\alpha][T_2] - k_a^\alpha K_d^\alpha[\alpha \cdot T_2] - k_a^\beta[\alpha \cdot T_2][\beta] + k_a^\beta K_d^\beta[\alpha \cdot T_2 \cdot \beta] \\
&- \frac{V_{\max, \text{pol}}[\alpha \cdot T_2]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
\frac{d[T_2 \cdot \beta]}{dt} &= k_a^\beta[T_2][\beta] - k_a^\beta K_d^\beta[T_2 \cdot \beta] - k_a^\alpha[\alpha][T_2 \cdot \beta] + k_a^\alpha K_d^\alpha[\alpha \cdot T_2 \cdot \beta] \\
\frac{d[\alpha \cdot T_2 \cdot \beta]}{dt} &= k_a^\alpha[\alpha][T_2 \cdot \beta] - k_a^\alpha K_d^\alpha[\alpha \cdot T_2 \cdot \beta] + k_a^\beta[\alpha \cdot T_2][\beta] - k_a^\beta K_d^\beta[\alpha \cdot T_2 \cdot \beta] \\
&- \frac{V_{\max, \text{pol}}[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
&+ \frac{V_{\max, \text{nick}}[\alpha\beta \cdot T_2]}{K_{\text{M, nick}} \left(1 + \frac{[U\alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha\beta \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\beta \text{inh} \cdot T_3]}{K_{\text{M, nick}}} \right)}
\end{aligned} \tag{A.2}$$

$$\begin{aligned}
\frac{d[\alpha\beta \cdot T_2]}{dt} &= \frac{V_{\max, \text{pol}}[\alpha \cdot T_2]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
&+ \frac{V_{\max, \text{pol}}[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
&- \frac{V_{\max, \text{nick}}[\alpha\beta \cdot T_2]}{K_{\text{M, nick}} \left(1 + \frac{[U\alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha\beta \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\beta \text{inh} \cdot T_3]}{K_{\text{M, nick}}} \right)} \\
\frac{d[T_3]}{dt} &= -k_a^\beta [\beta][T_3] + k_a^\beta K_d^\beta [\beta \cdot T_3] - k_a^{\text{inh}} [T_3][\text{inh}] + k_a^{\text{inh}} K_d^{\text{inh}T3} [T_3 \cdot \text{inh}] \\
\frac{d[\text{inh}]}{dt} &= -k_a^{\text{inh}} [T_3][\text{inh}] + k_a^{\text{inh}} K_d^{\text{inh}T3} [T_3 \cdot \text{inh}] - k_a^{\text{inh}} [\beta \cdot T_3][\text{inh}] + k_a^{\text{inh}} K_d^{\text{inh}T3} [\beta \cdot T_3 \cdot \text{inh}] \\
&+ k_a^{\text{inh}} \frac{K_d^{\text{inh}T1}}{K_d^\alpha} [\alpha][\text{inh} \cdot T_1] - k_a^{\text{inh}} [\alpha \cdot T_1][\text{inh}] + k_a^{\text{inh}} \frac{K_d^{\text{inh}T1}}{K_d^\alpha} [\text{inh} \cdot T_1][\alpha] - k_a^{\text{inh}} [\text{inh}][T_1 \cdot \alpha] \\
&- \frac{V_{\max, \text{exoNlong}}[\text{inh}]}{K_{\text{M, exoN}} \left(1 + \frac{[\alpha]}{K_{\text{M, exoN}}} + \frac{[\beta]}{K_{\text{M, exoN}}} + \frac{[\text{inh}]}{K_{\text{M, exoN}}} \right)} \\
&+ \frac{V_{\max, \text{polSD}}[\beta \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
\frac{d[\beta \cdot T_3]}{dt} &= -k_a^\beta [\beta][T_3] + k_a^\beta K_d^\beta [\beta \cdot T_3] - k_a^{\text{inh}} [\beta \cdot T_3][\text{inh}] + k_a^{\text{inh}} K_d^{\text{inh}T3} [\beta \cdot T_3 \cdot \text{inh}] \\
&- \frac{V_{\max, \text{pol}}[\beta \cdot T_2]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
\frac{d[T_3 \cdot \text{inh}]}{dt} &= k_a^{\text{inh}} [T_3][\text{inh}] - k_a^{\text{inh}} K_d^{\text{inh}T3} [T_3 \cdot \text{inh}] - k_a^\alpha [\alpha][T_3 \cdot \text{inh}] + k_a^\alpha K_d^\alpha [\alpha \cdot T_3 \cdot \text{inh}] \\
\frac{d[\beta \cdot T_3 \cdot \text{inh}]}{dt} &= k_a^\beta [\beta][T_3 \cdot \text{inh}] - k_a^\beta K_d^\beta [\beta \cdot T_3 \cdot \text{inh}] + k_a^{\text{inh}} [\beta \cdot T_3][\text{inh}] - k_a^{\text{inh}} K_d^{\text{inh}T3} [\beta \cdot T_3 \cdot \text{inh}] \\
&- \frac{V_{\max, \text{polSD}}[\beta \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot \text{inh}]}{K_{\text{M, polSD}}} \right)} \\
&+ \frac{V_{\max, \text{nick}}[\beta \text{inh} \cdot T_3]}{K_{\text{M, nick}} \left(1 + \frac{[\alpha\alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha\beta \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\beta \text{inh} \cdot T_3]}{K_{\text{M, nick}}} \right)}
\end{aligned} \tag{A.3}$$

$$\begin{aligned}
\frac{d[\beta inh \cdot T_3]}{dt} = & \frac{V_{\max, \text{pol}}[\beta \cdot T_2]}{K_{\text{M, pol}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot inh]}{K_{\text{M, polSD}}} \right)} \\
& + \frac{V_{\max, \text{polSD}}[\beta \cdot T_2 \cdot inh]}{K_{\text{M, polSD}} \left(1 + \frac{[\alpha \cdot T_1]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \beta]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\beta \cdot T_3 \cdot inh]}{K_{\text{M, polSD}}} \right)} \\
& - \frac{V_{\max, \text{nick}}[\beta inh \cdot T_3]}{K_{\text{M, nick}} \left(1 + \frac{[\alpha \alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha \beta \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\beta inh \cdot T_3]}{K_{\text{M, nick}}} \right)} \quad (\text{A.4})
\end{aligned}$$

$$\begin{aligned}
\frac{d[inh \cdot T_1]}{dt} = & k_a^{inh} [inh][T_1] - k_a^{inh} K_d^{inh T1} [inh \cdot T_1] \\
& - k_a^{inh} \frac{K_d^{inh T1}}{K_d^\alpha} [\alpha][inh \cdot T_1] + k_a^{inh} [\alpha \cdot T_1][inh] - k_a^{inh} \frac{K_d^{inh T1}}{K_d^\alpha} [inh \cdot T_1][\alpha] + k_a^{inh} [inh][T_1 \cdot \alpha]
\end{aligned}$$

Model Equations of the IFFL circuit

$$\begin{aligned}
\frac{d[U]}{dt} &= -k_a^U [U][T_1] + k_a^U K_d^U [U \cdot T_1] - k_a^U [U][T_1 \cdot \alpha] + k_a^U K_d^U [U \cdot T_1 \cdot \alpha] \\
\frac{d[T_1]}{dt} &= -k_a^U [U][T_1] + k_a^U K_d^U [U \cdot T_1] - k_a^\alpha [T_1][\alpha] + k_a^\alpha K_d^\alpha [T_1 \cdot \alpha] \\
\frac{d[\alpha]}{dt} &= -k_a^\alpha [T_1][\alpha] + k_a^\alpha K_d^\alpha [T_1 \cdot \alpha] - k_a^\alpha [U \cdot T_1][\alpha] + k_a^\alpha K_d^\alpha [U \cdot T_1 \cdot \alpha] \\
&\quad - k_a^\alpha [\alpha][T_2] + k_a^\alpha K_d^\alpha [\alpha \cdot T_2] - k_a^\alpha [\alpha][T_2 \cdot inh] + k_a^\alpha K_d^\alpha [\alpha \cdot T_2 \cdot inh] \\
&\quad - k_a^\alpha [\alpha][T_3] + k_a^\alpha K_d^\alpha [\alpha \cdot T_3] - k_a^\alpha [\alpha][T_3 \cdot Y] + k_a^\alpha K_d^\alpha [\alpha \cdot T_3 \cdot Y] \\
&\quad - k_a^{inh} \frac{K_d^{inhT3}}{K_d^\alpha} [\alpha][inh \cdot T_3] + k_a^{inh} [\alpha \cdot T_3][inh] - \frac{V_{\max, \text{exoNshort}} [\alpha]}{K_{M, \text{exoN}} \left(1 + \frac{[\alpha]}{K_{M, \text{exoN}}} + \frac{[inh]}{K_{M, \text{exoN}}} + \frac{[Y]}{K_{M, \text{exoN}}} \right)} \\
&\quad + \frac{V_{\max, \text{polU}} [U \cdot T_1 \cdot \alpha]}{K_{M, \text{polU}} \left(1 + \frac{[U \cdot T_1]}{K_{M, \text{polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{M, \text{polU}}} + \frac{[\alpha \cdot T_2]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{M, \text{polSD}}} + \frac{[\alpha \cdot T_3]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{M, \text{pol}}} \right)} \\
\frac{d[U \cdot T_1]}{dt} &= k_a^U [U][T_1] - k_a^U K_d^U [U \cdot T_1] - k_a^\alpha [U \cdot T_1][\alpha] + k_a^\alpha K_d^\alpha [U \cdot T_1 \cdot \alpha] \\
&\quad - \frac{V_{\max, \text{polU}} [U \cdot T_1]}{K_{M, \text{polU}} \left(1 + \frac{[U \cdot T_1]}{K_{M, \text{polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{M, \text{polU}}} + \frac{[\alpha \cdot T_2]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{M, \text{polSD}}} + \frac{[\alpha \cdot T_3]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{M, \text{pol}}} \right)} \\
\frac{d[T_1 \cdot \alpha]}{dt} &= k_a^\alpha [T_1][\alpha] - k_a^\alpha K_d^\alpha [T_1 \cdot \alpha] - k_a^U [U][T_1 \cdot \alpha] + k_a^U K_d^U [U \cdot T_1 \cdot \alpha] \\
\frac{d[U \cdot T_1 \cdot \alpha]}{dt} &= k_a^U [U][T_1 \cdot \alpha] - k_a^U K_d^U [U \cdot T_1 \cdot \alpha] + k_a^\alpha [U \cdot T_1][\alpha] - k_a^\alpha K_d^\alpha [U \cdot T_1 \cdot \alpha] \\
&\quad - \frac{V_{\max, \text{polU}} [U \cdot T_1 \cdot \alpha]}{K_{M, \text{polU}} \left(1 + \frac{[U \cdot T_1]}{K_{M, \text{polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{M, \text{polU}}} + \frac{[\alpha \cdot T_2]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{M, \text{polSD}}} + \frac{[\alpha \cdot T_3]}{K_{M, \text{pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{M, \text{pol}}} \right)} \\
&\quad + \frac{V_{\max, \text{nick}} [U \alpha \cdot T_1]}{K_{M, \text{nick}} \left(1 + \frac{[U \alpha \cdot T_1]}{K_{M, \text{nick}}} + \frac{[\alpha inh \cdot T_2]}{K_{M, \text{nick}}} + \frac{[\alpha Y \cdot T_3]}{K_{M, \text{nick}}} \right)}
\end{aligned} \tag{A.5}$$

$$\begin{aligned}
\frac{d[U\alpha \cdot T_1]}{dt} = & \frac{V_{\max, \text{polU}}[U \cdot T_1]}{K_{\text{M, pol}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, pol}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
& + \frac{V_{\max, \text{polU}}[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
& - \frac{V_{\max, \text{nick}}[U\alpha \cdot T_1]}{K_{\text{M, nick}} \left(1 + \frac{[U\alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha \text{inh} \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\alpha Y \cdot T_3]}{K_{\text{M, nick}}} \right)} \\
\frac{d[T_2]}{dt} = & -k_a^\alpha [\alpha][T_2] + k_a^\alpha K_d^\alpha [\alpha \cdot T_2] - k_a^{\text{inh}} [T_2][\text{inh}] + k_a^{\text{inh}} K_d^{\text{inh}T2} [T_2 \cdot \text{inh}] \\
\frac{d[\text{inh}]}{dt} = & -k_a^{\text{inh}} [T_2][\text{inh}] + k_a^{\text{inh}} K_d^{\text{inh}T2} [T_2 \cdot \text{inh}] - k_a^{\text{inh}} [\alpha \cdot T_2][\text{inh}] + k_a^{\text{inh}} K_d^{\text{inh}T2} [\alpha \cdot T_2 \cdot \text{inh}] \\
& - k_a^{\text{inh}} [\text{inh}][T_3] + k_a^{\text{inh}} K_d^{\text{inh}T3} [\gamma \cdot T_3] - k_a^{\text{inh}} [\text{inh}][\alpha \cdot T_3] + k_a^{\text{inh}} \frac{K_d^{\text{inh}T3}}{K_a} [\alpha][\text{inh} \cdot T_3] \\
& - k_a^{\text{inh}} [\text{inh}][T_3 \cdot Y] + k_a^{\text{inh}} \frac{K_d^{\text{inh}T3}}{K_d^\gamma} [\text{inh} \cdot T_3][Y] \\
& - \frac{V_{\max, \text{exoNlong}}[\text{inh}]}{K_{\text{M, exoN}} \left(1 + \frac{[\alpha]}{K_{\text{M, exoN}}} + \frac{[\text{inh}]}{K_{\text{M, exoN}}} + \frac{[Y]}{K_{\text{M, exoN}}} \right)} \\
& + \frac{V_{\max, \text{polSD}}[\alpha \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
\frac{d[\alpha \cdot T_2]}{dt} = & k_a^\alpha [\alpha][T_2] - k_a^\alpha K_d^\alpha [\alpha \cdot T_2] - k_a^{\text{inh}} [\alpha \cdot T_2][\text{inh}] + k_a^{\text{inh}} K_d^{\text{inh}T2} [\alpha \cdot T_2 \cdot \text{inh}] \\
& - \frac{V_{\max, \text{pol}}[\alpha \cdot T_2]}{K_{\text{M, pol}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
\frac{d[T_2 \cdot \text{inh}]}{dt} = & k_a^{\text{inh}} [T_2][\text{inh}] - k_a^{\text{inh}} K_d^{\text{inh}T2} [T_2 \cdot \text{inh}] - k_a^\alpha [\alpha][T_2 \cdot \text{inh}] + k_a^\alpha K_d^\alpha [\alpha \cdot T_2 \cdot \text{inh}] \\
\frac{d[\alpha \cdot T_2 \cdot \text{inh}]}{dt} = & k_a^\alpha [\alpha][T_2 \cdot \text{inh}] - k_a^\alpha K_d^\alpha [\alpha \cdot T_2 \cdot \text{inh}] + k_a^{\text{inh}} [\alpha \cdot T_2][\text{inh}] - k_a^{\text{inh}} K_d^{\text{inh}T2} [\alpha \cdot T_2 \cdot \text{inh}] \\
& - \frac{V_{\max, \text{polSD}}[\alpha \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot \text{inh}]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
& + \frac{V_{\max, \text{nick}}[\alpha \text{inh} \cdot T_2]}{K_{\text{M, nick}} \left(1 + \frac{[U\alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha \text{inh} \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\alpha Y \cdot T_3]}{K_{\text{M, nick}}} \right)} \tag{A.6}
\end{aligned}$$

$$\begin{aligned}
\frac{d[\alpha inh \cdot T_2]}{dt} = & \frac{V_{\max, \text{pol}}[\alpha \cdot T_2]}{K_{\text{M, pol}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
& + \frac{V_{\max, \text{polSD}}[\alpha \cdot T_2 \cdot inh]}{K_{\text{M, polSD}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
& - \frac{V_{\max, \text{nick}}[\alpha inh \cdot T_2]}{K_{\text{M, nick}} \left(1 + \frac{[U \alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha inh \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\alpha Y \cdot T_3]}{K_{\text{M, nick}}} \right)} \\
\frac{d[T_3]}{dt} = & -k_a^\alpha [\alpha][T_3] + k_a^\alpha K_d^\alpha [\alpha \cdot T_3] - k_a^Y [T_3][Y] + k_a^Y K_d^Y [T_3 \cdot Y] - k_a^{inh} [inh][T_3] + k_a^{inh} K_d^{inhT3} [inh \cdot T_3] \\
\frac{d[Y]}{dt} = & -k_a^Y [T_3][Y] + k_a^Y K_d^Y [T_3 \cdot Y] - k_a^Y [\alpha \cdot T_3][Y] + k_a^Y K_d^Y [\alpha \cdot T_3 \cdot Y] \\
& - k_a^{inh} \frac{K_d^{inhT3}}{K_d^Y} [\gamma \cdot T_3][Y] + k_a^{inh} [inh][T_3 \cdot Y] \\
& - \frac{V_{\max, \text{exoNshort}}[Y]}{K_{\text{M, exoN}} \left(1 + \frac{[\alpha]}{K_{\text{M, exoN}}} + \frac{[inh]}{K_{\text{M, exoN}}} + \frac{[Y]}{K_{\text{M, exoN}}} \right)} \\
& + \frac{V_{\max, \text{pol}}[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
\frac{d[\alpha \cdot T_3]}{dt} = & k_a^\alpha [\alpha][T_3] - k_a^\alpha K_d^\alpha [\alpha \cdot T_3] - k_a^Y [\alpha \cdot T_3][Y] + k_a^Y K_d^Y [\alpha \cdot T_3 \cdot Y] \\
& + k_a^{inh} \frac{K_d^{inhT3}}{K_d^\alpha} [\alpha][inh \cdot T_3] - k_a^{inh} [\alpha \cdot T_3][inh] \\
& - \frac{V_{\max, \text{pol}}[\alpha \cdot T_3]}{K_{\text{M, pol}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
\frac{d[T_3 \cdot Y]}{dt} = & k_a^Y [T_3][Y] - k_a^Y K_d^Y [T_3 \cdot Y] - k_a^\alpha [\alpha][T_3 \cdot Y] + k_a^\alpha K_d^\alpha [\alpha \cdot T_3 \cdot Y] \\
& + k_a^{inh} \frac{K_d^{inhT3}}{K_d^Y} [inh \cdot T_3][Y] - k_a^{inh} [inh][T_3 \cdot Y]
\end{aligned} \tag{A.7}$$

$$\begin{aligned}
\frac{d[\alpha \cdot T_3 \cdot Y]}{dt} &= k_a^\alpha [\alpha][T_3 \cdot Y] - k_a^\alpha K_d^\alpha [\alpha \cdot T_3 \cdot Y] + k_a^Y [\alpha \cdot T_3][Y] - k_a^Y K_d^Y [\alpha \cdot T_3 \cdot Y] \\
&\quad - \frac{V_{\max, \text{pol}} [\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
&\quad + \frac{V_{\max, \text{nick}} [\alpha Y \cdot T_3]}{K_{\text{M, nick}} \left(1 + \frac{[U \alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha inh \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\alpha Y \cdot T_3]}{K_{\text{M, nick}}} \right)} \\
\frac{d[\alpha Y \cdot T_3]}{dt} &= \frac{V_{\max, \text{pol}} [\alpha \cdot T_3]}{K_{\text{M, pol}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
&\quad + \frac{V_{\max, \text{pol}} [\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}} \left(1 + \frac{[U \cdot T_1]}{K_{\text{M, polU}}} + \frac{[U \cdot T_1 \cdot \alpha]}{K_{\text{M, polU}}} + \frac{[\alpha \cdot T_2]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_2 \cdot inh]}{K_{\text{M, polSD}}} + \frac{[\alpha \cdot T_3]}{K_{\text{M, pol}}} + \frac{[\alpha \cdot T_3 \cdot Y]}{K_{\text{M, pol}}} \right)} \\
&\quad - \frac{V_{\max, \text{nick}} [\alpha Y \cdot T_3]}{K_{\text{M, nick}} \left(1 + \frac{[U \alpha \cdot T_1]}{K_{\text{M, nick}}} + \frac{[\alpha inh \cdot T_2]}{K_{\text{M, nick}}} + \frac{[\alpha Y \cdot T_3]}{K_{\text{M, nick}}} \right)} \\
\frac{d[inh \cdot T_3]}{dt} &= k_a^{inh} [inh][T_3] - k_a^{inh} K_d^{inh T^3} [T_3 \cdot inh] - k_a^{inh} \frac{K_d^{inh T^3}}{K_d^\alpha} [\alpha][inh \cdot T_3] + k_a^{inh} [\alpha \cdot T_3][inh] \\
&\quad - k_a^{inh} \frac{K_d^{inh T^3}}{K_d^Y} [inh \cdot T_3][Y] + k_a^{inh} [inh][T_3 \cdot Y]
\end{aligned} \tag{A.8}$$

A.2 MODEL EQUATIONS AND DERIVATIONS CORRESPONDING TO CHAPTER 3

Expansion of the model of the IFFL circuit with explicit exonuclease activity

The exonuclease kinetics acting on the ssDNA in equations (A.5)-(A.8) are replaced by the corresponding complementary kinetics resulting from the additions below in equation (A.9).

$$\begin{aligned}
 \frac{d[exoN]}{dt} &= -k_{f,exoN}[exoN][U] + k_{f,exoN}K_{I,U}[U \cdot exoN] \\
 &\quad - k_{f,exoN}[exoN][T_1] + k_{f,exoN}K_{I,T1}[T_1 \cdot exoN] \\
 &\quad - k_{f,exoN}[exoN][T_2] + k_{f,exoN}K_{I,T2}[T_2 \cdot exoN] \\
 &\quad - k_{f,exoN}[exoN][T_3] + k_{f,exoN}K_{I,T3}[T_3 \cdot exoN] \\
 &\quad - k_{f,exoN}[exoN][\alpha] + k_{f,exoN}K_{M,exoN}[exoN \cdot \alpha] \\
 &\quad - k_{f,exoN}[exoN][inh] + k_{f,exoN}K_{M,exoN}[exoN \cdot inh] \\
 &\quad - k_{f,exoN}[exoN][Y] + k_{f,exoN}K_{M,exoN}[exoN \cdot Y] \\
 \frac{d[U \cdot exoN]}{dt} &= k_{f,exoN}[exoN][U] - k_{f,exoN}K_{I,U}[U \cdot exoN] \\
 \frac{d[T_1 \cdot exoN]}{dt} &= k_{f,exoN}[exoN][T_1] - k_{f,exoN}K_{I,T1}[T_1 \cdot exoN] \\
 \frac{d[T_2 \cdot exoN]}{dt} &= k_{f,exoN}[exoN][T_2] - k_{f,exoN}K_{I,T2}[T_2 \cdot exoN] \\
 \frac{d[T_3 \cdot exoN]}{dt} &= k_{f,exoN}[exoN][T_3] - k_{f,exoN}K_{I,T3}[T_3 \cdot exoN] \\
 \frac{d[exoN \cdot \alpha]}{dt} &= k_{f,exoN}[exoN][\alpha] - k_{f,exoN}K_{M,exoN}[exoN \cdot \alpha] \\
 \frac{d[exoN \cdot inh]}{dt} &= k_{f,exoN}[exoN][inh] - k_{f,exoN}K_{M,exoN}[exoN \cdot inh] \\
 \frac{d[exoN \cdot Y]}{dt} &= k_{f,exoN}[exoN][Y] - k_{f,exoN}K_{M,exoN}[exoN \cdot Y]
 \end{aligned} \tag{A.9}$$

Relative change in inverse degradation rate in TMII

Figure 3.7 depicts the relative change in the inverse degradation rate of the primers, which increases as a function of U , reflecting unproductive binding of the enzyme due to a decreased activity. The degradation rate of TMII is a function of the concentration of free input primer and is defined in equation (A.10).

$$D(U_{free}) = \frac{V_{max,exoN}}{K_{M,exoN} \left(1 + \frac{U_{free}}{K_{I,U}} \right)} \tag{A.10}$$

As the steady state value of the output strand and its peak value are both proportional to the inverse of the degradation rate (A.11)

$$\bar{Y}, Y_p \propto \frac{1}{D(U_{\text{free}})} \quad (\text{A.11})$$

the relative change in Y is proportional to the relative change in the inverse of the degradation rate. The inverse degradation rate is given by (A.12).

$$\frac{1}{D(U_{\text{free}})} = \frac{K_{M,\text{exoN}} \left(1 + \frac{U_{\text{free}}}{K_{I,U}}\right)}{V_{\text{max,exoN}}} = \frac{K_{M,\text{exoN}} K_{I,U} \left(1 + \frac{U_{\text{free}}}{K_{I,U}}\right)}{K_{I,U} V_{\text{max,exoN}}} = \frac{K_{M,\text{exoN}}}{K_{I,U} V_{\text{max,exoN}}} (K_{I,U} + U_{\text{free}}) \quad (\text{A.12})$$

Thus, the relative change in the inverse degradation rate relative to the input change is given in equation (A.13).

$$\frac{\left(\frac{\delta \left(\frac{1}{D(U_{\text{free}})} \right)}{\frac{1}{D(U_{\text{free}})}} \right)}{\frac{\delta U}{U}} = \frac{\frac{\delta U_{\text{free}}}{K_{I,U} + U_{\text{free}}}}{\frac{\delta U}{U}} = \frac{U}{K_{I,U} + U_{\text{free}}} \frac{\delta U_{\text{free}}}{\delta U} \quad (\text{A.13})$$

Constants of equation (3.17) in Chapter 3

$$\lambda = D^3 K_a^2 K_{\text{inh}} + 2D^2 K_a K_{\text{inh}} k_a T_1 + D k_a T_1 (K_a k_{\text{inh}} T_2 + K_{\text{inh}} k_a T_1) \quad (\text{A.14})$$

$$\bar{Y}_{\text{max}} = \frac{1}{\lambda} K_{\text{inh}} k_a T_1 k_Y T_3 (D K_a + k_a T_1) \quad (\text{A.15})$$

$$K = \frac{1}{\lambda} D K_a K_{\text{inh}} K_U k_a T_1 k_Y T_3 \quad (\text{A.16})$$

$$K' = \frac{D}{\lambda} (2D^2 K_a^2 K_{\text{inh}} K_U + 2D K_a K_{\text{inh}} K_U k_a T_1 + K_a K_U k_a T_2 k_{\text{inh}} T_2) \quad (\text{A.17})$$

$$K'' = \frac{1}{\lambda} D^3 K_a^2 K_{\text{inh}} K_U^2 \quad (\text{A.18})$$

A.3 DERIVATIONS CORRESPONDING TO CHAPTER 4

We here derive the system of ODEs that describes the dynamics of the network. The bimolecular autoactivation of trypsinogen and autocatalytic enzymatic conversion of trypsinogen to trypsin (*i.e.* 1 and 2 in Figure 4.2) results in three differential equations (A.19).

$$\begin{aligned}
\frac{d[Tg]}{dt} &= -k_{f,auto(Tg)}[Tg][Tg] - k_{f,auto(Tr)}[Tg][Tr] + k_{r,auto(Tg)}[Tg \cdot Tr] \\
\frac{d[Tr]}{dt} &= k_{f,auto(Tg)}[Tg][Tg] - k_{f,auto(Tr)}[Tg][Tr] + k_{r,auto(Tr)}[Tg \cdot Tr] + 2k_{cat,auto(Tr)}[TgTr] \quad (A.19) \\
\frac{d[Tg \cdot Tr]}{dt} &= k_{f,auto(Tr)}[Tg][Tr] - k_{f,auto(Tr)}[Tg \cdot Tr] - k_{cat,auto(Tr)}[Tg \cdot Tr]
\end{aligned}$$

The conversion of *proInh* to *intInh* (reaction 3 in Figure 4.2) can be described by three additional differential equations, resulting in the coupled ODEs system displayed in equation (A.20)

$$\begin{aligned}
\frac{d[Tg]}{dt} &= -k_{f,auto(Tg)}[Tg][Tg] - k_{f,auto(Tr)}[Tg][Tr] + k_{r,auto(Tr)}[Tg \cdot Tr] \\
\frac{d[Tr]}{dt} &= k_{f,auto(Tg)}[Tg][Tg] \\
&\quad - k_{f,auto(Tr)}[Tg][Tr] + k_{r,auto(Tr)}[Tg \cdot Tr] + 2k_{cat,auto(Tr)}[Tg \cdot Tr] \\
&\quad - k_{f,act}[Tr][proInh] + k_{r,act}[Tr \cdot proInh] + k_{cat,act}[Tr \cdot proInh] \\
\frac{d[Tg \cdot Tr]}{dt} &= k_{f,auto(Tr)}[Tg][Tr] - k_{r,auto(Tr)}[Tg \cdot Tr] - k_{cat,auto(Tr)}[Tg \cdot Tr] \quad (A.20) \\
\frac{d[proInh]}{dt} &= -k_{f,act}[Tr][proInh] + k_{r,act}[Tr \cdot proInh] \\
\frac{d[Tr \cdot proInh]}{dt} &= k_{f,act}[Tr][proInh] - k_{r,act}[Tr \cdot proInh] \\
&\quad - k_{cat,act}[Tr \cdot proInh] \\
\frac{d[intInh]}{dt} &= k_{cat,act}[Tr \cdot proInh]
\end{aligned}$$

Thus far, reactions have been modeled in which trypsin is the key enzyme. Next, inhibition of trypsin by active inhibitor is described. To that end, terms are added that describe formation of active inhibitor from intermediate inhibitor catalyzed by aminopeptidase (4, Figure 4.2) and irreversible inhibition of trypsin (5, Figure 4.2) resulting in equation (A.21).

$$\begin{aligned}
\frac{d[Tr]}{dt} &= k_{f,auto(Tg)}[Tg][Tg] \\
&\quad - k_{f,auto(Tr)}[Tg][Tr] + k_{r,auto(Tr)}[Tg \cdot Tr] + 2k_{cat,auto(Tr)}[Tg \cdot Tr] \\
&\quad - k_{f,act}[Tr][proInh] + k_{r,act}[Tr \cdot proInh] + k_{cat,act}[Tr \cdot proInh] \\
&\quad - k_{f,inh(I)}[Tr][I] \quad (A.21) \\
\frac{d[intInh]}{dt} &= k_{cat,act}[Tr \cdot proInh] - \frac{V_{max, delay}[intInh]}{K_{M, delay} + [intInh]} \\
\frac{d[Inh]}{dt} &= \frac{V_{max, delay}[intInh]}{K_{M, delay} + [intInh]} - k_{f,inh(Inh)}[Tr][Inh]
\end{aligned}$$

$$\begin{aligned}
\frac{d[Tg]}{dt} &= -k_{f,auto(Tg)}[Tg][Tg] - k_{f,auto(Tr)}[Tg][Tr] + k_{r,auto(Tr)}[Tg \cdot Tr] \\
\frac{d[Tr]}{dt} &= k_{f,auto(Tg)}[Tg][Tg] \\
&\quad - k_{f,auto(Tr)}[Tg][Tr] + k_{r,auto(Tr)}[Tg \cdot Tr] + 2k_{cat,auto(Tr)}[Tg \cdot Tr] \\
&\quad - k_{f,act}[Tr][proInh] + k_{r,act}[Tr \cdot proInh] + k_{cat,act}[Tr \cdot proInh] \\
&\quad - k_{f,inh(Inh)}[Tr][Inh] - k_{f,inh(intInh)}[Tr][intInh] - k_{f,inh(proInh)}[Tr][proInh] \\
\frac{d[Tg \cdot Tr]}{dt} &= k_{f,auto(Tr)}[Tg][Tr] - k_{r,auto(Tr)}[Tg \cdot Tr] - k_{cat,auto(Tr)}[Tg \cdot Tr] \\
\frac{d[proInh]}{dt} &= -k_{f,act}[Tr][proInh] + k_{r,act}[Tr \cdot proInh] \\
&\quad - k_{f,inh(proInh)}[Tr][proInh] - k_{f,\phi(proInh)}[proInh] \\
\frac{d[Tr \cdot proInh]}{dt} &= k_{f,act}[Tr][proInh] - k_{r,act}[Tr \cdot proInh] \\
&\quad - k_{cat,act}[Tr \cdot proInh] - k_{flow}[Tr \cdot proInh] \\
\frac{d[intInh]}{dt} &= k_{cat,act}[Tr \cdot proInh] \\
&\quad - k_{f,inh(\alpha_2 I)}[Tr][intInh] - k_{f,\phi(intInh)}[intInh] \\
\frac{d[Inh]}{dt} &= \frac{V_{max,delay}[intInh]}{K_{M,delay} + [intInh]} - k_{f,inh(Inh)}[Tr][Inh] - k_{f,\phi(Inh)}[Inh]
\end{aligned} \tag{A.22}$$

The system of differential equations ((A.20) and (A.21) combined), describes the regulations necessary for the short-range activation and long-range inhibition. However, also the inhibitory reactions of pro- and intermediate inhibitor (reaction 5, Figure 4.2), as well as aqueous non-enzymatic hydrolysis of *proInh*, *intInh* and *Inh* (reaction 6, Figure 4.2) need to be added resulting in equation (A.22).

Because the reactions are carried out in a flow reactor, the inflow of new material should also be incorporated. The inflow contains concentrations of trypsinogen $[Tg]_0$, trypsin $[Tr]_0$ and pro-inhibitor $[proInh]_0$ (reaction 7, Figure 4.2). The concentration of aminopeptidase is implicitly incorporated in the parameter $V_{max,delay}$ as $[Ap]_0$. If the reaction volume is V , and the flow rate is k_{flow} , the total system of differential equations for the synthetic chemical network is given in equation (4.1). A description of all the variables and parameters in this ODE model is given in Table 4.1.

SUMMARY

Modular Design and Analysis of Synthetic Biochemical Networks

Biological systems show self-organization and complex behaviors in both time and space, which is established by many interacting modules, connected in a hierarchical fashion. The scientific disciplines that focus on unravelling such braids of interactions within and between biochemical networks are called systems biology and synthetic biology. These disciplines differ in approach, namely the former attempts to infer the topology of intra- and intercellular regulatory circuits from large collections of data qualitatively. The latter focuses on (re)design and quantitative descriptions of relatively small, pre-programmed sets of molecular programs, both inside and outside of living cells. In this Thesis the focus lies on synthetic biology. Specifically, I show in this Thesis that for developments in synthetic biology, computer models that accurately describe synthetic reaction networks are an absolute necessity. We use such models to predict the behavior of novel biochemical networks, and also to understand, analyze and improve existing biochemical systems.

In **Chapter 1** we present an overview of the work that has been conducted in synthetic biology since the second half of the previous century. We thoroughly describe chemical reaction networks that mimic biological behavior, process information and function as model systems for providing critical insights into biological principles. Furthermore, we indicate a transition from using small, inorganic chemicals under harsh conditions to using biological macromolecules under much milder conditions.

In **Chapter 2** the focus is on computational design of DNA circuits operating far from equilibrium by means of enzymes. We describe the full process, from concept to the actual DNA molecules that need to be ordered. More specifically, we start with a concept that is translated to a mathematical model. Subsequently, the model is screened for certain temporal behavior, *i.e.* oscillations and adaptation. The resulting functional model is optimized in terms of robustness such that it is able to account for unexpected perturbations. All the information is then combined to deduce the exact base sequence of the DNA molecules.

Chapter 3 is a continuation of the analysis of the adaptive system designed in Chapter 2. New experimental findings that are incorporated in the computational model, has a very positive effect on the quality of the adaptive response. The complexity of the model is reduced to qualitatively obtain critical insight in the mechanism that is responsible for the improved result.

A biochemical network based entirely on enzymes and small organic molecules is described in **Chapter 4**. A computational model is used to examine the rates of certain step and contributes to the choice of the small molecules. The model is then characterized in terms of identifiability of parameters. These attributes are compared when we make use of experimental data showing oscillations only, and when we make use of experimental data that show oscillatory behavior, but also other types of temporal behavior.

The last study, described in **Chapter 5**, deals with a system of polymeric nanoparticles that shows spatial self-assembly while the solvent evaporates in a closed chamber. This process is simulated with a coarse-grained model. Moreover, there is statistical agreement between

experimental data and so-called spinodal-dominated dewetting of the solvent in the computational model, indicating nonequilibrium pattern formation.

In the **Epilogue** we emphasize the value of computational models in the presented studies and we indicate in what practical manner *in silico* techniques will continue to play a role within synthetic biology. Then we describe a number of current applications of the field and give an overview of current ethical aspects that are topics of discussion for this branch. The Thesis is concluded with a future perspective and potential novel applications of synthetic biology.

NEDERLANDSE SAMENVATTING

Modulair Ontwerp en Analyse van Synthetische Biochemische Netwerken

Biologische systemen vertonen zelforganisatie en complexe gedragingen zowel in de ruimte als in de tijd, wat bewerkstelligd wordt door vele hiërarchische modules van biochemische netwerken die interacties met elkaar vertonen. De wetenschappelijke disciplines die zich bezig houden met de ontrafeling van zulke interacties tussen en ook binnen biochemische netwerken zijn de systeembiologie en de synthetische biologie. Deze twee disciplines verschillen in aanpak. De systeembiologie richt zich vooral op de kwalitatieve beschrijving van de samenhang en het gedrag van grote systemen met behulp van grote hoeveelheden gegevens. De synthetische biologie houdt zich bezig met het (her)ontwerp en de kwantitatieve beschrijving van kleinere, voorgeprogrammeerde modules van biochemische reacties, zowel binnen de cel als daarbuiten in de reageerbuis. In dit proefschrift wordt dieper ingegaan op de synthetische biologie. In specifieke zin, laat ik in dit proefschrift zien dat voor de ontwikkelingen binnen de synthetische biologie, computermodellen die accurate beschrijvingen van de desbetreffende netwerken bewerkstelligen, onontbeerlijk zijn. Deze computermodellen worden enerzijds gebruikt om het gedrag van nieuwe biochemische netwerken te voorspellen, en anderzijds om reeds bestaande systemen beter te begrijpen, te analyseren en uiteindelijk te verbeteren.

In **Hoofdstuk 1** wordt een overzicht gepresenteerd van het werk dat sinds de tweede helft van de vorige eeuw op het gebied van synthetische biologie is gedaan. Chemische netwerken die biologisch gedrag imiteren, informatie kunnen verwerken en als modelsysteem dienen voor de verkrijging van een beter inzicht in de biologische principes worden uitvoerig beschreven. Bovendien wordt een transitie aangeduid van het gebruik van kleine anorganische chemicaliën onder extreme condities in de reageerbuis naar het gebruik van grotere biomoleculen onder mildere omstandigheden.

Hoofdstuk 2 beschrijft het computationele ontwerp van DNA circuits die uit thermodynamisch evenwicht worden gehouden door middel van enzymen. Het volledige proces—van concept tot de concrete biomoleculen die besteld moeten worden—wordt beschreven. Specifieker, er wordt begonnen met een concept dat vertaald wordt naar een wiskundig model. Het model wordt gescreend op de vertoning van complexe gedragingen in de tijd, in dit geval oscillaties en adaptatie. Dan wordt het werkende model geoptimaliseerd op robuustheid, zodat het beter bestand is tegen onverwachte perturbaties. Alle relevante informatie wordt gecombineerd en daaruit worden de precieze DNA sequenties afgeleid.

Hoofdstuk 3 is een verdieping van het systeem uit Hoofdstuk 2 dat adaptatie vertoont. Wanneer nieuwe experimentele bevindingen in het model geïncorporeerd worden, heeft dat een aanzienlijk positief effect op de kwaliteit van de adaptatie. De complexiteit van het model wordt gereduceerd om het effect kwalitatief te beschrijven en zo inzicht te verkrijgen in het mechanisme dat verantwoordelijk is voor het verbeterde resultaat.

Hoofdstuk 4 is de beschrijving van een biochemisch netwerk dat gebaseerd is op enzymen en kleine organische moleculen en oscillaties vertoont. Een computationeel model wordt gebruikt om sturing te geven aan de keuze van de molecule in het biochemische netwerk. Daarnaast wordt het model gekarakteriseerd in termen van de identificeerbaarheid en het effect van individuele

parameters. Deze attributen worden vergeleken wanneer enerzijds gebruik gemaakt wordt van slechts experimentele data die oscillaties laten zien, en anderzijds wanneer gebruik gemaakt wordt van experimentele data die zowel oscillaties als geen oscillaties laten zien.

De laatste studie, **Hoofdstuk 5**, focust op een systeem van polymere nanodeeltjes die ruimtelijke zelforganisatie vertonen terwijl het oplosmiddel verdampt in een afgesloten kamer. Dit proces wordt gesimuleerd met een coarse-grained model. Er is bovendien statistische overeenkomst tussen experimentele data en zgn. spinodaal-gedomineerde dewetting van het oplosmiddel in het model, wat duidt op patroonvorming uit thermodynamisch evenwicht.

In de **Epiloog** wordt de meerwaarde van het gebruik van computermodellen in de gepresenteerde studies bediscussieerd, en aangegeven hoe *in silico* technieken op een pragmatische manier een blijvende rol zullen spelen binnen synthetische biologie. Dan worden een aantal huidige applicaties van synthetische biologie genoemd en wordt een beschouwing gegeven van ethische aspecten die onderzoek binnen deze tak met zich mee brengt. We sluiten af met een speculatie over de toekomst en potentiële nieuwe applicaties van synthetische biologie.

CURRICULUM VITAE

Rik van Roekel was born on November 16th 1985 in Venlo, The Netherlands. After finishing his secondary education at Bisschoppelijk College Broekhin in Roermond, he went on to study Biomedical Engineering at Eindhoven University of Technology in Eindhoven. He graduated from the Bachelor's program *cum laude*, and continued in the field by doing a Master's program at the same university. During the Master's program, Rik did a research internship at the University of Birmingham in Birmingham, UK, in the group of Dr. F. Falciani, involving the inference of the genetic network underlying the ability of *E. coli* to adapt to temperature increase. His Master's thesis was carried out in the group of Prof. P. A. J. Hilbers under supervision of Dr. T. F. A. de Greef, and involved lattice-gas modeling in order to rationalize self-assembly of polymeric nanoparticles. He then had the opportunity to obtain a PhD in the field of synthetic biology in the same group, where he placed a strong emphasis on computational strategies. With the aid of various *in silico* approaches, he developed a computational methodology to rationally design cell-free synthetic network motifs capable of predefined high-order temporal dynamics. In the design process, he coupled the macroscopic level to the mesoscopic level in order to establish reliable descriptions of the motif behavior.



LIST OF PUBLICATIONS

International refereed journal publications

H. W. H. van Roekel, B. J. H. M. Rosier, L. H. H. Meijer, P. A. J. Hilbers, A. J. Markvoort, W. T. S. Huck and T. F. A. de Greef (2015), Programmable chemical reaction networks: emulating regulatory functions in living cells using a bottom-up approach, *Chemical Society Reviews* **44**, 7465–7483.

H. W. H. van Roekel, L. H. H. Meijer, S. Masroor, Z. C. Félix Garza, A. Estévez-Torres, Y. Rondelez, A. Zagaris, M. A. Peletier, P. A. J. Hilbers and T. F. A. de Greef (2015), Automated design of enzyme-driven DNA circuits. *ACS Synthetic Biology* **4**, 735–745.

S. N. Semenov, A. S. Y. Wong, R. M. van der Made, S. G. J. Postma, J. Groen, H. W. H. van Roekel, T. F. A. de Greef and W. T. S. Huck (2015), Rational design of functional and tunable oscillating enzymatic networks. *Nature Chemistry* **7**, 160–165.

H. W. H. van Roekel, P. J. M. Stals, M. A. J. Gillissen, P. A. J. Hilbers, A. J. Markvoort and T. F. A. de Greef (2013), Evaporative self-assembly of single-chain, polymeric nanoparticles. *Chemical Communications* **49**, 3122–3124.

DANKWOORD

Na vier jaar promotieonderzoek ligt er dus dit boekje, mijn proefschrift, waar veel tijd en energie is in gaan zitten en waar ik dan ook erg trots op ben. Hoewel het mijn proefschrift is, heb ik het zeker niet alleen gedaan. En dus is nu aan mij de taak om de mensen die op welke manier dan ook een rol hebben gespeeld in de totstandkoming van dit boekje nog eens expliciet te bedanken in dit meest gelezen deel van het proefschrift.

Ik begin bij jou, Peter, mijn promotor. Ik heb het voorrecht gehad om een eerste baan te hebben in jouw onderzoeksgroep. Met veel plezier en vrijheid heb ik kunnen werken aan fantastische onderwerpen binnen de synthetische biologie. Daarnaast ben je een baas met wie afspraken gemaakt kunnen worden, iets dat ik zeer waardeer. Ik dank je voor de geweldige werkomgeving, de informele sfeer binnen de groep, de discussiepunten en adviezen tijdens presentaties, je kritische blik op de inhoud en structuur van artikelen en het proefschrift en voor de interesse in dit werk waar je lovende woorden over hebt geuit.

Beste Tom, woorden schieten denk ik te kort hoe dankbaar ik ben dat ik onder jouw begeleiding een proefschrift heb mogen schrijven. Je bent enthousiast, daadkrachtig, pragmatisch, inspirerend en bovenal een lopende encyclopedie die voor elk probleem een literatuurbron wist te citeren die weer licht deed schijnen op het soms moeilijke werk. Ik heb heel veel van je geleerd en ik zal altijd met ontzettend veel plezier terug kijken op de in totaal zes jaar dat wij hebben samengewerkt. Ik denk dat we allebei enorm zijn gegroeid, zowel in persoonlijke als professionele zin.

For this next part I will switch to English. I am very grateful to the reading committee of this dissertation, Mikko, Wilhelm, Menno, Bert and Pieter Rein. Considering the ever busy agendas of full professors, I thank you all very much for thoroughly reading my dissertation and providing your enthusiastic words on the contents of it. Moreover, I thank you for your willingness to be my opponents during my defense ceremony.

Josien, bedankt dat je de voorzitter wil zijn van mijn promotiecommissie!

Wilhelm, ik noem jou nogmaals naast het feit dat je deel uitmaakt van de leescommissie. Dank je voor de inspirerende samenwerking die ik heb mogen ervaren met de mensen in jouw lab, Sergey, Albert, Martijn en Sjoerd in het bijzonder.

Bart, ik kijk met plezier terug op ons samenwerken tijdens het schrijven van mijn eerste paper en het schrijven van de review, welke respectievelijk hebben geleid tot hoofdstukken 5 en 1 van dit boekje. Ik heb bewondering voor hoe snel jij fouten kunt detecteren in stukken code, voor je prettige en duidelijke manier van communiceren en ik ben je dankbaar voor alle hulp. Je intelligentie, integriteit, kalmte en ook je goede gevoel voor humor zijn noemenswaardige eigenschappen die ik nooit zal vergeten.

Mark, jammer dat je op 10 december al een andere promotie moet bijwonen. Toch wil ik je graag bedanken voor je input en de prettige samenwerking. De wiskundige kijk op de materie gedurende de nuttige discussies heeft uiteindelijk een zeer belangrijke rol gespeeld bij de totstandkoming van dit proefschrift.

Beste Lenny, ik vond onze aanpak van model en experiment een gedenkwaardige combinatie om tot gezamenlijk succes te komen. Ik zal onze samenwerking blijven herinneren als heel productief en ook als zeer plezierig. Ik ben trots dat wij samen twee papers hebben gepubliceerd. Verder blik ik met een bijzonder brede glimlach terug op het reisje naar Taormina. Ik wens je heel veel succes met het afronden van je eigen proefschrift, maar met jouw instelling en vaardigheden twijfel ik er niet aan dat je dat gaat lukken.

Saeed, we shared numerous valuable discussions, lunch meetings, phone calls and e-mails over the last few years, some under more stressful situations than others. I am happy that we got to work together on the theoretical aspects of adaptation. I wish you the very best of luck in completing your own dissertation and pursuing your future endeavors.

Antonios, thank you for the wonderful collaboration which has proven immensely useful. I greatly admire how your theoretical expertise has led to a firm hypothesis that has shed new light on the complexity of adaptation. I am also very grateful for your useful advice on scientific writing.

Als onderdeel van mijn werkzaamheden heb ik het privilege gehad om een aantal goede studenten te mogen begeleiden. Sander en Zandra, jullie hebben tijdens jullie stages een substantieel deel van hoofdstuk 2 uitgewerkt, dank daarvoor. Het komt als geen verrassing dat jullie beiden zijn doorgegaan met een promotie, en ik ben ervan overtuigd dat dit jullie gaat lukken. Daniël, heel erg bedankt voor het opschonen van de spaghetti GUI en het schrijven van *modtoml*. Deze tools zijn onontbeerlijk geweest voor de succesvolle afronding van dit proefschrift. Ik wens je heel veel succes met je verdere afstuderen en met wat daarna komt. En Javiér, hartelijk dank voor de slimme en handige optimization tools die je gemaakt hebt. Gelukkig ben je terug bij de faculteit BMT, want met jouw skills en intelligentie ben je een ware aanwinst. Heel veel succes verder.

Huub, met name in het begin van mijn promotie hebben wij ook de nodige nuttige discussies gevoerd, hartelijk dank hiervoor.

Ik wil verder graag de leden en exleden van de CBio groep bedanken voor hun behulpzaamheid, hun bereidheid tot discussie, hun humor en hun gezelligheid. Joep en Joep, Christian, Fianne, Zandra, Anne, Huili, Ceylan (sorry for the Dutch), Yvonne, Silvia, Willem, Sander en Bram. Het uitwisselen van ideeën en het delen van de ups en downs van een promotie is voor mij heel nuttig en plezierig geweest. De PhD etentjes zullen me altijd bijblijven als speciaal en heel erg gezellig. Ook toveren de avondjes O'Sheas nog altijd een glimlach op mijn gezicht Joep! En Fianne, onze reguliere tripjes naar de Zwarte Doos en naar de stad moeten we er absoluut in houden.

Nu ik je toch noem Fianne, maak ik het voor de volledigheid maar even af door jou nog eens te noemen samen met Tom Smeets. Ik ben namelijk vereerd dat ik jullie mag aanstellen als mijn paranimfen. Dank jullie dat jullie dit voor mij willen doen. En dank jullie voor de leuke tijden die we hebben beleefd en ook nog zullen beleven!

Rina, zonder jou was het in 2015 niet meer gelukt! Ik dank je voor je moral support bij het moeizame proces van het uitzoeken van de promotiedatum, en ook voor alles wat je verder voor mij hebt betekend. Je bent altijd oprecht bereid om te helpen en mee te denken, en ik zal de tripjes naar jouw kantoor en de vakgroepuitjes absoluut gaan missen bij mijn volgende werk.

Ricky en Koen, hartelijk dank voor het brianstormen over en maken van de prachtige kافت van dit proefschrift! Koen, jij nog in het bijzonder bedankt voor alle ondersteuning, zowel technisch als sociaal. Ik heb bewondering voor de manier waarop je altijd oprechte interesse toont in mensen en hun werk, en waar nodig een helpende hand biedt.

Onze etentjes, spelavonden, kroegbezoekjes en natuurlijk de borrels op donderdag hebben voor de nodige afleiding gezorgd van het onderzoek. Jammer dat we dit allemaal uiteindelijk zonder Youri hebben moeten doen. Een welgemeend woord van dank gaat uit naar Alexander, Sander, Fianne, Jasper, Lucas en Stefanie.

Naast mijn Eindhovense vrienden, wil ik graag mijn vrienden aan het thuisfront bedanken voor al hun steun, hun liefde en aandacht, hun geduld, hun oprechtheid en hun vertrouwen in de afgelopen jaren. In het bijzonder geldt dit voor Mandy, Frank, Shirley, andere Frank, Sander, Tom, Elvira, Eline, Sanne en Rennae.

Barbara en Wouter, lieve zus en broer, bedankt voor alle mooie tijden, jullie adviezen, meedenken en de bijdrage aan mijn karaktergroei, en gewoon voor het zijn van mijn zus en broer. Sander en Cindy, dank dat jullie ook een belangrijke rol in mijn leven vervullen, en natuurlijk geldt dat ook voor mijn lieve nichtjes Juul, Esmée, Rowena, Bibi en Demy.

Het laatste stuk van dit dankwoord wijd ik aan jullie, lieve mam en pap. Ik ben jullie eeuwig dankbaar voor alle steun, de goede en welgemeende adviezen, het vertrouwen en alle onvoorwaardelijke hulp. De bloedband is voor mij heilig, familie is het belangrijkste dat er is. Zonder jullie was dit allemaal nooit gelukt. Ik hou van jullie.

Allemaal ontzettend bedankt!



