



HAL
open science

Functional analysis of artificial DNA reaction network

Alexandre Baccouche

► **To cite this version:**

Alexandre Baccouche. Functional analysis of artificial DNA reaction network. Organic chemistry. Université Sorbonne Paris Cité, 2015. English. NNT : 2015USPCB135 . tel-01578138

HAL Id: tel-01578138

<https://theses.hal.science/tel-01578138>

Submitted on 28 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ PARIS DESCARTES

École doctorale Médicament - Toxicologie - Chimie - Imageries
(EDMTCI 563)

LIMMS/CNRS IIS (UMI2820) Applied Microfluidic Systems Laboratory

Functional analysis of artificial DNA reaction network

Par Alexandre Baccouche

Thèse de doctorat de Chimie

Dirigée par le Dr Yannick Rondelez

Présentée et soutenue publiquement le 18/12/2015

Devant un jury composé de :

Dr Zoher GUEROUI	rapporteur
Dr Tom DE GREEF	rapporteur
Pr Olivia REINAUD	examinatrice
Pr Dominique COLLARD	examineur
Dr André ESTEVEZ-TORRES	examineur
Dr Yannick RONDELEZ	directeur de thèse



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Résumé : La gestion et transmission d'information au sein d'organismes vivants implique la production et le trafic de molécules *via* des voies de signalisation structurées en réseaux de réactions chimiques. Ces derniers varient selon leur forme, taille ainsi que la nature des molécules mises en jeu. Parmi eux, les réseaux de régulation génétiques nous ont servi de modèle pour le développement et la mise en place d'un système de programmation moléculaire *in vitro*.

En effet, l'expression d'un gène est majoritairement dominé par des facteurs de transcription, autres protéines ou acides nucléiques, eux-mêmes exprimés par d'autres gènes. L'ensemble forme l'interactome de la cellule, carte globale des interactions entre gènes et sous-produits, où la fonction du réseau est relié à sa topologie.

L'observation des noeuds et sous-architectures dénote trois mécanismes récurrents : premièrement, la nature des interactions est de type activation ou inhibition, ce qui implique que tout comportement non trivial est obtenu par une combinaison de noeuds plutôt que le développement de nouvelles interactions. Ensuite, la longévité du réseau est assurée par la stabilité chimique de l'ADN couplée à la chimiosélectivité des réactions enzymatiques. Enfin, l'aspect dynamique est maintenu par le constant anabolisme/catabolisme des intermédiaires et donc l'utilisation de combustible/énergie.

C'est suivant ces observations que nous avons développé un ensemble de trois réactions enzymatiques élémentaires : la «PEN-DNA toolbox». L'architecture du réseau, à savoir les connections entre les noeuds est médiée par la séquence de brins d'ADN synthétiques (appelés matrice), et trois enzymes (polymérase, nickase, et exonucléase) assurent la catalyse des réactions chimiques. La production et dégradation des intermédiaires consomme des désoxyribonucléotides triphosphates et rejette des désoxyribonucléotides monophosphate, dissipant ainsi le potentiel chimique.

Les réactions sont suivies grâce au greffage d'un fluorophore sur le brin matriciel et au «nucleobase quenching» qui intervient lorsqu'une base d'un intermédiaire se rapproche du fluorophore après hybridation sur le brin matriciel. L'activation correspond alors à la synthèse d'un brin output en réponse à un brin input, alors que l'inhibition survient lorsqu'un brin output s'hybride sur un brin matriciel, empêchant ainsi à l'input correspondant de s'y fixer.

Oscillations, bistabilité et mémoire sont des exemples de comportements implémentés en PEN-DNA toolbox, faisant appel à des architectures de plus en plus complexes. Pour cela, un réglage fin des concentrations en effecteurs (ADN et enzymes) est nécessaire, ce qui sous-tend l'existence de plusieurs comportements pour un même circuit, dépendant des conditions paramétriques. L'établissement d'une carte de chaque combinaison de paramètres avec le comportement global associé permettrait de comprendre le fonctionnement du réseau dans son ensemble.

ble, et donnerait accès à tous les comportements disponibles. Dans le cas d'un système dynamique non linéaire, une telle carte est un diagramme de bifurcation du système.

Pour explorer de manière exhaustive les possibilités d'un réseau dans un cadre expérimental raisonnable, nous avons développé une plateforme microfluidique capable de générer des gouttelettes d'eau dans l'huile à partir de quatre canaux aqueux différents. Ce dispositif nous donne accès, grâce à un contrôle fin des contributions de chaque canal aqueux, à des gouttelettes monodisperses (volume de l'ordre du picolitre) dont le contenu est différent pour chaque gouttelette. Nous avons adapté notre dispositif aux contraintes matérielles (design microfluidique, génération de gouttelettes à contenu différents et contrôlés, observation et stabilité à long terme) et techniques (tracabilité des gouttelettes et stabilité/compatibilité chimique).

Jusqu'à lors les diagrammes de bifurcation étaient calculés à partir de modèles mathématiques décrivant l'évolution des concentrations en effecteurs en fonction des cinétiques enzymatiques de chaque réaction. Le modèle était affiné au regard des données expérimentales, prises en différents points épars dans le domaine des paramètres. Ici, nous générons des millions de gouttelettes et chacune contient une combinaison de paramètres, c'est à dire un point du diagramme. Les gouttelettes sont localisées et leur évolution est suivie par microscopie confocale.

Nous avons appliqué cette technique sur une architecture bien connue, et avons obtenu le premier diagramme de bifurcation bidimensionnel et expérimental d'un système bistable, qui montre des particularités non décrites par les modèles mathématiques précédents. Nous avons par la suite repoussé les limites du dispositif en générant le diagramme de bifurcation tridimensionnel (et toujours expérimental) d'un système prédateur-proie basé sur les équations de Lotka et Volterra, connu du laboratoire. Nous avons pu étudier les effets du partage de ressources sur la croissance des proies et la prédation, facteur difficile à tester par le modèle mathématique.

Cette technique permet donc d'explorer en routine les possibilités de circuits à architectures nouvelles et déceler ainsi des comportements exotiques non mis en évidence par modélisation. Plus généralement, la programmation moléculaire, et plus spécifiquement les réseaux de réactions chimiques à base d'ADN sont une approche innovante pour modéliser les voies de signalisation cellulaire *in vitro*, retranscrire et étudier les interactions complexes qui sont à l'origine du Vivant.

Mots-clés : programmation moléculaire, réseau de réactions, microfluidique, bifurcation, microscopie confocale, bistabilité, prédateur-proie

Abstract: Information processing within and in between living organisms involves the production and exchange of molecules through signaling pathways organized in chemical reactions networks. They are various by their shape, size, and by the nature of the molecules embroiled. Among them, gene regulatory networks were our inspiration to develop and implement a new framework for in-vitro molecular programming.

Indeed, the expression of a gene is mostly controlled by transcription factors or regulatory proteins and/or nucleic acids that are themselves triggered by other genes. The whole assembly draws a web of cross-interacting genes and their sub-products, in which the well controlled topology relates to a precise function.

With a closer look at the links between nodes in such architectures, we identify three key points in the inner operating system. First, the interactions either activate or inhibit the production of the later node, meaning that non trivial behaviors are obtained by a combination of nodes rather than a specific new interaction. Second, the chemical stability of DNA, together with the precise reactivity of enzymes ensures the longevity of the network. Finally, the dynamics are sustained by the constant anabolism/catabolism of the effectors, and the subsequent use of fuel/energy.

All together, these observations led us to develop an original set of 3 elementary enzymatic reactions: the PEN-DNA toolbox. The architecture of the assembly, i.e. the connectivity between nodes relies on the sequence of synthetic DNA strands (called DNA templates), and 3 enzymes (a polymerase, a nickase and an exonuclease) are taking care of catalysis. The production and degradation of intermediates consume deoxyribonucleoside triphosphates (dNTP) and produce deoxynucleotide monophosphates leading to the dissipation of chemical potential. Reactions are monitored thanks to a backbone modification of a template with a fluorophore and the nucleobase quenching effect consecutive to an input strand binding the template. The activation mechanism is then the production of an output following the triggering of an input strand, and the inhibition comes from the production of an output strand that binds the activator-producing sequence.

Various behaviors such as oscillation, bistability, or switchable memory have been implemented, requiring more and more complex topologies. For that, each circuit requires a fine tuning in the amount of chemical parameters, such as templates and enzymes. This underlies the fact that a given network may lead to different demeanors depending on the set of parameters. Mapping the output of each combination in the parameter space to find out the panel of behaviors leads to the bifurcation diagram of the system. In order to explore exhaustively the possibilities of one circuit with a reasonable experimental cost, we developed a microfluidic tool generating picoliter-sized water-in-oil droplets with different contents. We overcame the technical challenges in hardware (microfluidic design, droplet gen-

eration and long-term observation) and wetware (tracability of the droplet and emulsion compatibility/stability).

So far, bifurcation diagrams were calculated from mathematical models based on the enzymes kinetics and the thermodynamic properties of each reaction. The model was then fitted with experimental data taken in distant points in the parameter space. Here, millions of droplets are created, and each one encloses a given amount of parameters, becoming one point in the diagram. The parameter coordinates are barcoded in the droplet, and the output fluorescence signal is recorded by time lapse microscopy. We first applied this technique to a well-known network, and obtained the first experimental two-dimensional bifurcation diagram of the bistable system. The diagram enlightens features that were not described by the previous mathematical model.

The next step was then to increase the potential of the device by adding one more dimension. We successfully performed the three-dimensional mapping of a richer network : the predator-prey system based on the Lotka-Volterra equations, previously developed in the laboratory. We prospected the effects of common resources on prey growth and predation, that would have been difficult to predict by upgrading the existing mathematical models.

Finally, using this technique, we aim at delving into the possibilities of circuits displaying new architectures or mechanism leading to exotic and/or unique behaviours. This new mechanism will open up the way to more elaborated architectures, as much as a convenient way to model non-trivial regulation pathways in cell. More generally, the building of self-organizing DNA chemical reaction networks are an innovative approach to the chemical origins of biological complexity.

Keywords : molecular programming, reaction networks, microfluidics, droplets, bifurcation, confocal microscopy, bistability, predator-prey.

Cette thèse est dédiée à Raymond Bordier, Mohammed Heidi Baccouche et Mouncef Baccouche.

Acknowledgments/Remerciements

First, I would like to express my gratitude to my supervisor, Yannick Rondelez, who gave me the opportunity to work with him, providing countless advices, suggesting clever ideas and thanks to whom I constantly learnt for three years. Yannick was always available despite his busy schedule, to guide, motivate, and correct me when necessary.

I would like to thank the members of the jury for being available to assess this work: Prof. Olivia Reinaud, Dr André Estévez-Torres, Prof. Collard as well as my two referees, Dr. De Greef, Dr. Gueroui.

I thank my sempai Adrien Padirac who taught me everything about experimental assembly of molecular programs and how to appreciate life in Japan, Anthony Genot for teaching me microfluidics, microscopy, and correcting my english (may the Rolling Stones never end), as much as Nathanaël Aubert, Kevin Montagne and Guillaume Gines for all the fruitful discussions on research, life and all the experience we shared. And I wish good luck to the newcomers: Adèle Drame-Maigné and Rémi Sieskind. A special thank to Alexis Vlandas, André Estévez-Torrès and Nicolas Bredeche for the fruitful discussions and good moments we spent.

I thank the Prof. Teruo Fujii and all Fujii Lab. members, who informed and guided me in this environment, with a special thank for Christophe Provin, Shohei Kaneda, Soo Hyeon Kim and Toshiro Maekawa. I would like to thank as well all my coworkers in LIMMS and IIS for all the precious moments we spent: Denis, Filiz, Mehul, Tomoko, Pierre, Cagatay, Guillaume P, Marie, Wenjin, Nathalie, Yannick T...

I would like to thank Nagatsuka-sensei and the Gisenkai members for allowing me in the Nakameguro branch of Tamiyaryu Iaijutsu and teaching me so much about iaido, Japan, and peace of mind. You are my family here.

Je tiens à remercier ma famille pour sa présence et son soutien inconditionnel pendant ces trois années (mention spéciale à ma mère pour le ravitaillement en fromage), ainsi que le Pr Lotz, le Dr Selle et le service d'oncologie médicale et thérapie cellulaire de l'hôpital Tenon sans qui je n'aurai pas pu faire cette thèse.

Un merci éternel à mes mentors : Diana Over, Gregory Thiabaud, Benoit Collasson, Olivia Bistri, Jean-Noël Rebilly et Luc Tamisier pour m'avoir donné une ligne de conduite et des exemples à suivre.

Un merci très spécial à Axelle pour ses relectures attentives, ainsi que pour avoir pris soin de Cicéron pendant plus de trois ans.

Enfin, je voudrais remercier tous les amis qui m'ont soutenu de près comme

de loin, de jour comme de nuit : Mutsumi, Virginie, Mathilde, Tiphaine, Juan, Alexandra, Rémi, Marine P, Johnson, Richa, Patrice, Walid, Yacine, Hajar, Youssef...

Plus particulièrement, je me souviens très bien de ces dimanches passés avec mon père à arpenter toutes les salles du Palais de la Découverte. Du temps que tu prenais pour expliquer à un enfant trop impatient les rouages les plus subtils de la physiologie, l'importance de la clinique pour établir un diagnostic, ainsi que la conduite à tenir face à un patient en insuffisance cardiaque... Plus généralement, je me souviens de tous les efforts que tu as déployés pour me transmettre ta passion pour les Sciences. Pour tout cela, merci papa.

Contents

Contents	9
List of Figures	13
List of Tables	17
1 Introduction	19
1.1 Structure of nucleic acids	20
1.2 Tools for molecular programming	23
1.3 Structural DNA nanotechnology	26
1.4 Molecular programming	29
1.4.1 RTRACS	31
1.4.2 Genelets	33
1.4.3 PEN DNA toolbox	34
1.4.4 Computation and modeling	44
2 Droplet-based parameter scanning	53
2.1 Introduction	53
2.1.1 Compartmentalizing	55
2.1.1.1 Co-flowing	60
2.1.1.2 Cross-flowing	61
2.1.1.3 Flow-focusing	61
2.1.2 Design	63
2.1.2.1 Design of fluidic resistances in the chip	66
2.2 Microfluidic device fabrication	68

2.2.1	Mold fabrication	69
2.2.2	Chip casting	70
2.3	Oil & Surfactants	71
2.4	Script generation, barcoding, and parameter space . .	75
2.4.1	Drawing a shape in parameter space	76
2.4.1.1	1 dimension: the line	76
2.4.1.2	2 dimensions: triangle and square . .	77
2.4.1.3	The cube	81
2.4.2	Sorting droplets	84
2.4.3	Barcoding	86
2.5	Observation chamber	89
2.5.1	Confocal microscope setup	95
2.6	Imaging	95
2.6.1	Time-lapse	96
2.6.1.1	Fast settings	98
2.6.1.2	Time constraint	98
2.6.1.3	Colors	99
2.6.1.4	Optical defects	99
2.6.2	Endpoint method	100
2.7	Conclusion	101
3	Mapping a bistable circuit	103
3.1	Introduction	103
3.1.1	bistability: two examples in biology	103
3.1.2	Introduction to bifurcation theory	106
3.1.3	Two PEN-DNA-toolbox examples	107
3.2	Results	113
3.2.1	Reaction assembly	113
3.2.2	Data analysis	115
3.2.2.1	Data collection	115
3.2.2.2	Image processing	117
3.3	Interpretation	121
3.3.1	Simulations	121

<i>CONTENTS</i>	11
3.3.2 Discussion	130
3.4 Conclusion	132
4 Mapping the predator-prey system	135
4.1 Introduction	135
4.2 Mapping the bifurcation diagram	138
4.2.1 Experimental implementation	139
4.2.2 Barcodes and reporters	142
4.2.3 Timelapse	145
4.2.3.1 Scanning time	145
4.2.3.2 Cost and storage	147
4.3 Results and analysis	147
4.3.1 Reaction assembly	147
4.3.2 Image processing	149
4.3.3 About the diagram	153
4.4 Discussion	158
4.4.1 Bifurcation analysis	161
5 Conclusion	169
Bibliography	175

List of Figures

1.1	Overview of the DNA helix	22
1.2	DNA operators for molecular programming	25
1.3	Timeline and selected breakouts in structural DNA nanotechnology	28
1.4	Implementation of a Hamiltonian path problem according to Adleman's strategy	29
1.5	Canonical Gene Regulatory Network	30
1.6	RTRACS: a RNA/DNA/enzyme molecular programming machinery	32
1.7	Genelet system	34
1.8	PEN-DNA toolbox: overview	36
1.9	General workflow of PEN-DNA toolbox-based network implementation	38
1.10	Biochemistry of PEN-DNA toolbox	40
1.11	Evaluation of several autocatalyst modules	43
1.12	Analog versus digital computation in cells	44
1.13	Mapping the dynamical behavior of a nonlinear chemical system	46
1.14	Schematic representation of the mechanism of the DNA predator-prey system	47
1.15	Bifurcation diagram of the 2-dimensional model obtained by LSA	51
2.1	Theoretical and experimental bifurcation diagram of a memory switch	55

2.2	Three ways to generate droplets	57
2.3	Droplet size control in flow-focusing	62
2.4	Microfluidic device for diagram mapping	64
2.5	Dust from solution/tubing trapped by the filters. . .	65
2.6	Preinjector and nozzle detail: microfluidic parameters	66
2.7	chip-prototyping methods	68
2.8	Marangoni effect on two approaching droplets	73
2.9	Non-ionic fluorosurfactant adapted from Krytox . . .	73
2.10	Strategy for deterministic triangle mapping	78
2.11	Pressure profiles for triangle mapping	79
2.12	strategy for deterministic square mapping	80
2.13	Pressure profiles for square mapping	82
2.14	Pressure profiles for random-walk cube mapping . . .	83
2.15	Droplet indexing: rail and anchors	84
2.16	Droplet indexing: direct readout	86
2.17	Molecular barcoding	88
2.18	Fluorescence spectra for the barcodes Dextran Cas- cade Blue and Alexa 647	89
2.19	Observation chamber developed by Hasatani <i>et al.</i> . .	91
2.20	Microscopy bloopers	94
2.21	Observation chamber filled with emulsion	94
2.22	Confocal principle	97
3.1	Positive feedback loop and bistability plot of MAPK pathway	104
3.2	Switching probability of a cell as a function of its initial growth rate and initial fluorescence	105
3.3	Bistable circuit with two autocatalysts : reactions net- work and implementation with PEN-DNA toolbox . .	109
3.4	Alternative bistable circuit : reactions network and corresponding implementation in PEN-DNA toolbox framework	110
3.5	Two bistable circuits and computed stability plots . .	112

3.6	Collection of fluorescence levels for each droplet . . .	116
3.7	Image processing workflow	118
3.8	Histogram of the distribution of raw fluorescence in droplets for a bistable circuit started in $\alpha\beta = 01$. . .	119
3.9	Experimental bifurcations in the bistable circuit . . .	120
3.10	1D experiment on bistable circuit	122
3.11	Simulated bifurcation diagrams of bistable circuit . .	125
3.12	Fitting performances of 3 kinetic models	129
4.1	Predator-prey system. Principle and PEN-DNA toolbox implementation	136
4.2	Reaction network of the predator-prey system	137
4.3	Principle of the PP mapping	139
4.4	Timetraces 9bp long PP system vs temperature. . .	141
4.5	Two nicking enzyme restriction site for two strategies	142
4.6	Four fluorophores used to map the diagram	142
4.7	Gradient of grass template in predator-prey	144
4.8	Scanning area limitation	146
4.9	Droplet array for the predator-prey system with Bst2.0 WarmStart polymerase	148
4.10	Number of droplets detected in each frame and max of jump lengths for each trajectory	150
4.11	Trajectories after tracking and reconstruction	152
4.12	Scanning result of the 3D experiment	154
4.13	Droplet array for the predator-prey system	155
4.14	Zoom in droplet array for the predator-prey system .	156
4.15	245 timetraces from the droplet array	157
4.16	Timetraces and bifurcation diagram of the predator-prey oscillator	159
4.17	Best oscillators	160
4.18	Bifurcation analysis of the model defined by equations	162
4.19	3-dimensional bifurcation diagrams of the 2 variables model along <i>pol</i> , <i>tem</i> , and <i>exo</i>	164

4.20	Detail on bifurcation diagram: plotting a line	165
4.21	Stochastic bursting in the 2 variables model for oscillators close to the extinction area	166

List of Tables

2.1	Summary of techniques for droplet emulsion	58
2.2	Dimensionless key numbers in microfluidic with ρ the density, U_0 the velocity, L_0 critical length, μ the viscosity, γ the surface tension and D the diffusivity	60
2.3	Commonly used oils and surfactants for bio-compatible droplet generation	75
2.4	Available fluorophores according to the laser wavelengths of the confocal microscope	97
3.1	DNA strands used for the bistable switch	113
3.2	Submixes composition for bistable diagram mapping	114
3.3	Acquisition parameters for confocal microscopy. The percentages indicate the laser power, “C.A” refers to the confocal aperture and “VBF” to variable barrier filter.	115
3.4	Default values of optimized parameters	128
4.1	Melting temperatures for grass/prey adducts	140
4.2	Candidates for predator-prey mapping and corresponding barcodes	143
4.3	Sequences of DNA strands used in the predator-prey network	147

Chapter 1

Introduction

Since its discovery in 1868 by Meischer [1], nucleic acid polymers, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are among the most fascinating biomolecules. DNA carries out multiple tasks regarding the storage, process and expression of genetic information. First described as phosphorous-rich and lacking sulfur compounds, and contained within the nuclei of cells, it took 85 years before the elucidation of its structure by Franklin, Crick and Watson [2].

The robustness of Watson-Crick base-pairing, coupled with advances in the solid-state synthesis of nucleic acids, has powered many revolutions in molecular biology: cloning [3], sequencing [4], polymerase chain reaction [5], RNA interference [6] and recently CRISPR [7], a powerful tool for genome editing which is widely seen as a trigger of future biochemical, medical and possibly societal revolutions.

But in the past 3 decades, molecular engineers have also come to realize that DNA could be used on its own to build and program at the nanoscale. So DNA computing was born and electrified the scientific community. The community of DNA computing has since grown considerably, extending beyond computing and morphing into the larger endeavor of “molecular programming” with two lofty goals: how can we program molecules to compute and control matters? What does it tell us about computation in the biological world? Actually, nature and molecular programmers face similar

conceptual constraints: they both need to make molecular circuits that use as little molecules as possible, while being robust and versatile and of course compliant with the law of physics. The gist of molecular programming is thus an unusual mixture of (obviously) chemistry and computer science, but also biology, physics, mathematics and information processing.

In this thesis I will introduce a new device for experimental mapping of bifurcation diagrams of chemical reaction networks assembled with the PEN DNA toolbox and show two examples of exploration in multi-dimensional parameter space. In chapter two, I will first introduce the basics of compartmentalization and microfluidics before describing the microfluidic device developed for high throughput generation of water-in-oil droplets with different contents, its design and micro-fabrication, the technical challenges underlying the choice of reagents, the strategy built to sort the droplets and their observation overtime. In Chapter 3 this process is applied to a bistable circuit to obtain its bifurcation diagram as a function of two parameters. The singularities of this map will be discussed. In Chapter 4 the same process is upgraded and assigned to a predator-prey system, leading to a three dimensional map. The complex behaviors displayed will be analyzed according to the reactivity of the species involved in the context of small compartments.

1.1 Structure of nucleic acids

DNA consists of a sequence of monomeric units -nucleotides- showing a 2-deoxyribose coupled with a phosphate group on the carbon 5 and a nucleobase on the carbon 1. The monomers are coupled by the hydroxyl in carbon 3 (C3) and the phosphate in carbon 5 (C5) of another nucleotide, forming the primary structure: a DNA strand with a succession of nucleotides (defined as the sequence of the strand). The secondary structure is a right handed double helical shape, obtained when two complementary strands meet in solution

and passively bind (hybridize) to form rather rigid double helices (Fig.1.1, a).

Several helical structures are found, starting from the B-DNA, the *in vivo* classical form. The A-DNA is quite similar to B-DNA, but shorter, and is found in dehydrated conditions. Z-DNA is a left-handed helix whose occurrence is function of the sequence. We will focus here on the most common architecture, that is the B-DNA (Fig.1.1, b). This canonical structure shows two grooves: a 2.2 nm wide groove a 1.2 nm wide minor groove and a 2 nm wide groove, giving easier access to the bases for interacting agents.

In DNA, the supramolecular recognition and pairing between two complementary strands follows the celebrated Watson-Crick base-pairing rules: an adenine nucleotide faces a thymine and share two hydrogen bonds (H-bonds), while a cytosine residue faces a guanine one with 3 H-bonds (Fig.1.1, a). This, added to the covalent linkage between two nucleotides from the same strand give a direction to the strands: from the free hydroxyl on C3 of the first nucleotide (3' end) to the phosphate group on C5 of the last one (5' end). Two strands are named antiparallel when their sequences are complementary and parallel, but go on opposite direction (Fig.1.1, c). The pitch of the helix contains 10.5 bases (3.32 nm) for a diameter of 2 nm. The non-covalent bonding allows the duplex to de-hybridize under heat or chemical denaturing agents. The stability of the duplex is ensured by the sum of H-bonds (10-20 kJ.mol⁻¹ per H-bond) and the stacking occurring between the bases inside the helix (~10 kJ.mol⁻¹ per stacking), but other factors influence the stability, such as the temperature and the amount of salts inside the medium. As a negatively-charged molecule, the DNA structure is sensitive to the ionic force of the medium, and in particular to the presence of Na⁺ and Mg²⁺.

RNA shares the same backbone structure as DNA, except that the sugar keeps its hydroxyl residue in C2, and the thymine is replaced by an uracil motif. Because of the added constraints induced

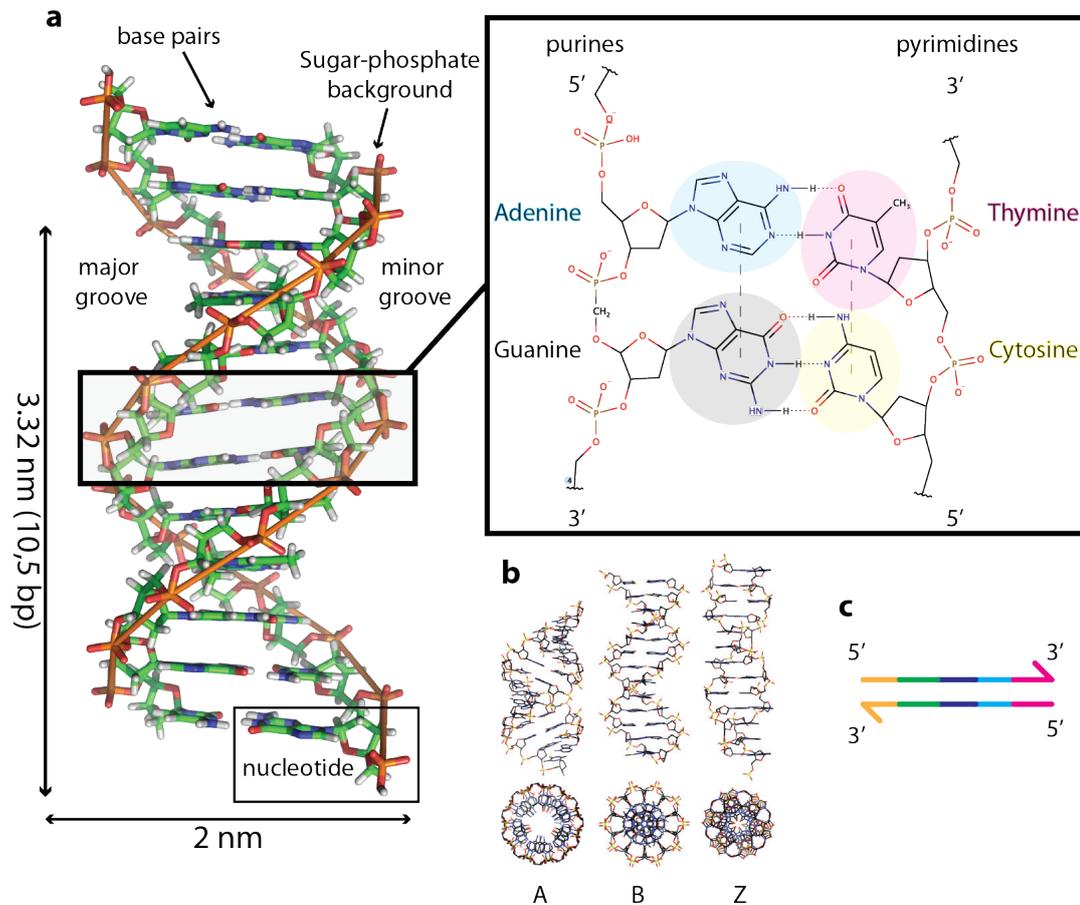


Figure 1.1: Overview of the DNA helix:(a) the double helix B structure results from the non covalent assembly of two DNA molecules. The double helical shape displays a major groove (width: 2.2 nm) and minor groove (width: 1.2 nm). Proteins interacting sequence-specifically with DNA preferentially bind on the major groove, since it provides a better access to bases -hence better sequence-specificity. The classical DNA duplex is 2 nm wide and the pitch is 3.32 nm long (i.e. 10.5 bases). Alternate structures exhibits different parameters. The bases are paired according to Watson-Crick base-pairing: an adenine residue from one strand is connected to a thymine residue from the second strand by two hydrogen bonds. Similarly, a cytosine residue from one strands is linked to the guanine residue of the second strand with 3 hydrogen bonds. This makes the strands of the double helix antiparallel. The bases are arranged in co-axial layers leading to highly stabilizing stacking interactions and are covalently bond from the hydroxyl group of the carbon 3 to the next nucleotide by the phosphate group of the carbon 5. (b) Alternative architectures for the DNA duplex: A, B and Z. A-DNA is the form found in RNA/RNA or RNA/DNA duplexes, often in dehydrated medium. B-DNA is the most common architecture *in vivo*. Z-DNA is a rarer form, found in high salt concentrations and require a particular pattern in the sequence. Non duplex structures, such as triplex or quadruplex DNA also exist. Image from Wikipedia. (c) Schematic representation of two antiparallel half-arrows. The head of the arrow represents the 3' end of the strand, bearing the OH group available for further polymerization of the strand (hence the direction of the arrow, which indicates the growth direction in the case of enzymatic polymerization). Each color represents a domain, i.e. a sequence of nucleotide encoding one bit of information in engineering approaches based on DNA self-assembly.

by the hydroxyl group, the structure of the base paired polymer is mostly A-DNA-like, with an exaggerated, deep major groove and wide minor groove. The structures found in RNA are various and complex, but the most interesting point is the number of tertiary structures (stabilized with metal ions such as Mg^{2+}), which are directly involved in the numerous functions of RNA in cell: information processing (translation, protein synthesis), regulation (microRNAs, riboswitches and riboregulators), RNA processing itself, viral mechanism (reverse transcription, double stranded RNA), catalysis (ribozymes, ribosome). This ubiquity has led to the RNA hypothesis for origin of life [8].

1.2 Tools for molecular programming

The specificity of nucleic acids comes from their possibilities in terms of structure and reactivity. Indeed, the first remarkable point is the controlled self-assembly, or hybridization: two complementary DNA strands will spontaneously assemble when they meet, following Watson-Crick base-pairing rules (Fig.1.2, a). The stability of the duplex is expressed as a sum of all interactions of the base pairs and stacking. The melting temperature (T_{melt}) of a given duplex (or sometime by extension for a single sequence, assuming it is paired with its perfect complementary strand) is defined as the temperature required for half of the population to be single stranded. The T_{melt} therefore depends also on the concentration of the strand in solutions and the salinity of the buffer (which must be specified for the T_{melt} value to make sense).

Currently, our understanding of the base-pairing self-assembly is quite limited, even for a «simple» hybridization. The most used model to describe the transition from single stranded DNA (ssDNA) to double stranded DNA (dsDNA) is a coarse-grained nearest neighbor model (n-n) proposed by SantaLucia *et al.* [9]. Finer model such as force field include much more degrees of freedom, and are thus

impractical for computational purposes. The model by SantaLucia reproduces the experimental melting temperatures for duplexes up to 16 base pairs (bp) with only 2.3 °K deviation. In the n-n model, stability is defined in terms of two successive base pair doublets. The effect of base-pairing, stacking and solvation will be considered for each doublet and the stability is then calculated for each sequence. An useful approach relies on the two-state melting theory (which assume the absence of significant intermediates between fully dissociated state and fully bound state during the hybridization-dehybridization reaction):

$$\frac{[AB]}{[A] + [B]} = \exp(-\beta(\Delta H_{AB} - T\Delta S_{AB}))$$

where ΔH_{AB} and ΔS_{AB} are the duplex melting transition enthalpy and entropy, coming from the sum of all n-n interactions and base-pairing, while β is the nucleation parameter containing the concentration dependence plus the external factors involved in duplex formation. Once all nearest neighbor parameters are tabulated, this simple approach can predict the approximate T_{melt} of arbitrary duplex, but the two-states model is not fine-grained enough to thoroughly study the dynamics of DNA systems, neither can it predict the effects due to the topology/geometry.

An application of DNA pairing programmability is the modulation of the kinetics of strand exchange of a duplex with a ssDNA invasion strand, using an overhang adjacent to the helical domain of the duplex. This toehold architecture (Fig.1.2, 2a) was proposed for the first time by Yurke et al. [10] in a paper describing a «DNA tweezer». The tweezer is a molecular motor acting like a scissor using DNA strands as fuel. The core of the tweezer on its open state is made of three strands A, B and C. B and C partially hybridize on A, occupying 18 bases each, 24 dangling bases are left on B (5' end) and C (3' end) (Fig.1.3, green, a). The F strands closes the tweezer by hybridizing the dangling edges of B and C (Fig.1.3, green, b), Thus,

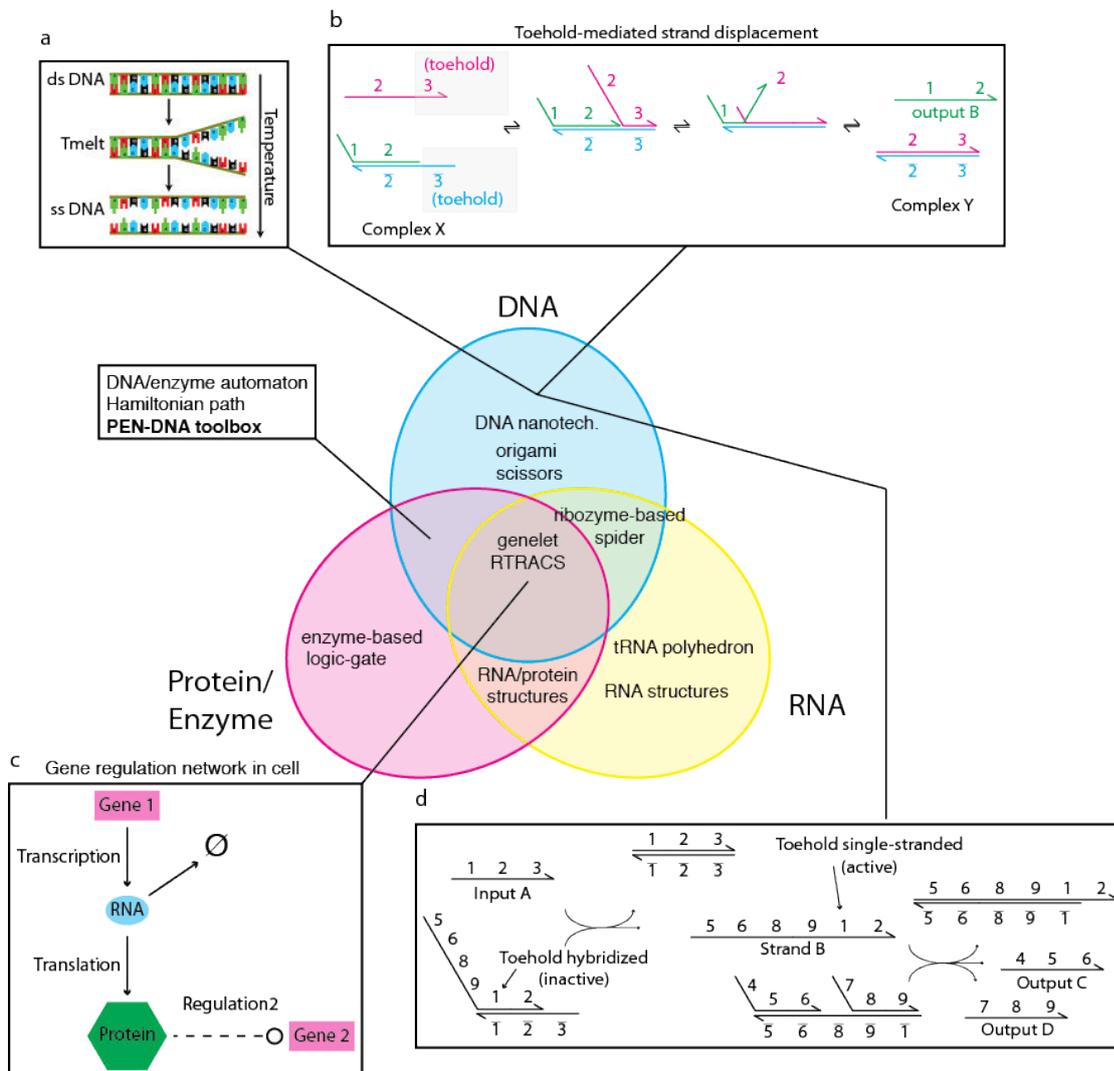


Figure 1.2: DNA operators for molecular programming (MP). (a) hybridization-denaturation by temperature. Under a given temperature, the DNA duplex spontaneously de-hybridizes and the single strands (ssDNA) are free to interact with other strands or through single strand specific enzymatic reactions. The melting temperature (T_{melt}) is directly linked to the sequence involved and can be calculated thanks to online algorithms. (b) Toehold-induced strand-displacement[10]: it is possible to release a ssDNA from a duplex that is partially unpaired. The input strand will first bind the unpaired part of the duplex and the output will be released because of the instability of the complex $X + input$. (c) *In vivo* gene regulation networks are a cellular inspiration for the control of DNA reactivity, by the mean of enzymes. These chemical reactions are irreversible, and typically far from-equilibrium. (d) Toehold mediated strand displacement cascades: using the thermodynamics of hybridization-denaturation, the complementarity of the sequences and given a set of partially complementary strands, one input strand can dehybridize a duplex for which a part of the sequence is shared and finally release an output sequence.

an overhang part on F allows the hybridization of \bar{F} , that completely hybridizes with F , therefore releasing the tweezer in its open state (Fig.1.3, green, c). The F strand is the fuel because of the higher stability of the duplex compared to the other species. By extension of this principle, it is possible to cascade the masking/unmasking of toeholds and release a strand duplexed with a longer one along a controlled path of strand displacements, as shown in Fig.1.2, b). This toehold-mediated strand displacement, quantitatively analyzed by Zhang and Winfree [11], is a powerful tool and a breakout in the field of DNA nanotechnology. Using the toehold feature increases the rate of displacement by up to 10^6 times compared to the invasion of a blunt duplex [12]. Recent approaches have been proposed to decrease the leak and may therefore bring further usability to this approach.

Also, an other feature usable to engineer DNA comes from the observation of what is occurring *in vivo*: enzymatic process. Many enzymes are involved in the production, modification or digestion of nucleic acid polymers. For example, in case of genetic regulation, as shown in Fig.1.2 4, a gene will see its transcriptional activity (the production of RNA copies of its sequence by a RNA polymerase) regulated by effectors (often proteins) produced upstream in the regulation pathway. These networks of regulations define a lot of computational tasks essential for the cell survival (historically the first example described was the Lac operon [13][13]).

1.3 Structural DNA nanotechnology

The crystallographer Ned Seeman was the first to perceive in the 80s that the self-assembly of DNA could herald a new kind of nanotechnology. Using only the hybridization/denaturation and sequence recognition properties, Seeman and the pioneers of the field started to engineer molecular shapes made out of DNA, such as a tetrahedron [14] (Fig.1.3).

Seeman hypothesized that DNA strands with the right sequences ought to self-assemble into a 3D grid, forming a rigid framework that he expected to use to firmly latch proteins. In other words, he only needed to make a 3D DNA crystal to bypass protein crystals. It took Ned Seeman and colleagues many years to demonstrate the first 2D DNA crystal [15] from the initial publication of the tetrahedron design and ten more years to achieve his dream of a 3D DNA crystal (structural studies of proteins are still undergoing). But meanwhile, the idea of using synthetic DNA for non-biological purposes had blossomed into a burgeoning field that discovered how to fold a plethora of nanostructures: smileys and maps [16], twisted jars and meshed spheres [17], Moebius strips [18], the qwerty keyboard and its emoticons... Much of this revolution was fueled by the invention of DNA origamis by Paul Rothemund [16] (Fig.1.3): a technique to fold a long DNA scaffold into a prescribed thanks to the assistance of a swarm of helping strands called staples. This expansion reached the field of robotics and lead to the development of tweezers [10] or DNA walkers, such as the DNA spider [19] (Fig.1.3)

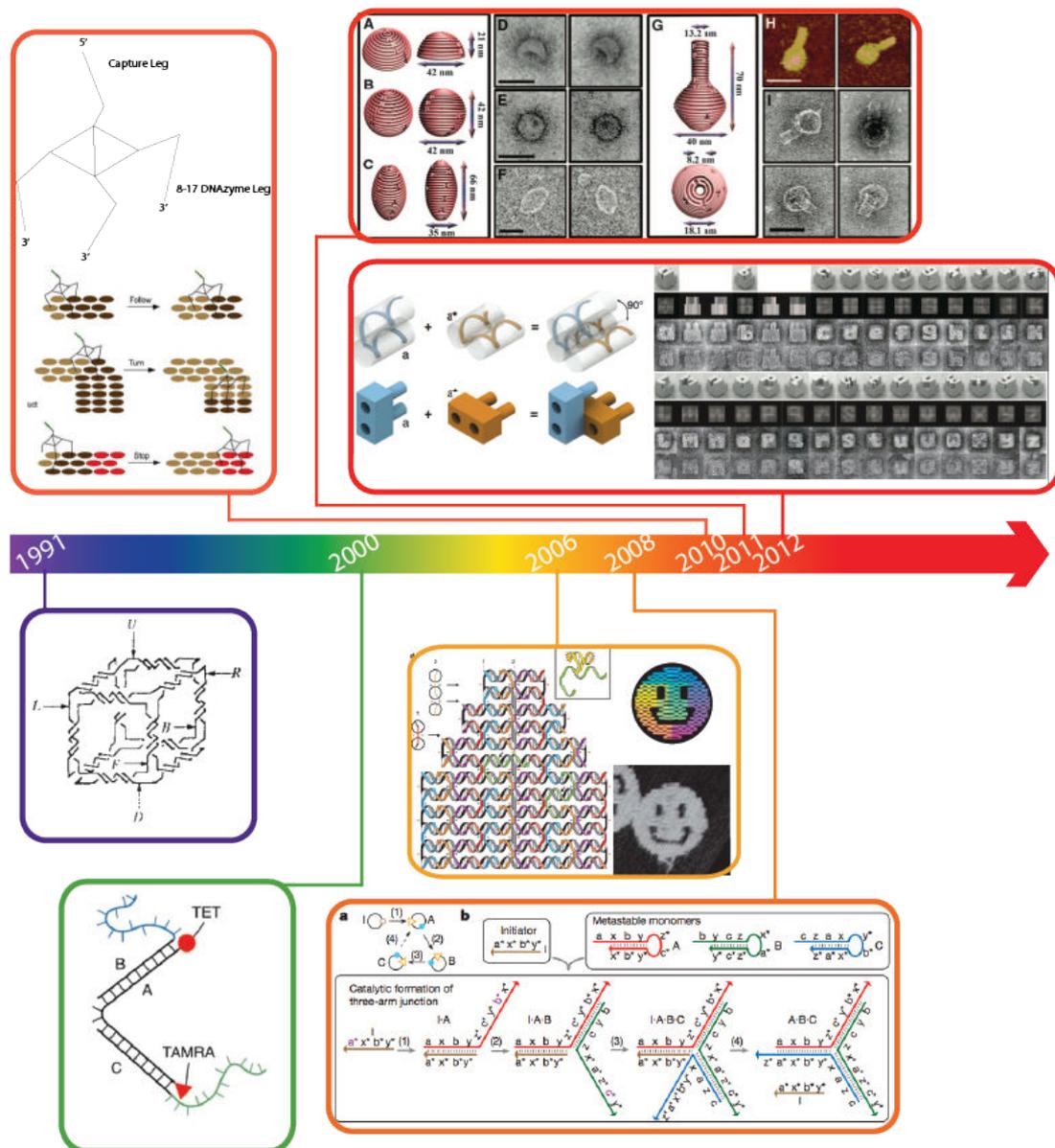


Figure 1.3: Timeline and selected breakouts in structural DNA nanotechnology. **1991:** Chen and Seeman designed and characterized a set of DNA strands self assembled into a cubic shape [14]. The design involves partially complementary strands. **2000:** Yurke *et al.* design and implement the «DNA tweezer» [10], opened or closed by consuming other DNA strands. **2006:** a revolution. Paul Rothemund synthesizes the DNA origami [16], consisting of a long circular DNA strand (scaffold) and hundreds of smaller ones (staples) designed to a planar shape the scaffold by successive hybridizations (left). Famous smiley shape, imaged by AFM (right). This work generalized the use of DNA as a building block for molecular architecture/robotics. **2008:** Peng Yin *et al.* describe a self-assembly of three and four arms junction [20]. (a) three arm junction network. (b) Detailed mechanism showing the initiator strand and monomers with toeholds. **2010:** Lund *et al.* are describing the walk of a DNA based molecular motor named DNA spider (up) on a track made of a modified origami (down) [19]. **2011:** time to go 3D: Han *et al.* use curved origamis to design 3D shapes [17]. **2012:** Ke *et al.* developed self-assembled tile for multiple 3D design [21] .

1.4 Molecular programming

The proteiform DNA stunned again the world in 1994 when Leon Adleman (the “A” in the RSA cryptography protocol) used it to solve a complex computational task [22]. He attacked the Hamiltonian Path problem: given a graph with a list of cities and road connections between, is there a path that starts from a given city, goes through each city once and finishes at a specified terminus city (Fig.1.4, a). The Hamiltonian path is an archetypal NP-complete problem; there is (probably) no quick algorithm to crack it and the brute force method of testing all possible paths is thought to be the best.

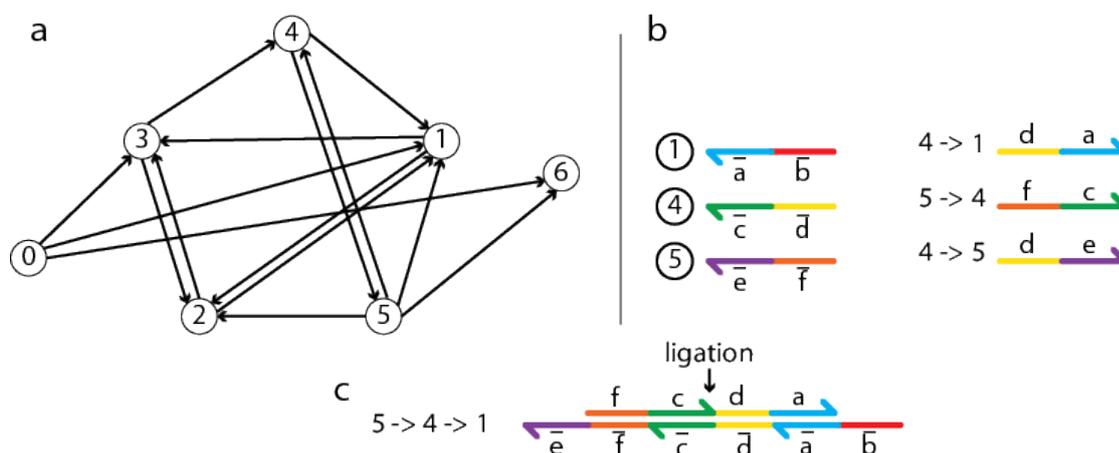


Figure 1.4: Implementation of a Hamiltonian path problem according to Adleman's strategy. (a) Circled numbers are the cities and arrows the existing airplane connections. How to find the path which goes from 0 to 6 and stops only once in each city? (b) Cities and roads are implemented with DNA strands, the sequence a is complementary to the sequence \bar{a} . Airport departure and arrival becomes the 5' end of the starting city and the 3' end of the arrival city. (c) Flight strands hybridize on the corresponding city strands and a ligase binds two adjacent strands. The resulting DNA molecule that starts with the strand 0 and ends with the strands 6 and has a length of 7 cities is the solution of the Hamiltonian path. Strands are then recovered through conventional molecular biology process.

Adleman encoded the cities as DNA strands and the connections between them as strands partially complementary to the city strands (a road strand forms a splinter between the two city strands that it connects). Hence, each city strand can bind to many different flights

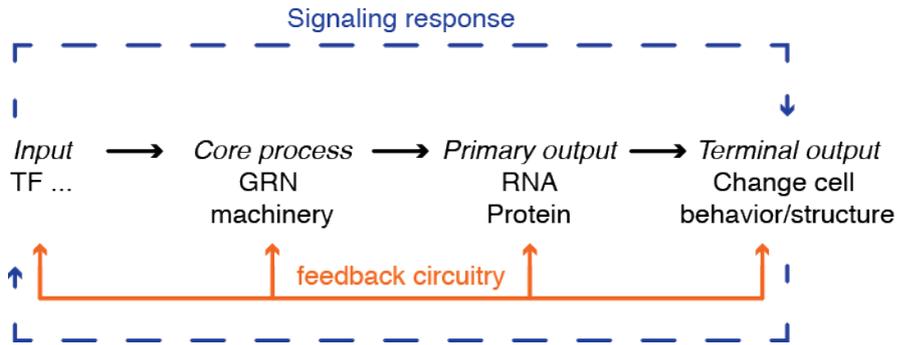


Figure 1.5: Canonical Gene Regulatory Network: an external input initiates the signal transduction to the core component complex composed of proteins interacting with DNA to produce the primary output (RNA or protein) itself interacting with other cellular structure, resulting in a change within the cell phenotype. Direct and indirect feedbacks (orange) might occur at any level and are crucial to maintain the efficiency of the GRN. Signaling responses and feedbacks (blue dashed) lay beyond GRN boundaries (e.g. chemotaxis process), squeezing the DNA machinery to target other structures.

strands by Watson-Crick pairing and conversely each road strand can bind to many different cities. The succession of road-to-city and city-to-road hybridizations yields a long DNA complex encoding one possible path (Fig.1.4, b). The magic of chemical parallelism took place when Adleman mixed the road and city strands. An almost infinity of possible paths were spontaneously generated by hybridization: the paths that loop through each city twice, the paths that end up nowhere, the paths that incessantly bounce between two cities.... Because of the number of molecules involved, Adleman was guaranteed to find the Hamiltonian path among this vast collection of paths, if it existed (Fig.1.4, c). Aided by classical tools from biochemistry and sequencing, he extracted and sequenced the strands encoding the Hamiltonian path, yielding a molecular solution to his problem.

In cells, Gene Regulatory Networks provide a completely different, more general framework for the implementation of computing tasks at the molecular level [23, 24, 25, 26, 27]. Indeed genes can be wired in networks of cross interactions through, for example, the expression of regulatory proteins or transcription factors (Fig.1.5).

Only very recently have similar general chemical reaction networking frameworks been described *ex vivo* [26, 27]. Similar to what has been done for neural networks, whose fundamental features have been abstracted into the computational framework of Artificial Neural Network circuits (ANN), these approaches focus on the most important dynamic properties of biological GRN to define functional *in vitro* models. Therefore it has become possible to construct *in vitro* reaction networks with well-controlled topologies [25], targeting a precise dynamic function and reproducing biological architectures. Extracting the essential dynamic features of GRN from the viewpoint of dynamical systems, one is left with a set of collective molecular transformations that can be linked in such a way that the product of one either activates or inhibits the production of another [26, 28]. A second important feature is that the network linking these reactions is hardcoded in the sequence of stable DNA strands (genes and promoters). The long-term stability of these DNA species stands in sharp contrast with the dynamic behavior of their products (RNA and proteins) that are constantly produced and degraded/diluted. Therefore, the maintenance of a constant flux of energy through the system, together with a precisely controlled reactivity landscape are also essential features of GRN. While genetic regulation in cells uses a bewilderingly complex molecular machinery, these three essential characteristics have served as inspiration in a small number of *in vitro* simplified schemes able to implement *artificial* regulatory networks: RTRACS, the genelets approach and the PEN DNA toolbox.

1.4.1 RTRACS

One example of molecular program involving DNA, RNA and enzymes, inspired by retroviral replication was proposed by the group of Prof Suyama at the University of Tokyo. The Reverse-transcription-and-TRanscription-based Autonomous Computing System (RTRACS) processes a RNA input into a RNA output thanks to DNA-and-

enzymes machinery (Fig.1.6, a). The hardware is composed of a reverse transcriptase to transform the input into the equivalent DNA sequence, a DNA polymerase for DNA strand processing and a combination of RNA polymerase and DNA-RNA converter to generate the output. The system keeps its dynamics thanks to the RNase.

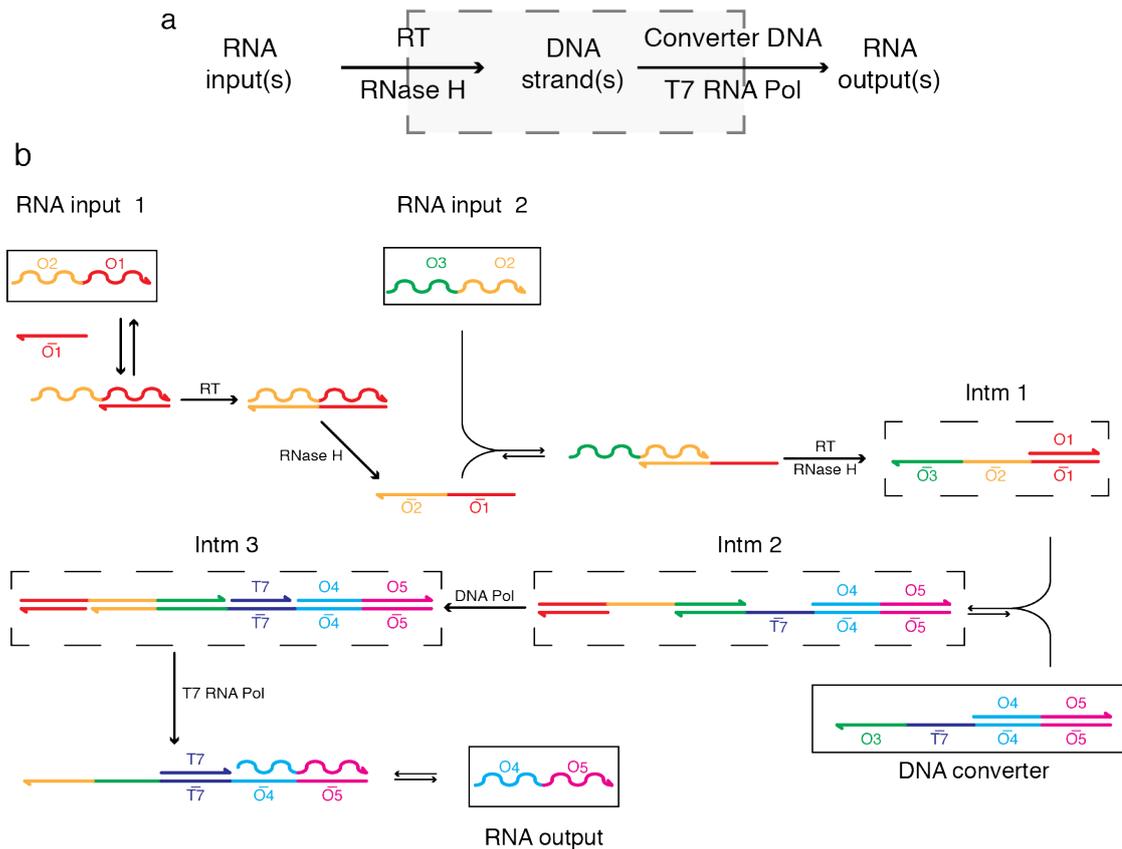


Figure 1.6: RTRACS: a RNA/DNA/enzyme molecular programming machinery. (a) Global framework: a RNA input is processed after reverse transcription in DNA and the molecular computation ends with the expression of a RNA output by the T7 RNA polymerase (T7 RNA Pol). (b) Implementation of an AND gate: wavy lines stand for RNA strands and straight ones show DNA strands. Two RNA inputs lead to the intermediate species 1 (Intm 1), fulfilling the condition imposed by the AND gate (i.e. both inputs are required to express the DNA sequence $\bar{O}3$). Intm1 then partially hybridizes on the DNA converter to Intm2, on which the $\bar{T}7$ sequence is single stranded, generating a substrate for the DNA Polymerase (DNA Pol). Intm1 and the O4O5 sequences are hiding the rest of the converter from DNA Pol. The goal is the production of Intm3, the trigger of output. Indeed the T7 sequence is the promoter of T7 RNA Pol, which recognizes the Intm3 pattern and synthesizes the RNA output.

The first experimental implementation was an AND gate [29] (Fig.1.6, b). The AND gate needs both inputs to deliver the output.

The strategy lays in the condition of generation of the Intm3, the precursor of the output. The sequence $\bar{O}3$ is the key to the output production: given the sequence of the DNA converter, knowing that the pattern for the T7 RNA Pol has to have its promoter double stranded (adduct T7 $\bar{T}7$), only the sequence $\bar{O}3$ allows the polymerization along the converter to get the T7 strand. Moreover, because the primer DNA is $\bar{O}1$, input 1 and 2 are required to produce the DNA strand containing $\bar{O}3$ at its 3' end. Later, a NAND gate was experimentally implemented [30] and a more general logic gate was developed as a way to implement other logic functions [31]. It should be possible to implement oscillators [32], or more complex systems mimicking cell computation. As a step toward artificial cells, some of these systems were embedded in liposomes [27].

1.4.2 Genelets

Closer to the real cell information management, but still avoiding protein translation, the genelet MP is mimicking the *in vivo* regulation of genes as a the substrate of computation. Such an approach, developed by Kim and al [26], is not a complete mimic of what is happening in cell, but more an analogy. Here also, the outputs are RNA strands, generated from a DNA-engraved program thanks to a RNA polymerase. A RNase degrades the inputs and outputs over-time, sustaining the system out of equilibrium. The computation stops after exhaustion of the NTP fuel.

A genelet is defined as a duplex of DNA containing two domains: the output sequence and nicked promoter. The genelet can be off or on if the promoter sequence is, respectively, partially or fully double stranded (Fig.1.7, a). The activation of a genelet by the hybridization of the DNA activator on the OFF genelet starts the production of the output with the RNA polymerase. The inhibition mechanism occurs when the promoter sequence binds to a complementary RNA designed for that purpose. As for RTRACS, the system is kept out of

equilibrium by degrading the RNAs formed with a RNase (Fig.1.7, a).

The network is built on cascading genelets together thanks to a rational design of the sequences. Designing two repressive modules, Kim and Winfree implemented a bistable switch circuit [26] and some oscillators [28] just by changing the topologies of the networks.

1.4.3 PEN DNA toolbox

The previous examples are inspired by *in vivo* processes. In that lineage, our group has developed another molecular programming scheme inspired by gene regulation networks, but even more abstracted than the two previous examples. We tried to simplify the operating system by removing the RNA steps. Indeed, the reactivity of DNA strands in particular their interaction with a DNA polymerase can be regulated with complementary DNA strands acting as triggers. Because there is only one polymer type in this system (DNA, no RNA) one needs to find a way to protect some of the

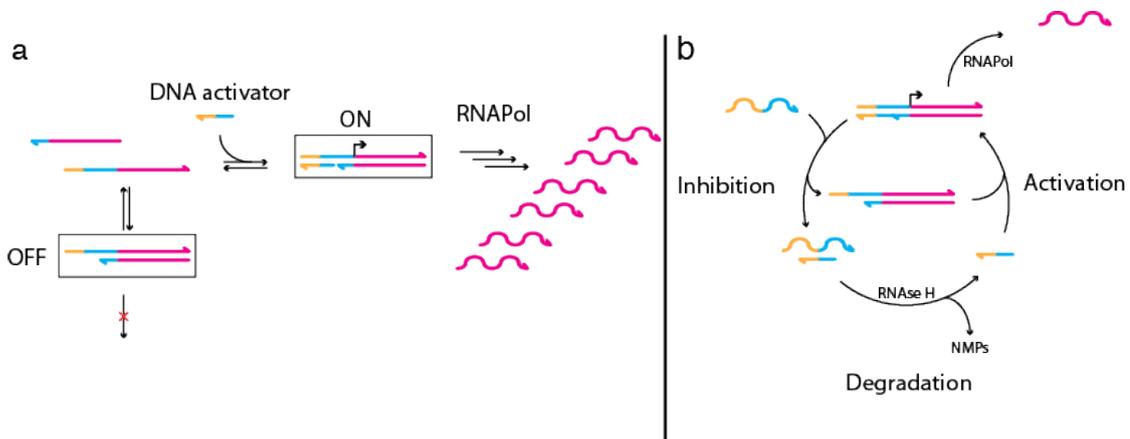


Figure 1.7: Genelet system. (a) General workflow: genelets are composed of short ds-DNA with a nicked T7 promoter (OFF). As soon as the promoter is active by hybridization of the DNA activator (ON) the RNA Polymerase generate RNA transcripts from the template. These RNA transcripts will connect the genelets within each others. (b) Elementary reactions for genelet assembly: activation corresponds to the hybridization of the complete promoter and RNA synthesis. Inhibition occurs when an output RNA from an other genelet binds the DNA activator strand, thus inhibiting its hybridization on the inactive genelet. Finally, the presence of RNase H -degrading only RNAs- keep the system out of equilibrium.

species, while letting others be dynamically degraded. As with the other programs, it is possible to cascade the reactions by a thorough design of the sequences involved.

This program, the Polymerase/Exonuclease/Nicking DNA toolbox (PEN DNA toolbox), uses a reduced set of only three DNA polymerizing-depolymerizing enzymes: a polymerase, a nicking enzyme and an exonuclease. Globally, the concerted activity of these three enzymes starts from deoxynucleotide triphosphates (dNTP) as substrates, that are converted to deoxynucleotide monophosphates (dNMP) through a polymerizing-depolymerizing process (direct conversion from dNTP to dNMP is kinetically blocked). It thus corresponds to a dissipative process going down a chemical potential; because our experiments are set in a closed tube (i.e. no mass transfer possible), the system will stop working once all the dNTPs are exhausted. The three basic blocks function the same way (Fig.1.8, a):

- first, an input strand transiently binds a template on its input site.

- second, the primer/template structure is recognized by a polymerase and elongated to a stable full duplex.

- the nicking enzyme then releases the input and output from the template: they are now able to diffuse and bind other templates.

In their free, single stranded form, inputs and outputs are recognized and degraded by the exonuclease. The exonuclease we use is a processive enzyme and does not lead to the accumulation of partially degraded intermediates, which would be detrimental to the reaction. The concentration of any species that is not actively produced by a template will therefore eventually decay; the concentration of a species involved in a positive feedback loop of sufficient strength will reach a non-zero steady-state.

Depending on the sequence of the output (for a given network), the reaction can be:

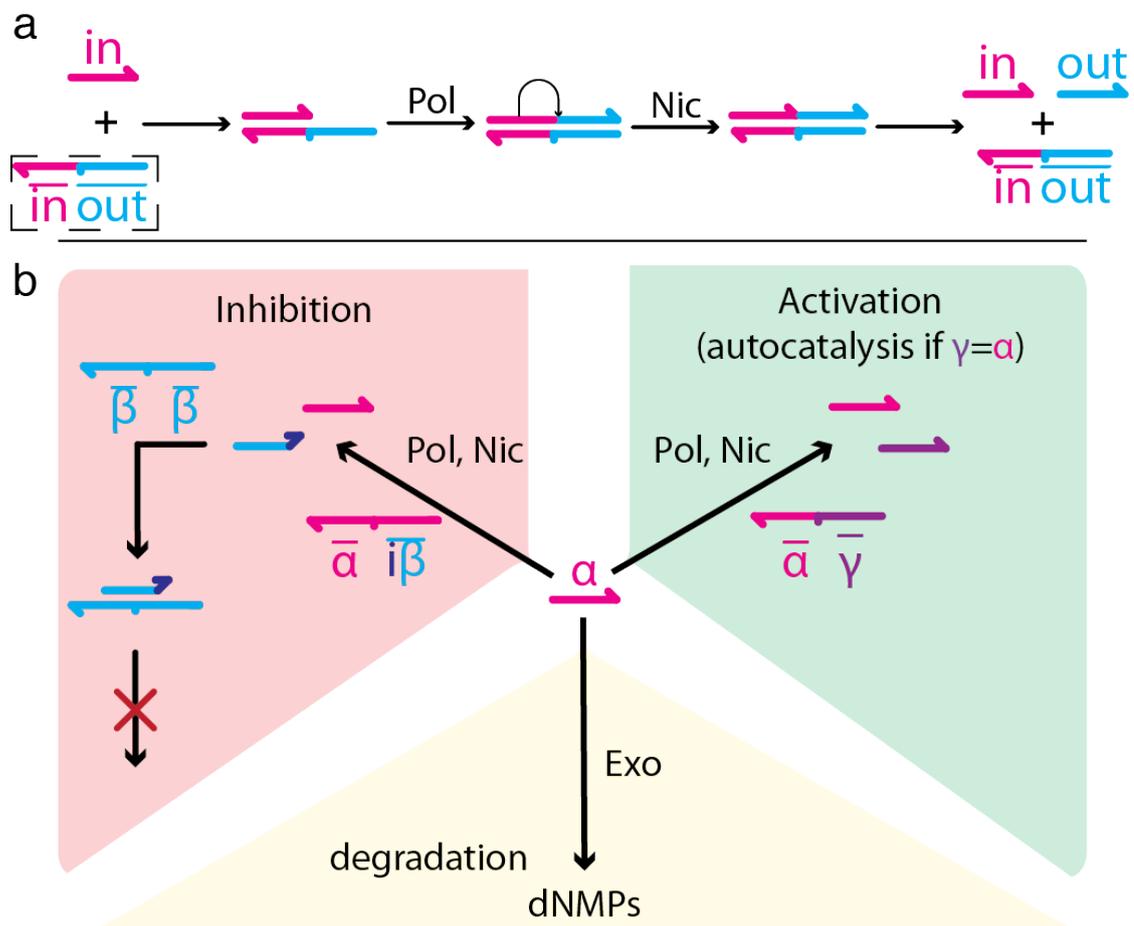


Figure 1.8: PEN-DNA toolbox: three cascable enzymatic reactions between DNA strands using three enzymes: polymerase, nickase, exonuclease. (a) General workflow: an input sequence hybridizes on a template (dashed box) before being processed by a polymerase. The resulting duplex shows the recognition site for the nicking enzyme and the nicking product (output) can interact with other templates in solution. Depending on the sequence of the output, the transition input-output can be an activation (the output is the input of another template, or autocatalysis if the input and output sequences are the same) or an inhibition (the output is a partial complementary strand to an activation template with a mismatch, preventing the template to be processed by any enzyme/input). Finally, the system remains dynamical thanks to the exonuclease which constantly degrades outputs and inputs (templates are protected from degradation using two to four terminal phosphorothioate backbone modifications).

- an activation (if the output of this node is the input of an other one, Fig.1.8, b)
- an autocatalysis (in the special case where input and output have the same sequence)
- or can correspond to the generation of an inhibitor when the output released is partially complementary to a target template but have some mismatches at the 3' end, hence do not trigger polymerization (Fig.1.10, a).

These inhibitor molecules are also dynamic because they are produced (by the action of the polymerase and nicking enzyme on other templates) and subject to degradation (by the exonuclease, when they are in single-strand, free-floating form). The modular construction of the system allows the designer to “wire” the templates together so that they control each other’s activity by exchanging small activators or inhibitors. This modularity is central to the building of complex systems; it allows the cascading of elementary modules into precise network topologies. The cascability is greatly facilitated by the fact that the DNA strands being exchanged are small (typically 10 to 20 bases long) and do not form complicated secondary structures. This range of length theoretically limits the number of species that can be mixed together before unwanted interactions start to occur (see below). Although for medium scale, biologically relevant networks, choosing DNA sequences while avoiding unwanted interactions is entirely feasible. With a correct selection of the sequences, the product of any reaction can be used as the activator of any other module. For that, it is enough to design (and order to a DNA-synthesizing company) the corresponding templates with the *ad hoc* sequences and modifications. In a similar way, it is always possible to define an inhibiting module targeting any activation template. This modularity opens the road to the rational molecular encoding of a variety of dynamic behaviors.

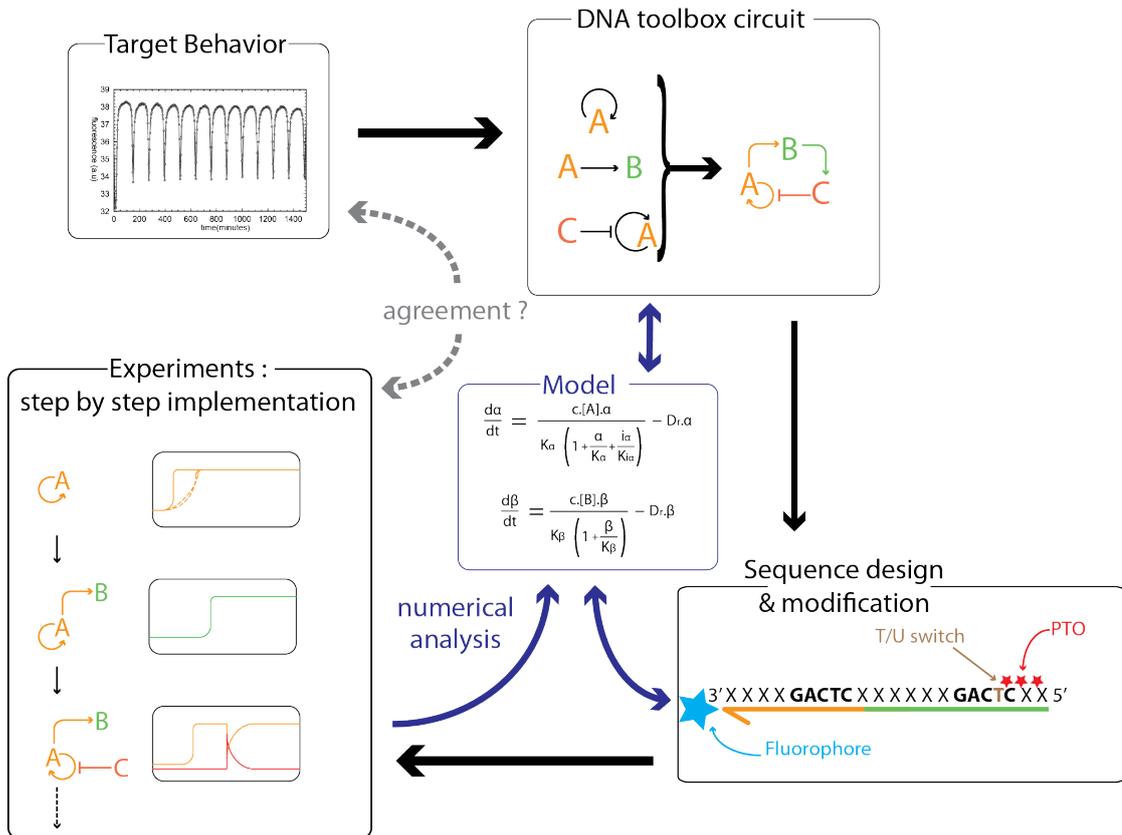


Figure 1.9: General workflow of PEN-DNA toolbox-based network implementation. First, a putative network is drawn using PEN-DNA toolbox module, supposingly with the target behavior. From this network, a model of the system can be built and the required DNA species can be designed. Once the sequences have been synthesized, molecular implementation is carried out step by step, starting with one or two nodes, then adding the nodes one by one, adjusting the experimental conditions in the same time. The intermediate experimental results provide measured parameters that can be fed into the mathematical model, which will be used to predict experimental results and optimize the conditions until the molecular network achieves the targeted behavior.

The sequence of the template is the concatenation of the input (trigger) and the output's (product) complementary sequences (in the order 5'-output complementary sequence-input complementary sequence-3'). We design inhibitors so that they hybridize to the targeted template with a dissociation constant about two orders of magnitude higher than that of the input strand, enough to outcompete them, but not too high in order to maintain a reasonable rate of spontaneous de-hybridizing. Inhibitors should not be sensitive to nicking and hence their design will depend on the geometric peculiarities of the nicking enzyme that is used. For example, whether the nickase nicks within, before or after the recognition site will have an impact on the design process. In the following we discuss the case of the nicking enzyme *Nt.bstNBI*, which recognizes a 5 base site and nicks 4 bases downstream. While different possibilities are available, we generally work with so called 7-6 inhibitors, which, among their 15 bases, have 13 bases complementary to the targeted activating template, with 7 bases binding to its 3' input region and 6 to its 5' output region (Fig.1.10, a). This distribution between the input part and the output part prevents the formation of a nickase recognition site on the inhibiting sequence. The two unpaired bases at the 3' end prevent the extension of the template-inhibitor partial duplex by the polymerase. Other designs for inhibitors (8-6, 8-5, 6-7 etc) are possible. However, too high coverage of the input sequence will lead to an unwanted recognition site for the nicking enzyme, or a too high stability on the target (leading to an irreversible inhibition) while too low coverage will not result in a high enough affinity for the inhibitor compared to the input.

The input sequences correspond to the triggers that will prime any DNA polymerization reaction of the DNA toolbox. Their length ranges from 10 to 15 bases, with the upper limit set by the working temperature (they have to dynamically hybridize and de-hybridize) and the lower limit by the sequence requirement of the nicking enzyme (a duplex that does not extend at least one bp beyond the

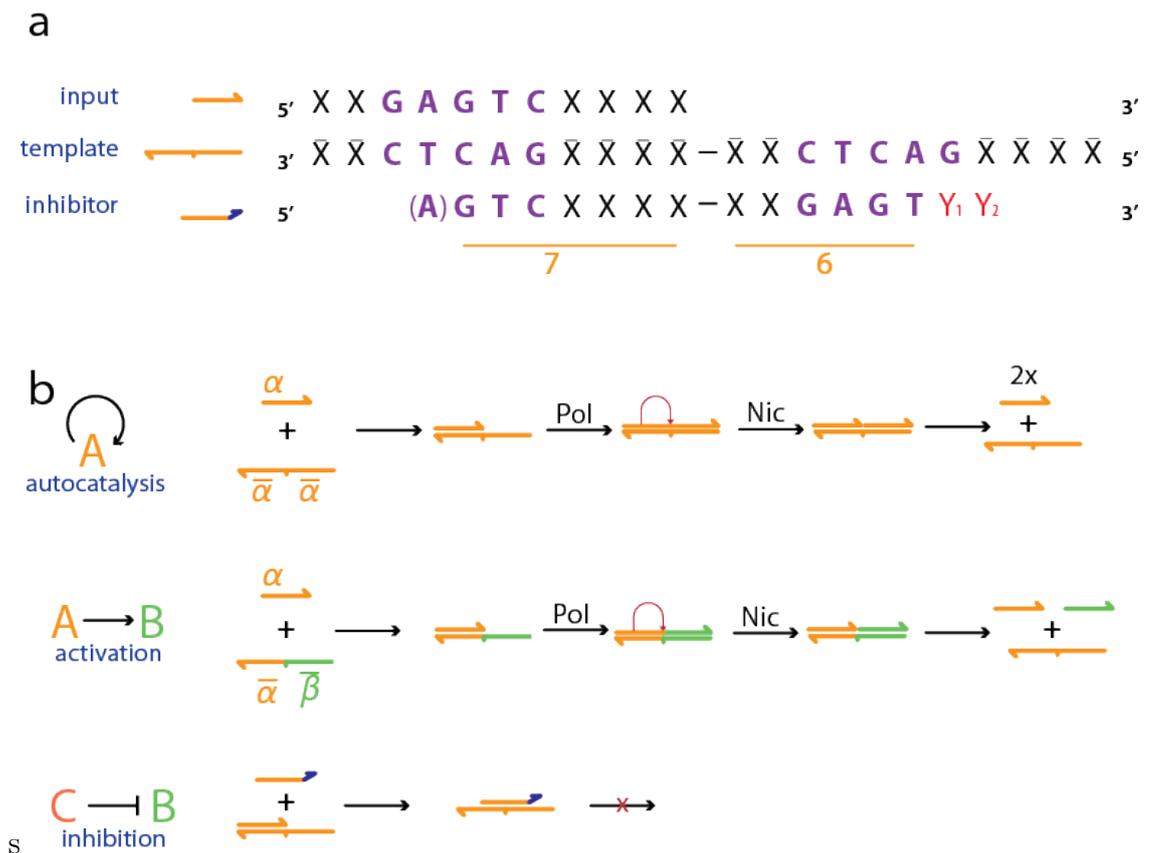


Figure 1.10: Typical set of DNA strands necessary for a network when using Nt.BstNBI as the nicking enzyme and their biochemistry within the DNA toolbox. (a) The Nt.BstNBI recognition site is shown in violet. The template illustrated here is an autocatalyst triggered by α , and producing itself. Regarding the 7-6 inhibitor, two mismatches are introduced at the 3' end (red). The nomenclature comes from its position on the template: 7 bases binding to the input position and 6 bases to the output position.

recognition site is poorly processed by the nicking enzyme Nt.BstNBI, which cuts 4 bases downstream of its 5 base cognate sequence; hence inputs should have at least 10 bases). We typically use templates that are 11 bases long, in which 5 bases -corresponding to the recognition site- are fixed. Considering two DNA species to be distinct if at least two of their bases are different, we can design more than 1000 (4^5) different sequences from which a number should be discarded, using the following filters:

- the sequence (or the dual repeat) should not contain a parasitic nicking enzyme recognition site,
- the melting temperature of the sequence should be neither too high nor too low and typically close to the experimental temperature (i.e. should not be composed of too many C/G nor A/T),
- the sequence should not allow stable secondary structures to form (this could lead, among other uncontrolled behaviors, to self-triggering in the case of a self-fold with a matched 3' end) or unwanted interactions between two input species (primer dimerization). Several softwares can be used to predict the formation of secondary structures, such as the NUPACK software suite [33].

In fact, adjusting the melting temperature of the various DNA species involved in the network is one way to tune the circuit parameters at a local level (together with changing the templates' concentrations). Therefore, those adjustments can be very important for the proper function of the network. However, it has been reported that the performance of a given primer in terms of amplification is not simply related to its thermodynamic stability (Fig.1.11). The sequence dependence of exponential DNA amplification [34] is still not well understood, but has been the subject of a recent study [31], in which

the authors characterized the performance of about 400 autocatalytic templates. They notably observed that GA or AG dimer-rich sequences were poorly performing. The rules proposed by Qian et al. have to be considered for future design of autocatalytic templates [?].

For each design of a new activator sequence, we experimentally evaluate its relative performance by conducting the following procedure:

- We first perform a simple autocatalytic amplification (Fig.1.11, left), using a template that bears the new sequence both as input and output. Since we only want to check the amplification potential, the mixture contains the polymerase and nicking enzyme, but no exonuclease.
- For all sequences that do yield a reasonable exponential amplification, we conduct an evaluation of their dNTP consumption rate in steady state (therefore in a mixture including the exonuclease) in the presence of a limited amount of dNTPs, e.g. 40 μ M (Fig.1.11, right). To compare the autocatalytic modules, it is possible to calculate their normalized plateau time, Pt , relative to the template X:

$$Pt = \frac{N \cdot T}{N_x \cdot T_x}$$

with N the number of limiting dNTPs per polymerized oligonucleotide and T the plateau time. This number provides a global estimation of the ability of one particular sequence to display a fast turnover (which in turn is dependent on many factors, such as the sequence's stability, but also its affinity for the various enzymes). Plateau times are listed in Fig.1.11. In the present case, it might be wise to avoid working with template d (which is slow) or r (which shows a tilted plateau, for unknown reasons).

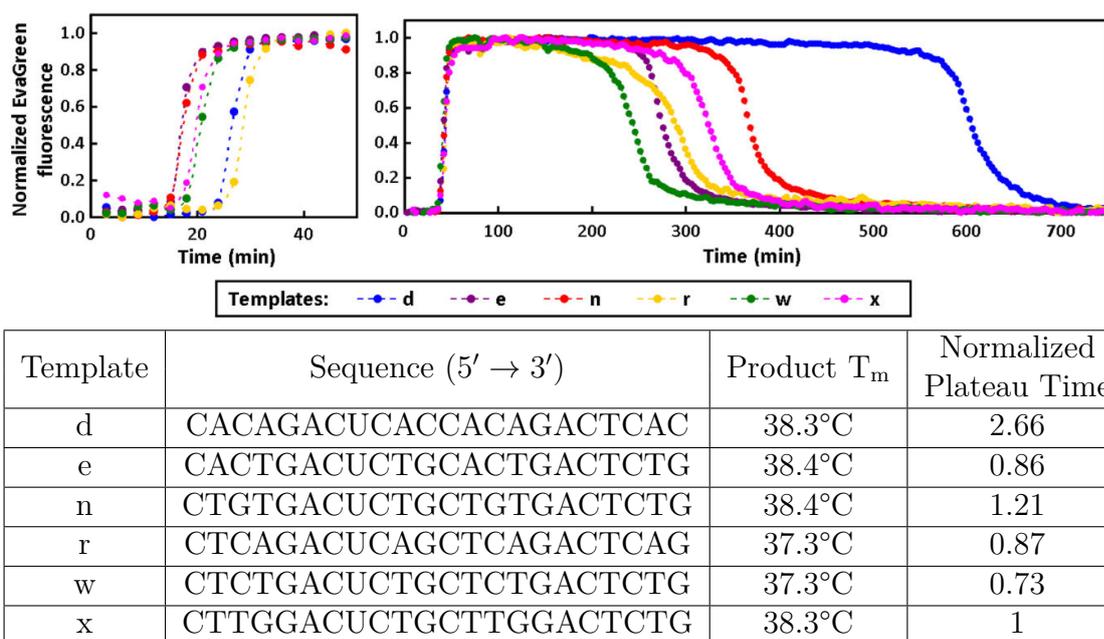


Figure 1.11: Comparing the performance of various autocatalytic templates by measuring the turnover of dNTPs. The left panel shows the signal amplification phase using autocatalytic templates d, e, n, r, w and x. The right panel shows typical plateaus in the fluorescence signal indicating that the production and destruction of DNA species by the module has reached a dynamic equilibrium, which ends when the dNTPs are exhausted and the signal drops back to its initial value. The duration of the plateau gives an estimation of the ability of the sequence to dissipate energy by fast enzymatic turnovers. The table lists the sequences of the templates (modifications are not shown), the T_m of their products and the normalized plateau time.

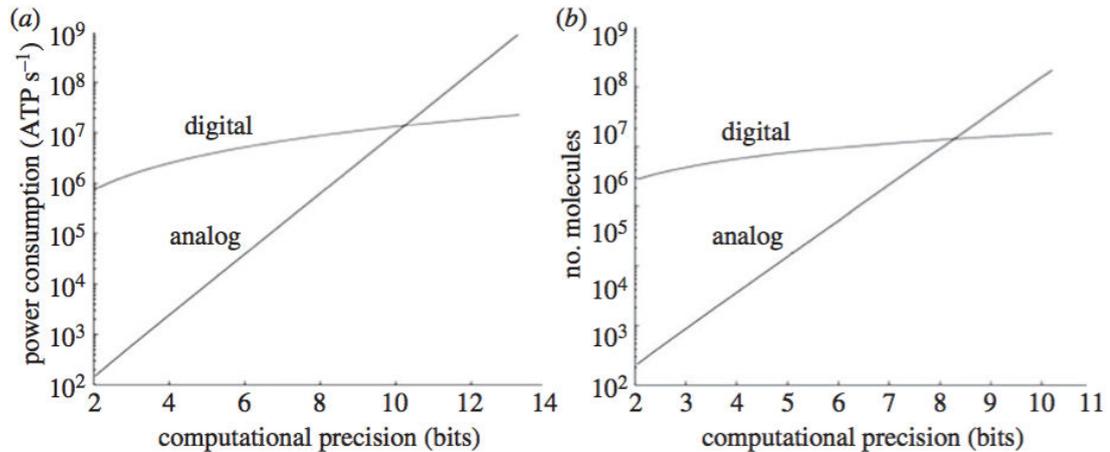


Figure 1.12: Analog versus digital computation in cells. (a) The figure shows the power costs for doing addition in cells with a genetic circuit. (b) The figure shows the molecular protein number required to perform an addition in cells with a genetic circuit. From Sarpeshkar 2014[38]

1.4.4 Computation and modeling

Despite taking inspiration from cellular regulation, initial proof of principle in artificial chemical reaction networks have targeted Boolean logic, with the examples of the AND gate in RTRACS and genelets, or the easy development of NOT gate in PEN DNA toolbox [35]. Such networks use complex biomolecules and circuits to output digital results interpreted as 0 or 1, that is, with a low informational content. It is however clear that in most cases cells are not computing digitally. For instance, the regulation of the concentration of many proteins is done on a continuous set between its minimum and maximum amount and not on a discrete one [36]. Accordingly the network mapping revealed by systems biology approaches, are reminiscent of neural networks architectures, rather than engineered electronic circuits. Sarpeshkar *et al.* studied in details the possibilities of analog computation in cell, in the context of synthetic biology circuits [37, 38].

They demonstrated for *Saccharomyces cerevisiae* that the power consumption cost of analog logic is significantly less than the one of

digital computation, so that the yeast would have to engage 100% of its power budget just to add two numbers at 10-bit precision (Fig.1.12a). On the other hand, when they compare the computational precision with the number of molecules involved, it appears that the increase of precision in computation in the case of digital addition requires several order of magnitude more proteins than in the analog version. Such high changes in the amount of energy or number of molecules becomes easily toxic for the cell itself (high number of copies of molecules can hinder signal pathway and a high metabolism enhances the apparition of reactive oxygen species).

It is therefore very tempting to also try to build analog circuits using artificial molecular programming approaches [39, 40]. Yet analog circuits are more delicate to conceive than their digital counterparts. Besides circuit topology, the dynamic is determined by an additional layer of continuous parameters: concentrations of species, activities of enzymes, temperature... Continuous changes to these parameters can lead to discontinuity in function, a counter-intuitive phenomenon known as bifurcation. How can one then map the functional landscape of analog circuits to their design parameters? In other words, how can we see their bifurcation diagram (Fig.1.13)? Manual tuning of parameters -discrete, hypothesis-driven and bulky- is poorly suited to the continuous, black-box and gigantic nature of the space that needs to be explored. To understand this better, we analyze below the dynamics a small artificial molecular network [41].

Let us consider a PEN DNA toolbox network in which two autocatalytic loops are linked together (the second autocatalytic compound uses the first as a precursor) and both outputs are being degraded overtime by the exonuclease. This in fact implements a predator-prey system [41] as it has been described by Lotka and Volterra almost a century ago [42, 43]. The aim is to map the bifurcation diagram of the network according to three parameters.

In the molecular implementation of this network, preys (N) and predators (P) are small ssDNA strands and obey three rules:

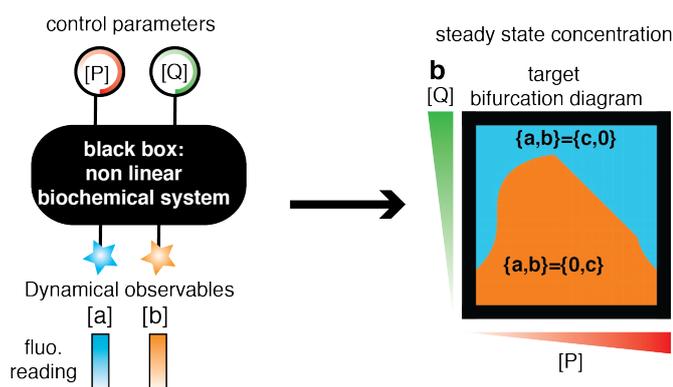


Figure 1.13: Mapping the dynamical behavior of a nonlinear chemical system: (left) the dynamics of a nonlinear biochemical system is parameterized by the concentrations of some control species (catalysts or cofactors such as enzymes, DNA strands, salts). Experimentally, it is necessary to create many systems with different parameter sets and infer their dynamics (for example by fluorescent reading -steady state levels or time traces- of some dynamical observables) to give the bifurcation diagram of the system (right).

- prey growth (an autocatalytic production of the prey strand on a grass template (G)),
- predation (elongation of the prey strand into the predator strand)
- and natural decay (degradation of ssDNA by exonuclease. G is protected from the degradation).

The main interactions are described in Fig.1.14.

In the original publication, the authors develop the mathematical model using a kinetic description of the three biochemical reactions, written as follows (with some modifications):

- the prey growth is a combo process comprising elongation and nicking steps (Fig.1.14, 1-5). It is reasonable to consider that the polymerase works in first order regime, therefore the rate-limiting step at low substrate concentration. However, the following reaction (the nicking event) will impose an upper limit on the maximum rate of the process, so it should be written as a Michaelis-Menten-like expression. Because the nicking enzyme has a low turnover rate, the prey production is mostly limited by the nicking step, the previous ones being in the first order

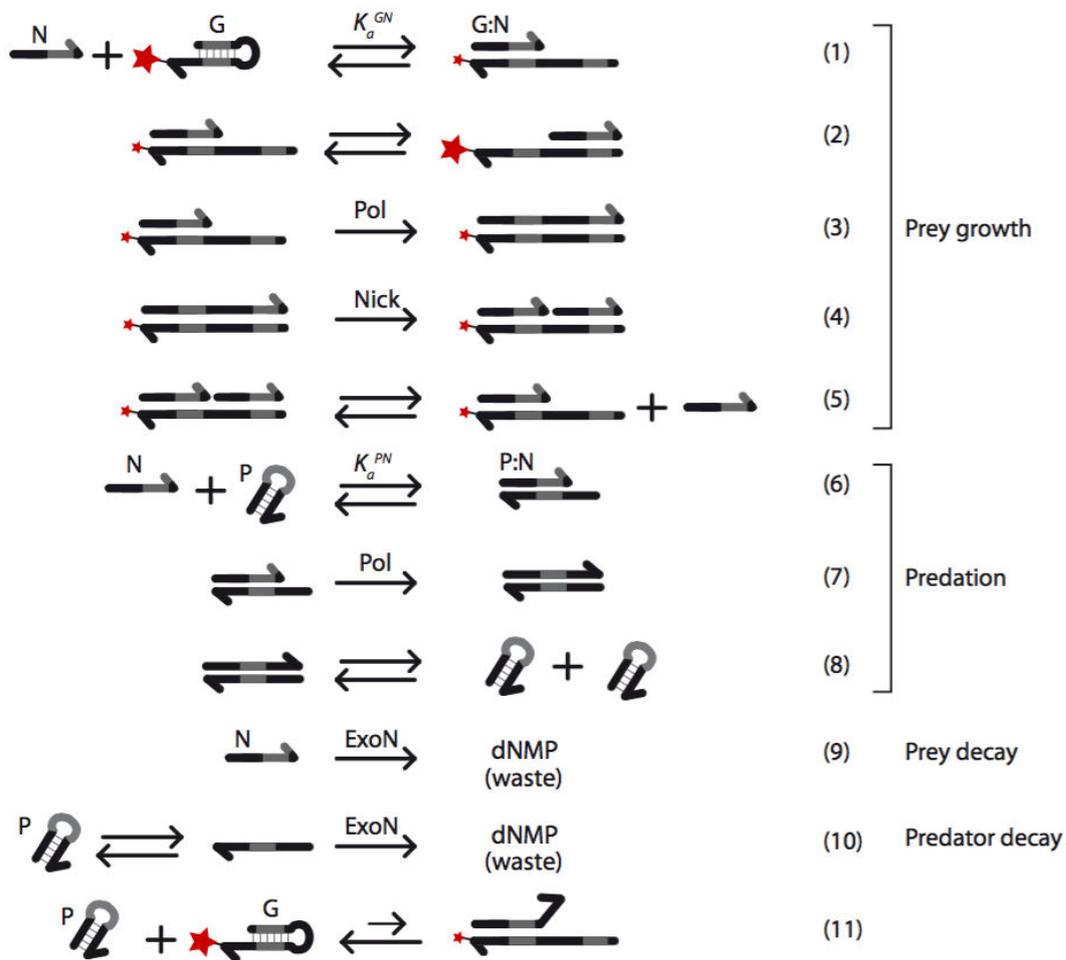


Figure 1.14: Schematic representation of the mechanism of the DNA predator-prey system. Gray and black parts in the strand sequences show how the predator's palindromic sequence is reflected on the prey and grass strands. The star represents the fluorescent dye attached to the 3' end of the template and the modulation of the intensity of the dye depending on the hybridization state of the template (single strand state has a higher fluorescence, whereas it is partially quenched in double strand form). Reaction (11) shows that the predator can also have a contribution to the fluorescence shift in the prey channel. Reproduced from Fujii & Rondelez, 2013[41]

regime. The expression of prey growth becomes:

$$\varphi_{N \rightarrow N} = k \cdot pol \cdot \frac{G : N}{K + pol \cdot G : N}$$

where G:N is the concentration of prey-template duplex. Then, considering the fast equilibrium in dehybridization, we can express G:N as the product of the free template and preys (G and N), and since the binding constant of such complexes are small and N and G are mostly free, the resulting equation gives:

$$\varphi_{N \rightarrow N} = k_1 \cdot pol \cdot \frac{G \cdot N}{1 + b \cdot pol \cdot G \cdot N}, \text{ with } k_1 = K_a^{GN} \frac{k}{K} \text{ and } b = \frac{K_a^{GN}}{K}$$

- the predation is a simple elongation (Fig.1.14, 6-8) of the duplex N:P, whose binding constant is low enough for the duplex to be mostly dissociated at working temperature. Assuming that the extension of the duplex is a first-order and rate-limiting step for the predation. We write: $\varphi_{N \rightarrow P} = k' \cdot pol \cdot P : N$. Thus, applying the same assumptions on binding constants of the duplexes, we have:

$$\varphi_{N \rightarrow P} = k_2 \cdot pol \cdot P \cdot N$$

- finally, the degradation is solely due to the action of the processive exonuclease on both prey and predator strands (Fig.1.14, 8). This one-step enzymatic reaction can be described using Michaelis-Menten kinetics:

$$\frac{dX}{dt} = rec \cdot k_{cat,x} \frac{X}{K_{m,X} + X} = V_m \frac{X}{K_{m,N} + X}$$

Since the same enzyme processes both strands, there is competition between preys and predator for the enzymatic activity. Considering also that predators have a lower K_m et V_{max} , the equations become:

$$\varphi_{N \rightarrow \emptyset} = rec \cdot k_N \frac{N}{1 + \frac{P}{K_{m,P}}}, \text{ with } k_N = \frac{k_{cat,N}}{K_{m,N}}$$

$$\varphi_{P \rightarrow \emptyset} = rec \cdot k_P \frac{P}{1 + \frac{P}{K_{m,P}}}, \text{ with } k_P = \frac{k_{cat,P}}{K_{m,P}}$$

We can now write the complete equations as two coupled ordinary differential equations (ODE) by summing the contributions of prey growth, predation and degradations:

$$\frac{dn}{d\tau} = \frac{pol \cdot tem \cdot n}{1 + \beta \cdot pol \cdot tem \cdot n} - pol \cdot p \cdot n - \lambda \cdot exo \frac{n}{1+p}$$

$$\frac{dp}{d\tau} = pol \cdot p \cdot n - exo \frac{p}{1+p}$$

with the dimensionless parameters $\tau = \frac{t}{t_c} = t \cdot k_2 \cdot pol \cdot K_{m,P}$, $p = \frac{P}{K_{m,P}}$, $n = \frac{N}{K_{m,N}}$, $\lambda = \frac{k_n}{k_P}$, $g = \frac{G}{G_0} = \frac{k_1 \cdot G}{k_2 \cdot K_{m,P}}$ and $\beta = \frac{b \cdot k_2 \cdot K_{m,P}^2}{k_1}$ and pol , tem and exo the dimensionless concentrations of three enzymes.

Because of the non-linearities in prey growth and decay, the analytical treatment is not immediate in this case, but if we consider the saturation on all enzymes (i.e. $\beta \cdot pol \cdot tem < 1$), the ODE describing the prey derivative can be re-written as:

$$\dot{n} = pol \cdot tem \cdot n(1 - \beta \cdot pol \cdot tem \cdot n) - pol \cdot p \cdot n - \lambda \cdot \delta \frac{n}{1+p}$$

To assess the number of solutions (i.e behaviors) of this system and their stabilities, we focus on the equilibrium points, meaning the points where both derivative are equal to zero (i.e. the position in phase space where there is no temporal evolution) and the eigenvalues of the Jacobian, the linearization of the system near these equilibrium points. According to the sign of the real part of the eigenvalues, the corresponding point is declared stable or unstable. In this particular system, setting \dot{n} and \dot{p} to 0 reveals four points. Looking at the eigenvalues of the community matrix around these points, we found that three of them have a stability domain:

- the first point $(0, 0)$ corresponds to the extinction of both species, stable for $\lambda \cdot exo > tem \cdot pol$

- the second point $(\frac{tem \cdot pol - exo \cdot \lambda}{\beta \cdot pol^2 \cdot tem^2}, 0)$ shows the case of the extinction of predators and the stable production of preys. This point is stable for $\lambda \cdot exo < tem \cdot pol$ and $\lambda \cdot exo > tem \cdot pol(exo \cdot tem \cdot pol - 1)$
- the third one $(\frac{1+tem-\frac{\sqrt{\Delta}}{pol}}{2(pol^2 \cdot tem^2)}, -1+tem-\frac{\sqrt{\Delta}}{2 \cdot pol})$, where both prey and predator coexist is never stable over its domain of existence, with

$$\Delta = \sqrt{(pol - pol \cdot tem)^2 - 4pol(-pol \cdot tem + exo \cdot tem^2\beta + exo \cdot \lambda)}$$

- the fourth, $(\frac{1+tem-\frac{\sqrt{\Delta}}{pol}}{2(pol^2 \cdot tem^2)}, -1+tem+\frac{\sqrt{\Delta}}{2 \cdot pol})$ corresponds also to a coexistence of both preys and predators species. This solution has an unstable region (corresponding to the oscillatory area) and a stable region (corresponding to damping oscillations), both defined by analytical expressions.

Using those results, it is possible to compute the theoretical bifurcation diagram. The four solution's domains are calculated and the stable parts are superimposed to reveal the diagram, as shown in Fig.4.18.

The linear stability analysis gives a compact description of the experimental dynamics of the *mathematical* system. Moreover, we will see in Chapter 4 that the diagram in Fig.4.18 is quite close to the one obtained experimentally. We distinguish 3 main domains (Fig.4.18, c):

- at low productivity (low pol and high exo), we find the extinction region: no species can survive
- at high productivity (high pol and low exo) we define the stable coexistence. Oscillations occurs at intermediate pol.
- finally, an Hopf bifurcation delineates the region of stable limit cycles.

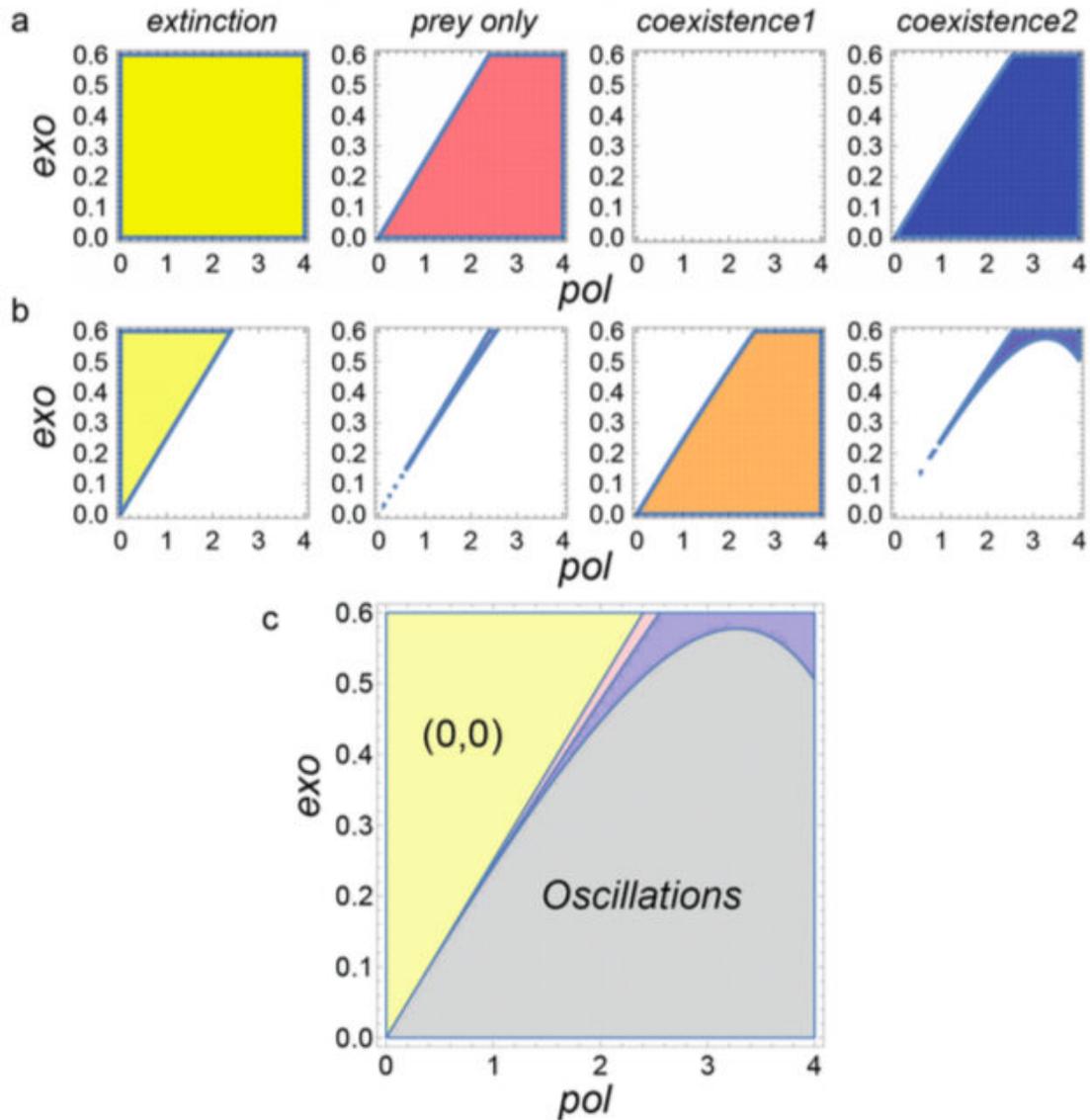


Figure 1.15: Bifurcation diagram of the 2-dimensional model obtained by linear stability analysis. (a) regions of existence of positive solutions. (b) Stability of solutions. (c) Bifurcation diagram showing the Hopf bifurcation around the oscillatory region (gray). Yellow area corresponds to the extinction regions ($P=N=0$), pink area to prey-only steady state and blue area to stable coexistence between prey and predators (after some oscillatory transients).

However, we must not forget that the mathematical analysis uses a highly idealized description of the actual chemistry running in the tube. For example, it considers only two dynamic species, the prey and the predator, whereas at least 6 are involved (predator, prey, grass template, polymerase, exonuclease and nicking enzyme) and much more if we would consider the individual steps of the enzymes (for example, each intermediate during the polymerization of successive dNTP on a template), are actually present. Therefore the problem is double: on the one hand, can we trust linear stability analysis to provide us with an actual description of the system's functional capabilities? It would be possible to refine the model but we would have then to resort to numerical predictions and an increasing number of parameters.

On the other hand, the challenge in realizing an experimental diagram, hence free oneself from assumptions, is to set the mesh of the diagram as fine as possible, without consuming too much chemicals, or using too much time. It is quite easy to control the reaction volumes and amounts of reagents while working with volumes around $10 \mu L$ but a simple 10×10 grid would use 1 mL of reaction medium and would still be rather crude in term of spatial resolution. Here, the use of microfluidic to generate picoliter sized micro-reactors with unique combination of parameters reduces the cost since we can have more than 10000 points for less than $100 \mu L$ engaged. Moreover this can be done in only a couple of hours.

Chapter 2

Droplet-based parameter scanning

2.1 Introduction

Several advances occurred recently in artificial DNA reaction networks, a blossoming branch of molecular programming which focuses on programming artificial DNA circuits. Such systems can achieve complex tasks, as signal processing, diagnosis or computation, while displaying various behaviors such as boolean logic, multistability or oscillations [22, 26, 44, 39, 45, 25].

Despite their apparent disparity in term of size, molecular substrate or function, it appears that networks share some recurrent patterns, a topology that encodes crucial information-processing tasks [46, 47]. Alon and his group studied thoroughly the importance and recurrence of meaningful, sub-network topologies. Considering a triad of connected effectors, they targeted all possibilities for this triad and screened networks from very different fields: sensory transcription networks for gene expression in E.Coli, yeast, signal transduction pathways in different organisms, links between webpages on the world wide web... Considering each superfamily of networks, some architectures take precedence over the species considered [46].

These networks process an initial set of reagents (inputs) through a cascade of chemical reactions to yield a product (the output), whose

state encodes the result of a computation on the inputs. In order to orchestrate a consistent response to the input, the reactivity of each node in the network must be finely adjusted. Typically, it should not be so high as to produce unwarranted output (leak) in absence of appropriate input, nor should it be so low that downstream nodes can barely distinguish the output of the node from the ambient noise. Yet in nonlinear networks, slightly tuning a few parameters (concentration, pH, temperature...) can produce a completely different output from the same input. In other words, if we visualize a circuit as a point in a parameter space, its function (the output it produces) can be very sensitive on its position. Mathematically, we say that nonlinear circuits bifurcate: their long term dynamic can abruptly veer in response to minute change in parameters. Bifurcation is what makes nonlinear systems so rich from an intellectual and practical point of view. We have several options to understand in a deeper way the influence of a parameter on the computed output. First, one can resort to simulations and establish a model of the network based on known kinetics and thermodynamic of the molecular species. The second option is simply empirical, brute force exploration of the parameter space by performing a large number of experiments with different parameters. Obviously, those approaches are complementary and mutually nurturing: an approximate model will help us to initially target an interesting network structure and a particular zone of the parameters space, and in turn feedbacks from experiment results will inform and improve the initial model.

An example is shown in Fig.2.1, where models and experiments were used to chart the behavior of a bistable switch. In this experiment, the concentration of two species (two autocatalytic templates) were varied together and the resulting dynamic recorded. However, due to the very limited resolution of the experimental diagram (only 41 points), bifurcation frontiers are hardly distinguishable. We would like to sample many more points so as to visualize precisely the bifurcation of a DNA circuits and understand better its dynamic. How

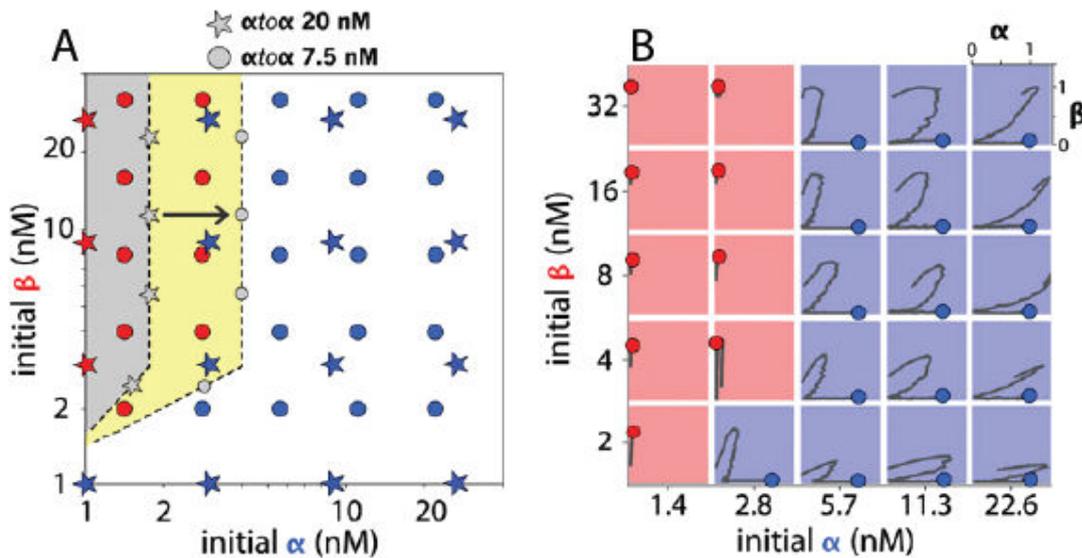


Figure 2.1: Theoretical (A) and experimental (B) bifurcation diagrams of a bistable switch (as seen in Chapter 3), in which two steady-states are possible: A (yellow and gray) or B (white). The calculated attraction basins (α in grey, β in yellow) according to the initial conditions (α -reach or β -reach medium). Dots and stars stand for the experimental points. (B) Experimental trajectories in different combinations of initial parameters. For each trajectory, the axis correspond to the change of the parameters. After transients, the system stabilizes in either state A or B. From [45]

many points should we aim for? A rule of thumb is that the precision of sampling will be eventually limited by reading errors on the concentration of DNA species, which is typically 1-10% when concentration are read by fluorescence. Varying two concentration parameters by steps of 1% requires testing $100^2 = 10^4$ experimental conditions! It is two orders of magnitudes larger than what we can do realistically in a wet lab with classical techniques.

2.1.1 Compartmentalizing

Since we aim to test tens of thousands of experimental conditions, it is no more realistic to prepare DNA circuits by manually pipetting and mixing microliters volumes: it would take days and waste liters of reagents. The sheer scale of work and reagent needed forces us to

think about newer ways to reduce volumes and automatize preparation. Micro reactors -a concept that rose in the mid 80's- combine the ability to simultaneously generate a large number of vessels with different experimental conditions from a limited amount of reagents. One of the initial incentive was the possibility of synthesizing compounds using a minimal amount of reagents while drastically increasing the safety of dangerous or explosive reactions [48].

An immediate consequence of shrinking the dimensions of reactors from millimeter to micrometer is that the surface-to-volume ratio explodes accordingly. Indeed, when the open surface of $200\mu\text{L}$ of mixture contained in a classical PCR tube is around 140 mm^2 , a $50\mu\text{m}$ wide water-in-oil droplet shows an accessible surface of 0.031 mm^2 for a volume of 500 pL (i.e. a 4.10^5 times smaller volume). This may lead to problems if the reaction carried inside the compartment is sensitive to the presence of interfaces. Another, possibly more beneficial, consequence is that the heat exchange being a function of the surface of the compartment, the temperature of the droplet vary smoothly with the global temperature of the continuous phase, making the control of the temperature easier.

The appealing features of micro reactors (decreased consumptions of reagents, faster temperature control and higher throughput) stimulated the development of convenient methods to generate microreactors [49, 50, 51, 52][49, 50, 51, 52, 53]. Emulsions provide a simple and economic route to compartmentalization into microreactors: when two immiscible fluids (say oil and water) are properly mixed together, physical instabilities force one of the fluid to partition into droplets (the dispersed phase) suspended in the other fluid (the continuous phase).

The generation and management of emulsion can be done through microfluidics, which is essentially a field dedicated to miniaturized plumbing and fluidic manipulation. This comes from fine control of each hydrodynamic flows, thus controlling the formation (deformation and break up) of each individual droplet, structures the gener-

ation of monodisperse emulsion quite easy.

As powerful as they can be, these methods still have some issues. For instance, a high heat transfer value also enhance the surface chemistry and surface effect on the walls of the reactor, meaning that some local parasitic reactions might prevent the chemicals within to properly react (i.e. as they were designed to). Another drawback is the control of size and dispersion of the emulsion: a top-down approach (for instance vortexing a biphasic solution of continuous and dispersed phase) leads to high polydispersity in compartment size. Even if the generation of an emulsion by microfluidic devices provides a good control over size and uniformity of the droplet compared to other emulsification processes (vortexing, or membrane-based techniques), setting the parameters for such generation can be delicate, depending on the considered device.

Microfluidic droplets is a now mature field, and used for a wide range of applications, such as study of heat transfer [54], screening of the conditions for crystal growth [53], single molecule enzymology [55], rare mutation detection by PCR [56], chemical oscillating reactions [57].

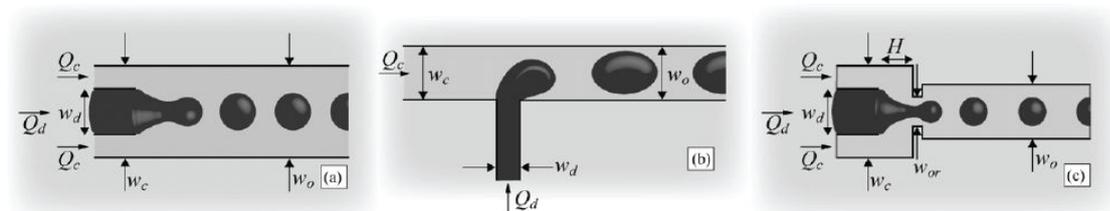


Figure 2.2: Three ways to generate droplets: co flowing(a), cross flowing(b) and flow focusing(c) streams. From [58]

Generation of droplets can be divided in two types of methods: active and passive, summarized in Table 2.1. In active methods, a physical process (beyond the mere flowing of a fluid inside microchannels) is employed to induce instabilities. Active methods comprise techniques such as spraying, atomization, membrane emulsification or inkjet printing. The simplest and oldest active method is vortex-

ing (simply shaking vigorously a solution of continuous and dispersed phase), yet its high polydispersity is undesirable for many applications. On the other hand, passive techniques merely rely on the channels' geometry to induce instabilities between two immiscible fluids and generate droplets. Since the size of droplets is fully determined by geometric and hydrodynamic parameters, which are held constant during generation, passive methods can achieve a very high monodispersity.

Three main geometries have been proposed for passive droplet formation: co-flowing [59], cross-flowing [60] or flow focusing [61]. Their mode of operation is similar: two (or more) microchannels carrying unmiscible phases meet at a junction, where hydrodynamic instabilities occur and break apart the dispersed phase. Their differences lay in the geometry of the junction and the relative flowrates, these parameters change the way the continuous phase will pinch the dispersed phase.

Technique	Active/passive	yield	monodispersity
spraying	Active	++	-
vortexing	Active	++	-
membrane emulsification	Active	+	+
inkjet printing	Active	-	++
flow-focusing	passive	++	++
cross-flowing	passive	+	++
co-flowing	passive	++	++

Table 2.1: Summary of techniques for droplet emulsion

To understand droplet formation, it is important to introduce some classical numbers from fluid mechanics (listed in Table 2.2). First of all, we introduce the well-known Reynolds number (Re) which compares inertial and viscous forces. Consider a fluid flowing with a velocity U_0 in a microchannel with a typical dimension L_0 (usually the width or height). Two forces apply on the fluid: the inertial force (whose magnitude is on the order of the gradient of kinetic energy $\rho U_0^2/L$), and the viscous force deriving from the shear

stress exerted by the walls on the fluid (on the order of $\mu U_0/L^2$). Taking the ratio of these forces yields the Reynolds number:

$$Re = \frac{\rho U_0 L_0}{\mu}$$

Turbulences empirically appears over $Re \sim 2000$ (at least in the classical duct experiments) and in aeronautic the Reynolds number of jet planes can easily exceed 10^6 . Conversely, the Reynolds number of microfluidic devices fall on the other side of the scale. Due to their micrometrical dimensions and moderate velocities, Re in microchips rarely exceeds 1. In other words viscosity quickly dampens spatial gradients of velocities at the micron scales. This low Re has an important consequence for design: the flows operate in the so-called Stoke regime. Convective forces can be completely neglected and the flow obeys a linear, time-reversible equation. The flow dynamic is then determined by two competing forces: the surface tension between continuous and dispersed phases and viscosity. Surface tension γ tends to minimize the interfacial area but viscous stress tends to stretch it. Intuitively the competition of these conflicting forces generates hydrodynamic instabilities and control the size of droplets. Their relative strength is measured by the Capillary number (Ca) defined as:

$$Ca = \frac{\mu U_0}{\gamma}$$

Finally, since we want to generate droplets with controlled contents, the question of diffusion versus convection arises. Let's consider two flows injected alongside the same channel. The required distance for the homogenization of the two flows is a function of the diffusion part and the velocity of the flows. The time needed for the diffusion of the particles across the channel is a function of the width of the channel and the diffusivity D so that $t \sim \frac{w^2}{D}$. In the same time, the particles will have moved down the channel, by a distance $L \sim \frac{U_0 w^2}{D}$. The ratio of these two components gives the Peclet

Number	Expression	Relation
Reynolds	$Re = \frac{\rho U_0 L_0}{\mu}$	inertial-viscous force
Capillary	$Ca = \frac{\mu U_0}{\gamma}$	viscous force-surface tension
Peclet	$Pe = \frac{U_0 L_0}{D}$	convection-diffusion

Table 2.2: Dimensionless key numbers in microfluidic with ρ the density, U_0 the velocity, L_0 critical length, μ the viscosity, γ the surface tension and D the diffusivity

number:

$$Pe = \frac{U_0 L_0}{D}$$

This number will be helpful to predict the effect of diffusion in the design. For instance, a ion with a diffusivity of $2000 \mu\text{m}^2 \cdot \text{s}^{-1}$ flowing at $100 \mu\text{m} \cdot \text{s}^{-1}$ in a $100 \mu\text{m}$ channel will need 5 channel widths to significantly diffuse. With these numbers, we can look in more details on the three passive break-up techniques.

2.1.1.1 Co-flowing

The system consists of two concentric channels. The inner channel contains the dispersed phase and pours into an outer channel containing the continuous phase (Fig.2.2, 1a). The dispersed phases is forced through a narrow orifice into the continuous phase; the shearing forces tear apart the dispersed phase and induce the droplet formation. Co-flowing exhibits two regimes of droplet generation : dripping (in which droplets are formed immediately next to the end of the capillary tip) and streaming (break up does not occur immediately but after the dispersed phase has flowed downstream over some distance). The switch between the pinching and jetting regime is driven by the ratio of flow velocities (a higher flowrate for the dispersed phase enhances jetting, a higher flowrate for the continuous phase favors dripping). The exact transition depends on Re (a viscous flow stabilizes the liquid-liquid interface and therefore the jet formation..), and Ca , since interfacial tension is the main force

applied on the droplet at the end of the tip. Co-flowing leads to highly monodisperse droplets, whose size is controlled by the diameter of the inner capillary. Although co-flowing was used for one of the first demonstration of microfluidic droplet generation, it is not in wide use now. One of the reason is that it is relatively delicate to fabricate two coaxial channels, microfluidic methods being more adapted at «extruding» 2D shapes that are projected by lithography.

2.1.1.2 Cross-flowing

Cross-flowing is one of the most common method to produce droplets, laying on a T-shaped device where the dispersed phase is crossed by the continuous phase (Fig.2.2, b). Here ratio between the pressures in the dispersed phase channel (Q_d) and the pressure in the continuous phase channel (Q_c), as well as the width of the channels (w_d w_c) are key parameters which determine the rate and size of droplets. A low or high $\frac{Q_d}{Q_c}$ means the invasion of one channel by the other, so that $\frac{Q_d}{Q_c} \sim 1$. The resulting shear forces elongate the dispersed channel and break up when $Ca \sim 1$. Generating droplets using T-junctions is easy and the monodispersity is good. The formation of droplet is possible over a wide range of flowrates, and the frequency is regular.

2.1.1.3 Flow-focusing

Flow focusing is by far the most common geometry to generate droplets. The dispersed phase flows in a central channel and meets the continuous phase at a symmetric junction (Fig.2.2, c) The two phases are then evacuated through a small constriction located downstream: the nozzle. The flow of the continuous phase Q_c is higher than the flow of the dispersed phase Q_d , creating viscous stresses on the dispersed phase inside the nozzle which breaks down the dispersed phase into droplets. This technique was pioneered by Anna *et al.* [61]. The targeted droplet size is not much more dependent on the ratio of $\frac{Q_c}{Q_d}$ than the nozzle's size w_{or} (Fig2.3). This flexibility

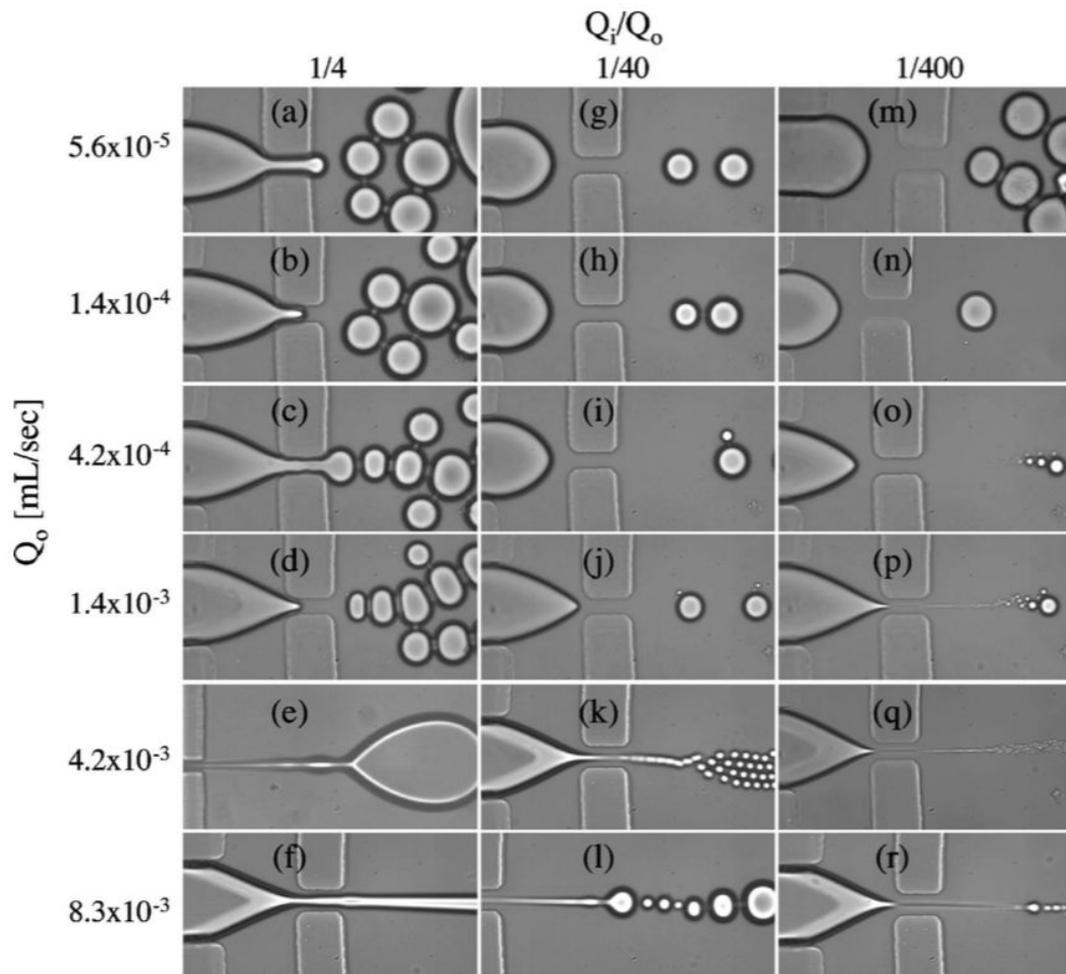


Figure 2.3: Range of droplet size according to the ratio of flowrates of the continuous phase (Q_i) and dispersed phase (Q_o). Reproduction from [61]

in range of sizes despite the fixed geometry of the device makes the flow-focusing method more practical than the two others.

2.1.2 Design

In this chapter we will describe the design, fabrication and control of a microfluidic chip used in high throughput droplets generation with different contents. Each droplet represents a point in the parameter space, and the whole droplets cover the parameter space with a fine mesh for exhaustive experimental mapping of parameter space.

The chip (Fig.2.4) comprises 4 identical aqueous inlets and one oil inlet. The design shows three parts. First, the inlet area in which a filtration system designed to prevent clogging. The filters consist of an array of tightly packed pillars(Fig.2.4, blue) whose spacing was chosen to match that of the constriction ($12.5 \mu\text{m}$) of the nozzle, and block dust or fragments of PDMS torn off during processing(Fig.2.5).

We inserted coil-shaped resistances in the oil and aqueous channels prior to the flow-focusing junction (Fig.2.4, pink) in order to mitigate invasion from other channels and place the working pressure in the range accessible by our controller (see below for details).

The four aqueous channels are merged in one channel before crossing the oil channel, thus focusing the flow at the entrance of the $50 \mu\text{m}$ nozzle, where the droplets are actually formed. The pre-injector channel upstream of the junction is $300 \mu\text{m}$ long. Its length was chosen long enough to visualize the co-flow of aqueous channels, and short enough to minimize their contact time. The serpentine channel after the flow-focusing (Fig.2.4, yellow) junction improves the mixing of droplets[62] and allows an appropriate coating of the droplet to prevent further merging [63]. Droplets are collected from the outlet in a pipetman tip. Closing the end of the tip with the thumb, we recover the emulsion and transfer it from the microfluidic device to a PCR tube while crafting the observation chamber.

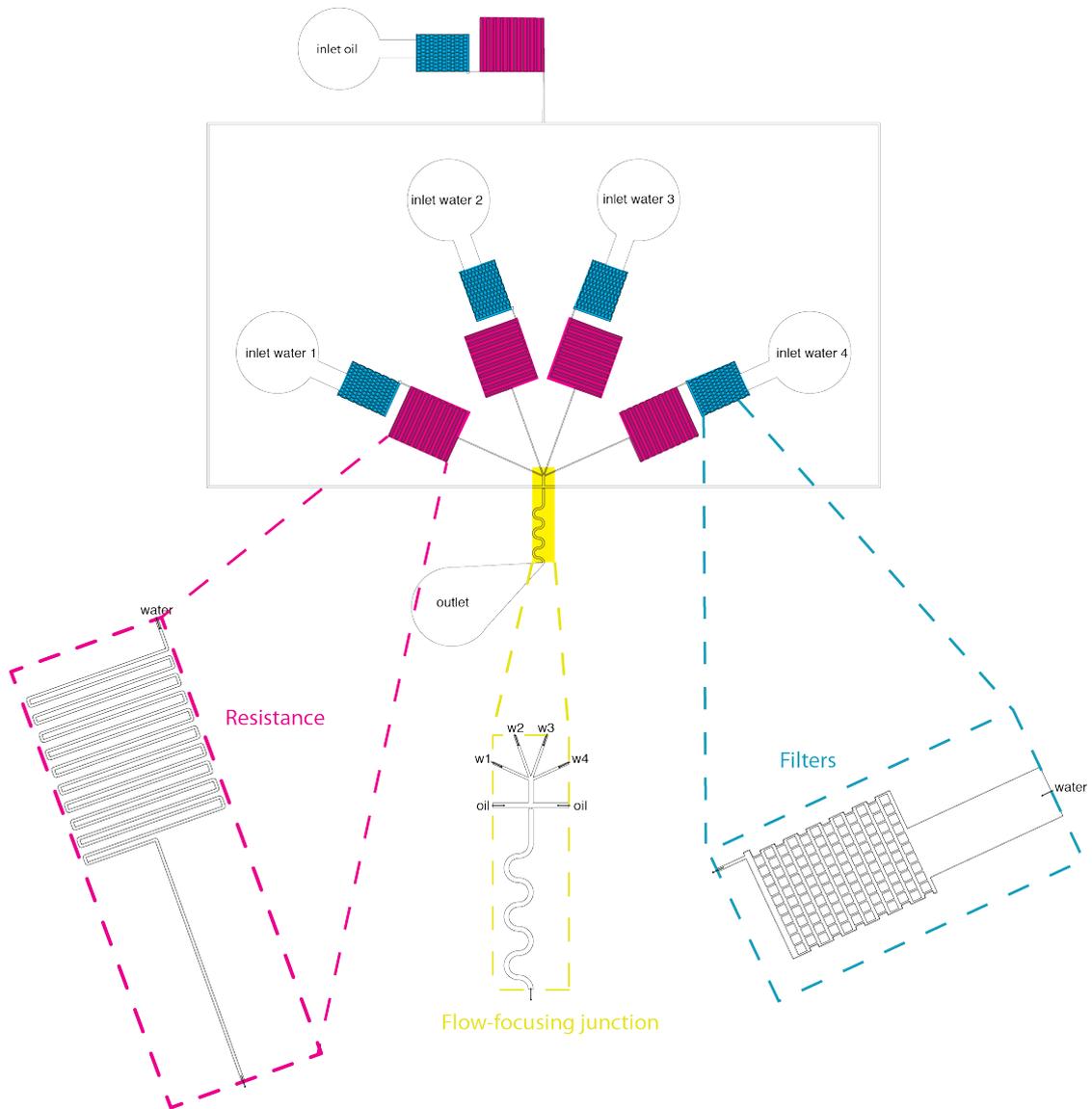


Figure 2.4: Microfluidic circuit developed for the bifurcation diagram mapping. 4 water inlets converge before the nozzle where the oil flow generate the droplets. Highlights: filter area (blue), pre-nozzle serpentine (pink) and post-nozzle serpentine (yellow).

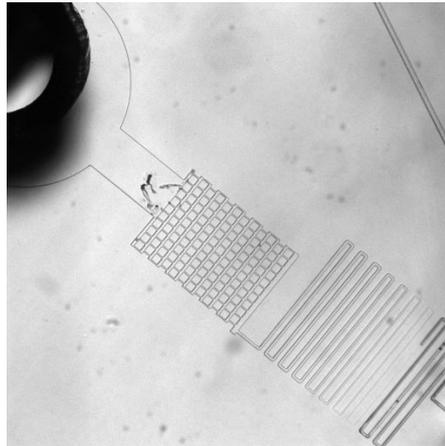


Figure 2.5: Dust from solution/tubing trapped by the filters.

Microfluidic chips can be actuated by two means: flow control or pressure control. In our experiments, we need to scan a large number of combinations of concentrations, and thus we need a quick actuation. We chose pressure control because it is noticeably faster and more responsive [64]. In an incompressible fluid, pressure jumps propagate instantly throughout the chip - allowing us to quickly modulate pressures near the junction by varying the pressures applied on the inlets. Indeed, for incompressible fluids ($\text{div}(v) = 0$) the pressure field satisfies an equation which does not contain any time-derivative: the pressure Poisson equation. In the laminar regime of microfluidic this equation reduces to the Laplace equation $\Delta P = 0$. The pressure is then an harmonic function whose value inside the chip is entirely determined by the chip's boundaries and the pressure applied on inlets and outlets. By contrast, jumps in flow rate do not propagate immediately -even in incompressible fluids- when channels are not rigid (which is known to be the case for PDMS and can be worsened by the elastic behavior of tubing or connections). Elastic channels act as fluidic capacitors that absorb jumps in flow rates and drastically slow down response times. But while pressure control allows for fast actuation, flow rates can become negative (backflows). Care need to be taken when programming the pressure profile to ensure

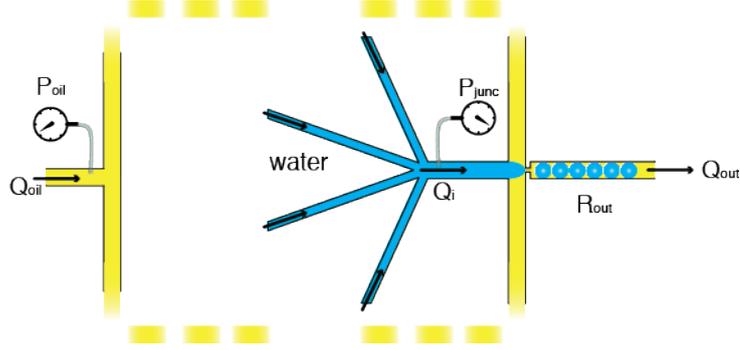


Figure 2.6: Preinjector and nozzle detail: microfluidic parameters. P_{oil} , Q_{oil} are the pressure and flow-rates applied to the oil channels, Q_i the flow-rate of the aqueous channel, P_{junc} the pressure in the preinjector, Q_{out} the flow-rate downstream the junction and R_{out} the corresponding resistance.

that the content of a channel does not invade significantly another (slight invasion of a channel is still needed to ensure that the species it carries sometime reach 0 nM).

2.1.2.1 Design of fluidic resistances in the chip

For sake of generality, we consider N aqueous channels (in our chip $N = 4$). All pressures are defined with respect to the atmospheric pressure P_o (pressure at the outlet). We neglect for this model the resistance of the preinjector, which is short. Let P_i be the pressure of the aqueous inlet i , Q_i its flow rate and R its hydrodynamic resistance (identical for all aqueous channels), whose value can be estimated for rectangular cross-section channels (*i.e.* $w > h$):

$$R \approx \frac{12\mu L}{h \cdot w^3 \left(1 - \frac{0.630w}{h}\right)}$$

where h is the height of the channel, w its width, L the total length and μ the viscosity of the fluid. It appears that the resistance is highly sensitive to the length of the channel and its width.

Similarly we define P_{oil} , Q_{oil} and R_{oil} the pressure, *total* flow rate and *equivalent* resistance of the two oil channels. Let P_{junc} be the pressure at the junction, Q_{out} the flow rate downstream of the junc-

tion and R_{out} the corresponding resistance (Fig2.6). The pressure of the aqueous channel P_i and oil channel P_{oil} are specified by the user through the pressure controller. On the other hand the flow rates (Q_i , Q_{out} and Q_{oil}) and the junction pressure P_{junc} are unknown variables. Those quantities are related linearly by

$$Q_i = \frac{(P_i - P_{junc})}{R}$$

$$Q_{oil} = \frac{(P_{oil} - P_{junc})}{R_{oil}}$$

$$Q_{out} = \frac{(P_{junc})}{R_{out}}$$

In order to avoid channel invasion, the flow rates Q_i must remain positive and thus $P_i \geq P_{junc}$ (similar conditions also hold for the oil channels). We now investigate the constrains these equations place on the chip. From the conservation of volume for an incompressible flow, we have

$$Q_{out} = \sum_i Q_i + Q_{oil}$$

and thus

$$\frac{(P_{junc})}{R_{out}} = \frac{1}{R} \sum_i (P_i - P_{junc}) + \frac{P_{oil} - P_{junc}}{R_{oil}}$$

we can extract the junction pressure

$$P_{junc} = \frac{\sum_i P_i + P_{oil} \frac{R}{R_{oil}}}{\frac{R}{R_{out}} + N + \frac{R}{R_{oil}}}$$

Note that since the sum of aqueous pressures $\sum_i P_i$ and the oil pressure P_{oil} are kept constant, P_{junc} also remains constants during droplet generation. In the regime $R \ll R_{out}, R_{oil}$, the junction pressure is simply the average of the pressures of aqueous inlet

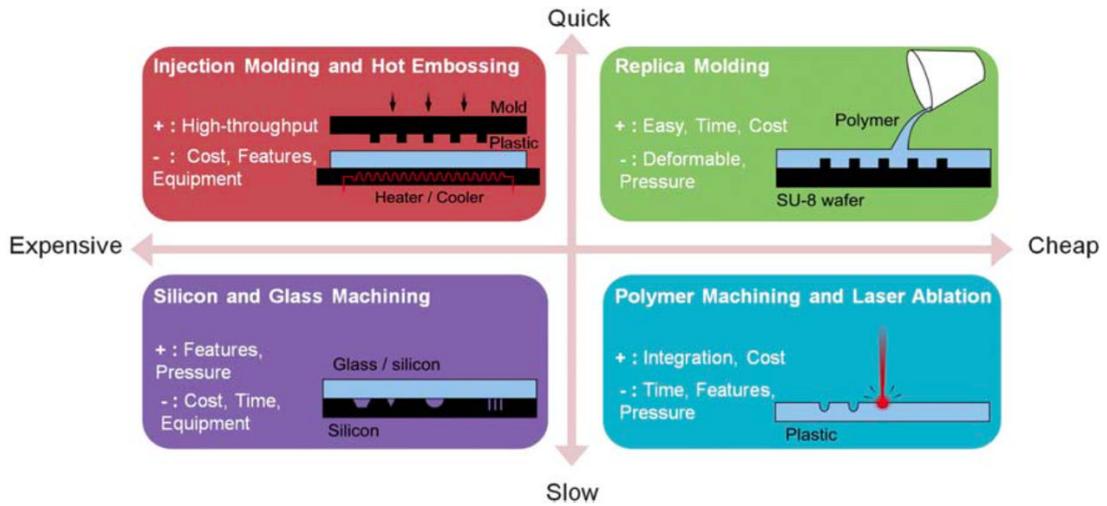


Figure 2.7: Four chip-prototyping methods, from [65].

$$P_{junc} = \frac{1}{N} \sum_i P_i$$

Since it is impossible for all the aqueous pressures to exceed their mean, at least one of the aqueous channel must be invaded. To avoid this, we dimensionalized the resistances of the oil channels R_{oil} and the aqueous channels R to be noticeably larger than the resistance of the chip after the junction R_{out} (we have $R_{oil} \simeq R$ and $R \simeq 80R_{out}$).

2.2 Microfluidic device fabrication

Thanks to progress in microfabrication, there are now many options available to pattern a microfluidic design into a material chip. The nature of the fluids, the features of the design, the size of the droplets, the cost and amount of emulsion will dictate the choice of one material over another. The main specification that is sought is a cheap, rapid generation of the emulsion from a limited amount of reagents. Historically, the first microfluidic devices were fabricated by lithographing and etching channels into a glass substrate. The method is laborious as the fabrications steps must be repeated in a

clean room each time a new batch of chip is need. This difficulty in fabricating chips held back the initial development of microfluidic. The situation changed in the late 90s with the introduction of replica molding by Georges Whitesides and colleagues [66]. They showed that polydimethylsiloxane (PDMS) -a polymer with elastomeric properties- could replicate the shape of a master mold upon curing, with a resolution of dozens of nanometers. The properties of PDMS, such as gas permeability and control of the hydrophobicity, have made it a golden standard for (molecular) biology microfluidics. The polymer has allowed hundreds of labs in the world to cheaply and effortlessly fabricate microfluidics chips.

Fig.2.7 shows an overview of the main techniques mentioned above. In the case of glass micromachining, the main drawbacks are the complexity of the process and its cost, but these devices resist high pressures. Laser-based ablation or drilling on plastic polymers are far cheaper methods, but their channels are not well defined and rugged. Soft-lithography techniques are more common, because of their simplicity. This technique has been developed and tuned for microfluidics by Whiteside's group [67, 66] and are interesting in terms of cost, throughput and accuracy in the design.

2.2.1 Mold fabrication

This step patterns a silicon wafer using a negative photoresist with the mask. The patterned wafer becomes a negative mold of the fluidic circuit. A clean, dustless, silicon wafer is covered with negative photoresist uniformly spread on the wafer by spincoating. The choice of the photoresist, its amount and the spincoating procedure are lead by the aimed thickness of the microfluidic channels. Since the rheology of the photoresist is depending on the temperature of the room, one may follow the instructions given by the furnisher. In our case, we target a thickness of $55 \mu\text{m}$ ($\pm 10 \mu\text{m}$), so we used the negative photoresist SU-8 2075 (MicroChem) on a 4 inches silicon

wafer. The spincoater is set on 2000 rpm for 25 s.

The spincoat process is followed by a pre-baking step on a Triplet TH-900 Hot Plate: 1 min at 65 °C, then 6 min at 95 °C and a slow return to room temperature. The resulting assembly is then ready for UV-exposure. The mask is tightly laid on the hardened photoresist, and exposed for 20 seconds (6000 W, 150 mJ) under a Super High Pressure UV lamp (Ushio USH-250LT, USA). Here again, heating time, temperature and exposure time are driven by the final thickness and lamp power.

The next step is the post-exposure baking, in which the photo-activated compounds diffuse to surface and trigger the polymerization of the photo-resist. Time and temperature define the final thickness, and according to the literature the settings should be 1 min at 65 °C and 6 min at 95 °C for 55 μm .

The photoresist then soaks 5 min in a solution of SU-8 developer (MicroChem). Parts that were not exposed to UV remain polymerized and the UV-patterned surface is revealed. A second bath of developer and a rinse with isopropanol complete the revelation of the mold. During this step, the mold should remain immersed in solution to avoid the clogging of clots on the surface. Finally, we apply a Teflon coating on the mold to protect it from attrition/aging, and it makes the PDMS peeling easier. A mold can be used ~ 100 times before significant erosion occur. Actually breaking, rather than overuse, is the main reason a mold needs to be replaced.

2.2.2 Chip casting

Chips come from polymerization of PDMS on the mold. First a cross linking agent is added to the PDMS (SILPOT 184 W/C, Dow Corning Toray, Japan) with a typical ratio of 10% (w/w), mixed thoroughly for 30 s at 2000 rpm (Thinky ARE 310) and poured in a Petri dish with the mold. Typically, we use 72 g of PDMS base and 8 g of cross linking agent for one batch, the final thickness of the chip

is ~ 8 mm. Thicker chips will facilitate the tubing process. After 20 minutes under reduced pressure to degas the mixture, the PDMS is cured at $75\text{ }^{\circ}\text{C}$ for one hour and a half. Once hardened, the PDMS cake is peeled off and immediately put on a plastic sheet (peeled face in contact with the plastic) to prevent any clogging of the channel by dust in the air. Inlets and outlets are opened thanks to a clean hole-puncher with a diameter slightly smaller than the external diameter of the tube. The punching should be neat in order to avoid any tear of the PDMS, because it would lead to fluidic clogging/problems later in the process. Just before bonding, both PDMS surface and a glass slide are cleaned with an O_2 plasma by reactive ion etching process (RIE-10NR, Samco International): 50 SCCM, 20 Pa, 75 W for 5 s. This allows a covalent bond between the resulting silanols of the PDMS surface on the glass slide [66]. This step is critical because it is the most probable chance to catch dust particle on the PDMS or glass, leading to an unusable chip.

Finally, a baking step ($190\text{ }^{\circ}\text{C}$, 5 h) is used to make the whole channel surface (glass and PDMS) hydrophobic and improve bonding [68].

2.3 Oil & Surfactants

The size of droplets can be easily adjusted in a microfluidic device to fall within a range useful for single cell biology, biochemistry and molecular programming ($1\text{-}100\text{ }\mu\text{m}$). But in addition to size, droplets must meet other important specifications for biological applications: they must be stable against coalescence over a long time, biocompatible, and reagents should not leak between them. Here the choice of oil and surfactant will prove crucial to meet these requirements.

Two physical effects (Ostwald ripening and coalescence) conspire to destabilize and coalesce droplets arrays. Coalescence occurs after drainage of continuous layer between two droplets. Ostwald ripening is a passive, thermodynamic effect. Consider a polydisperse emulsion,

in which larger droplets are more stable than smaller ones thanks to their lower surface tension. Because of this imbalance, smaller droplets tend to transfer material to larger ones to reduce their surface tension, further increasing the size of the larger droplets. The average radius R of droplets will evolve in function of the temperature (T), surface tension (γ), solubility (c_∞) of the transiting particle, its molar volume (ν) and its diffusion coefficient (D). Here R_g is the ideal gas constant [69]:

$$R^3 - R_0^3 = \frac{8\gamma c_\infty \nu^2 D}{9R_g T} t$$

SURFace ACTive AgeNTs (named hereafter surfactants) are designed to prevent both coalescence and Ostwald ripening. They consist of amphiphilic molecules, that is molecules that carry both hydrophobic and hydrophilic moieties. Because of this, they preferentially sit at interface of droplets, which reduces the global surface tension between the phases. The surfactant effect is quantified by the amount of surfactant molecules on the droplet's surface (Γ , surface concentration), the surface tension and its bulk concentration (c) [70]:

$$\Gamma = -\frac{c}{R_g T} \frac{d\gamma}{dc}$$

Surfactants prevent coalescence thanks to the Marangoni stress produced when two droplets approach. As two droplets get close, the resulting flows expel surfactants, which increases surface tension. In turn, the surface tension gradient drives the convection of surfactants to replenish the depleted area, thus slowing drainage and stabilizing droplets (Fig.2.8).

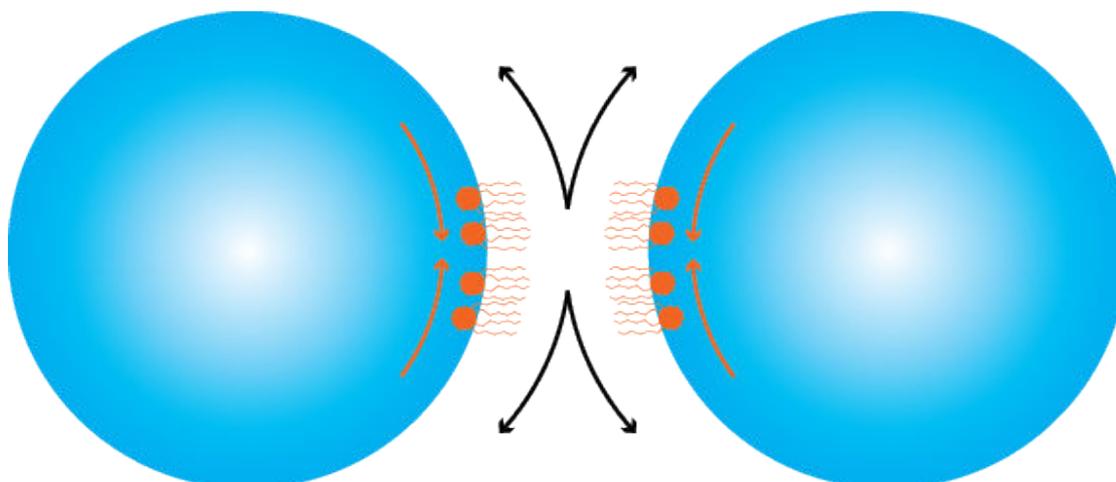


Figure 2.8: Marangoni effect on two approaching droplets: the droplets (blue) are approaching and creates a drainage of the immiscible phase in between (black arrows) before coalescence. As a result, a Marangoni flow appears on the surface of the droplets (orange arrows) thus concentrating the adsorbed surfactant (orange) on the apex of the droplet. The hydrophobic tails then repel the droplet, stabilizing the emulsion by increasing the drainage time. Adapted from [63]

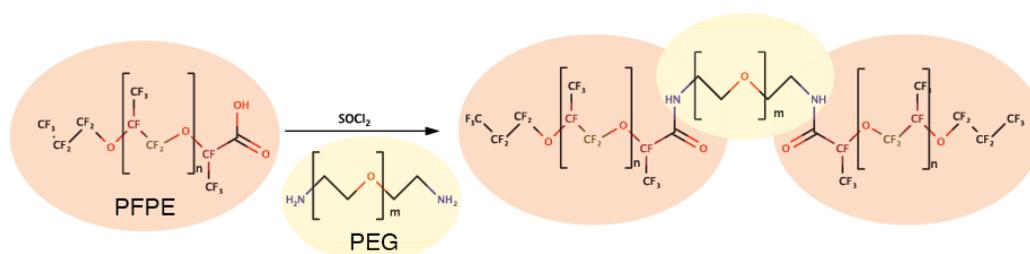


Figure 2.9: Non-ionic fluorosurfactant adapted from Krytox: perfluorinated polyethers (PFPE) coupled with polyperfluoropropylene glycol (PEG)

Concerning the continuous phase, two kinds of oil are used for microfluidic-driven molecular biology: hydrocarbon oils, fluorocarbon oils and silicon oil. Hydrocarbon oils are used for PCR applications or directed evolution of enzymes, but hydrophobic compounds can easily diffuse through the oil phase, which limits their use. Fluorocarbon oils are both hydrophobic and hydrophilic, with low solubility of the reagents for the aqueous phase, and they limit the swelling of PDMS during the emulsion production process. Fluorinated oils, together with fluorocarbonated surfactant, are widely used for cell analysis or gene amplification [71][72]

The nature of the oil has a strong effect on the generation and handling of the droplets. For instance, viscous oils, such as mineral oils, requires a high oil-to-water pressure ratios during generations. The material of the microfluidic device is another important criteria: in case of PDMS-based devices, silicon oil are unusable because of their solubility in PDMS (which is why we did not consider such oils). In this case, hydrocarbon oils provide an alternative (although care should be taken about possible diffusion of reagents in it, as mentioned above) [73]. Temperature also affects the physicochemical properties of oils and surfactants, which may destabilize the emulsion. Therefore, a careful choice of oil, surfactant, aqueous reagents and chip material should be made upon designing the experiment in order to provide an appropriate control over the size of droplets, the leakage of reagents and the stability of the emulsion.

Another practical difference between oils is their densities. In the case of hydrocarbon or fluorocarbon oil, the emulsion lays on top of the oil phase. In the case of mineral oil, the emulsion is heavier and sit below the oil phase. Since we want to process the emulsion off chip, the use of fluorinated oil is more comfortable to collect the “cream”. An additional point is that the lower viscosity of hydrocarbon oils leads to a smaller Ca , so the pinch off occurs a smaller oil/water pressures ratio and provides a finer tuning over the droplet size. Commonly used oils, and available surfactants are listed in Table 2.3. We chose the fluorinated oil HFE7500 (3M Novec) as the continuous phase for all experiments.

Fluorosurfactants suitable for use with fluorinated oil have been developed by several academics groups [56, 63, 74]... Commercially available fluorosurfactants are sold as Krytox by DuPont. Their structure consists of a fluorocarbon tail like perfluorinated polyethers (soluble in the oil phase) and ionic headgroups (soluble in the aqueous phase) such as poly(perfluoropropylene glycol)-carboxylates. Unfortunately, ionic headgroups interact with charged biomolecules like DNA or RNA or proteins.

Oil	Mineral oil (Sigma Aldrich)	Silicone oil AR20 (Sigma Aldrich)	FC-40 (3M)	HFE7500 (3M Novec)
density (g.mL ⁻¹)	0.84 at 25°C	1.01 at 20°C	1855	1.614 at 25°C
viscosity at 25°C (cps)	~ 30	~ 2	4.1	0.77
boiling point (°C)	?	> 140	165	128
Surfactants used	none/ ABIL EM90	none/ Triton X-100/ SDS/ABIL EM90	PF decanol/ PFPE-PEG	PFPE-PEG (Ran Tech, EA Raindance)
PCR compatibility	+	+	++	++

Table 2.3: Commonly used oils and surfactants for bio-compatible droplet generation, adapted from [63]

To bypass this issue, groups started to synthesize fluorosurfactant similar to Krytox, but lacking ionic headgroups [75] (Fig.2.9). A notable example is the surfactant from Raindance Technologies developed by Griffith and al [71] - with a derivative being commercially available (RanTech). As for us, we have used Jeffa1000 kindly provided by V. Taly and J.F. Bartolo. We also use Bovine Serum Albumin (BSA, New England Biolabs) to stabilize proteins and prevent non specific bindings, which typically occur at interface (microfluidic walls, oil/water interface...).

2.4 Script generation, barcoding, and parameter space

In our device, we map the bifurcation diagram of a given molecular system by preparing droplets with different concentrations. The content of droplets is tuned by adjusting the pressures applied on inlets. Let us focus now on the way those pressures profiles are programmed. The main constrain we have to follow is that droplets should be monodisperse: their volumes should be identical. Polydispersity complicates the comparison of fluorescence level between two droplets. Moreover, it has been reported that the confined environment of a droplet may influence the chemical reactivity; droplets

with different size may therefore exhibit different chemical dynamic [76]. To enforce monodispersity during droplet generation, we ensure that the total pressure applied on the aqueous channels is constant. Since the aqueous channels are identical (modulo fabrication error), this guarantees that the total aqueous flow rate is constant. To that end, we use an additional channel (the compensation channel containing the master mix) which tracks and compensates variation in flow rates in the other aqueous channels (the working channel).

2.4.1 Drawing a shape in parameter space

2.4.1.1 1 dimension: the line

The simplest case is when one parameter only is explored: one only needs to change one concentration. The droplets explore a 1D line in the parameter space. For the microfluidic generation, we need two channels: one channel containing the master mix and the parameter at maximal desired concentration, and another channel containing the master mix only.

The first step of the protocol for droplet generation is to empirically find the maximum (P_{\max}) and minimum pressures (P_{\min}), corresponding to the pressure for which one channel invades or is invaded by the other one. The script starts from an equal pressure in channel 1 and 2, and consists of a loop in which the pressure in channel 1 is gradually decreased and increased in channel 2 gradually until the P_{\min} in channel 1 is reached. Then the pressure in channel 1 is increased as much as the pressure in channel 2 is decreased until the invasion of channel 2. The pressure increment is $\frac{P_{\max}-P_{\min}}{nb_{iter}}$ where nb_{iter} is the number of iterations during one scanning. The higher nb_{iter} is, the higher the resolution of the line is, but the drawback is the time spent running the script. We usually use 20 iterations to ensure a good coverage over the pressure range.

2.4.1.2 2 dimensions: triangle and square

Adding a variable to the system means that we now have to screen the parameter space across a surface and not a line. Therefore an additional channel is required for the second parameter. We now have three channels : channel (Ch1) containing the parameter 1, channel 2 (Ch2) containing the parameter 2 and Channel 3 for the compensation (Ch3).

The first geometry we could achieve was the triangle shape. For this polygon, each vertex of the simplex corresponds to one channel invading the two others. As a starting point, we chose the one vertex corresponding to the Ch1 reaching its P_{\max} . On that point, (Fig.2.11, a) Ch2 and 3 are invaded and the pressure in Ch1 is decremented by $\frac{1.5(P_{\max}-P_{\min})}{nb_{iter}}$ (while Ch3 is incremented by the same amount, in order to keep the sum $P_{Ch1} + P_{Ch2} + P_{Ch3}$ constant (Fig.2.11, b); then, P_{Ch2} increases by the same increment (Fig.2.11, c) before the decrease of P_{Ch1} (Fig.2.11, d). Finally, the decrease of P_{Ch2} brings it to its original value (Fig.2.11, e). This pattern is iterated nb_{iter} times until the P_{Ch2} reaches the P_{\max} value. Here we name Ch2 the fast variable and Ch1 the slow variable because the pressure range for Ch2 was higher than the range for Ch1. The fast and slow variables are then exchanged, i.e. we start from the vertex in which $P_{Ch2} = P_{\max}$ and perform the same routine. The whole scan is completed ~ 10 times in order to obtain at least 30 μL of emulsion. Pressures and script are summarized in Fig.2.11.

The main advantage of this geometry is the simplicity of implementation, and the time spent in running the script. Indeed, finding the boundary pressures and perform one complete scanning will require less than 10 min, which means that the embedded chemicals will have less time to react before the encapsulation. The drawback comes from the fact the triangle shape itself: if two parameters are supposed to interfere with the global behavior of the network, the triangular shape give access to half of the possible combinations of

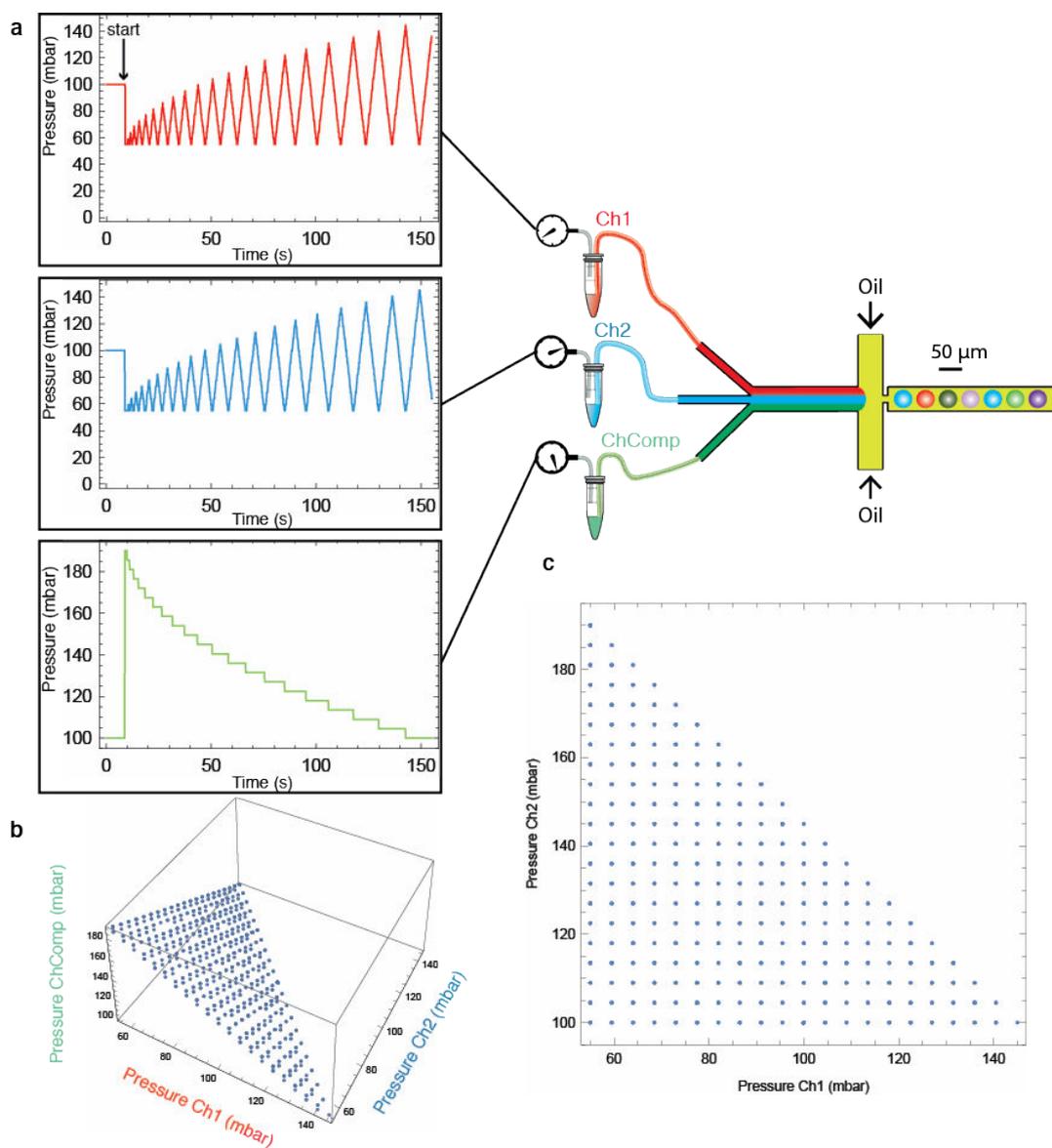


Figure 2.11: Pressure profiles for triangle mapping. for all aqueous channels, the script applies pressures (a) whose sum draws the simplex in pressure space (b). While projected according to the parameter-relevant pressures (i.e. pressures for which the aqueous solution contains a parameter) the shape of the triangle appears (c).

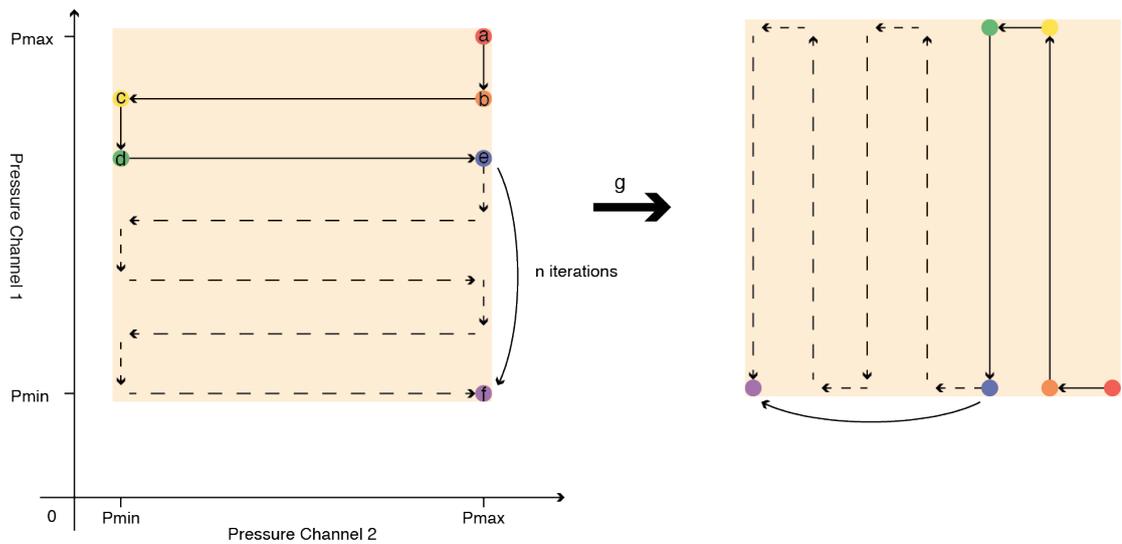


Figure 2.12: Pressure variations strategy for deterministic square mapping. The compensation channel (not shown) changes its pressure in the opposite way of the parameter channels to maintain the sum of pressures constant. The script operates from (a), where pressures in channel 1 and 2 are equal and the compensation channel invaded. After a reduction of pressure by half an increment value in channel 1 (channel 1 is the slow variable here) (b), the pressure in channel 2 (named as fast variable) is decreased until its P_{\min} value (c) and the pressure in channel 1 is reduced again of half increment (d). Finally, the pressure in channel 2 goes back to P_{\max} . This pattern is repeated n times until the pressure in channel 1 equals its P_{\min} value (f). After that, fast and slow variable exchange and the script is played again (g).

marized in Fig.2.13.

2.4.1.3 The cube

Adding a third parameter to the exploration of the bifurcation requires a fourth aqueous channel and means that the resulting diagram will be in three dimensions. Our first approach was a generalization of the square script. The volume is sliced in n -squares (typically 20), and we gradually decrease the pressure in the channel corresponding to the third parameter to switch from one slice to the next one. In order to thoroughly scan the whole cube, the 3 channels containing the parameters are in turn slow variable of the square, fast variable of the square, and sliced third dimension. This deterministic method is quite difficult to implement in terms of code, and one of the main drawback is the time required to scan the entire cube. In fact, since we repeat the previous script on 20 slices along 3 dimensions, one full scan lasts 40 minutes. During this time, chemical reaction could occur inside the reservoirs and corrupt the whole experiment. To reduce the scanning time, we put aside the deterministic approach for a Monte-Carlo sampling within the parameter space. We generate a random walk inside the $[0, 1]^3$ cube thanks to Mathematica software, that is rescaled in the scripting software into $[P_{min}, P_{max}]^3$ with the P_{min} and P_{max} values. The script is a sum of thousands of 3D increments $(\Delta_1, \Delta_2, \Delta_3)$, where each Δ_i is drawn independently from a Pareto distribution. Pareto distributions are set to ensure that the pressures sometimes move abruptly to new regions of the cube, offering a better coverage of the volume. The boundary conditions maintain the random walk inside the cube. The oil pressure is set to 700mbar. The pressure profiles and resulting volume in pressure space are shown in Fig.2.14.

Thanks to this new method for volume sampling by random-walk, the scanning time is lowered to 20 minutes (setting 3000 points within the cube, plus three calibration processes as described for the square

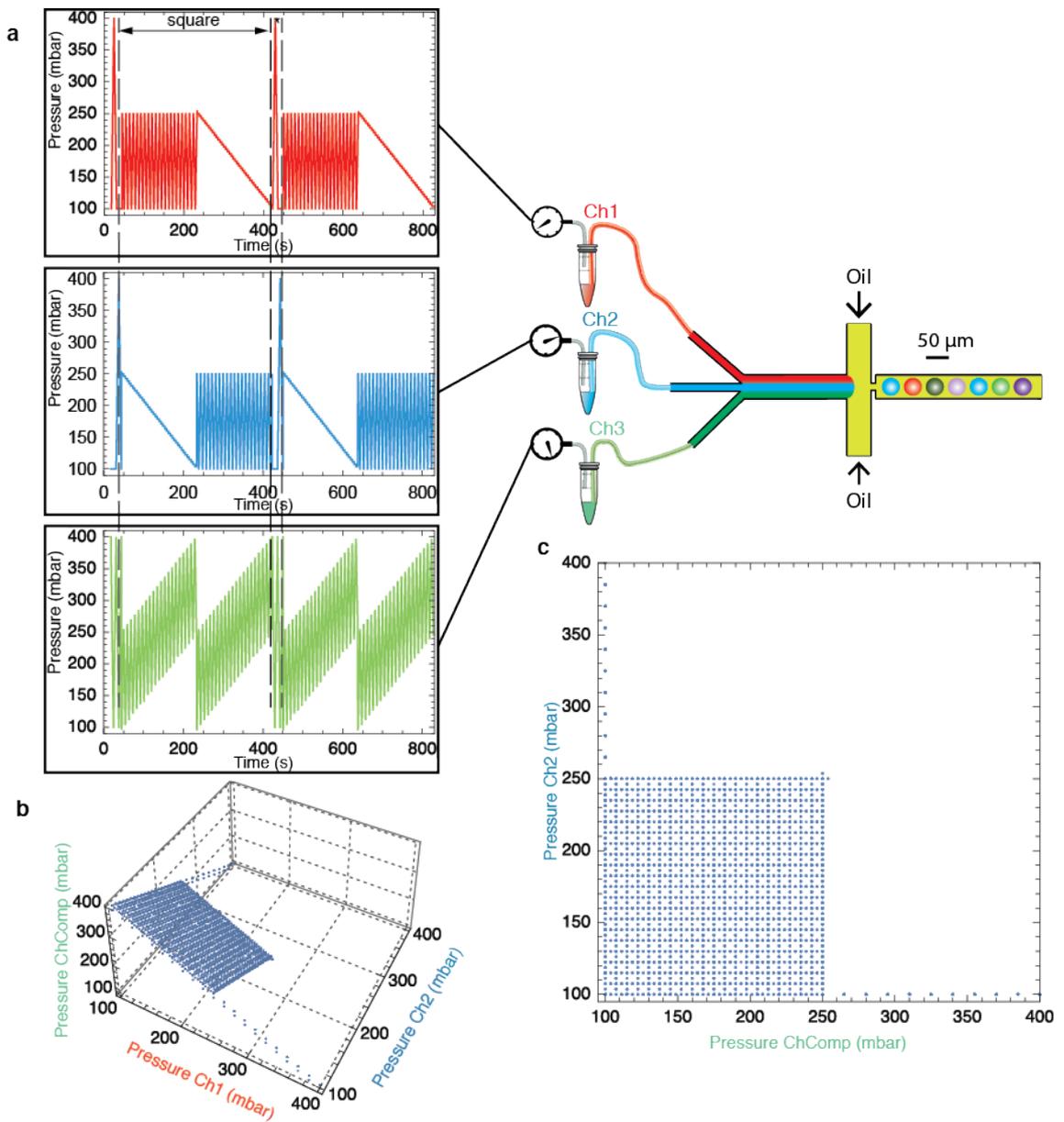


Figure 2.13: Pressure profiles for square mapping. Individual pressures channel by channel (a), square in pressure space with calibration arms (b) and projection in parameter-relevant pressures.

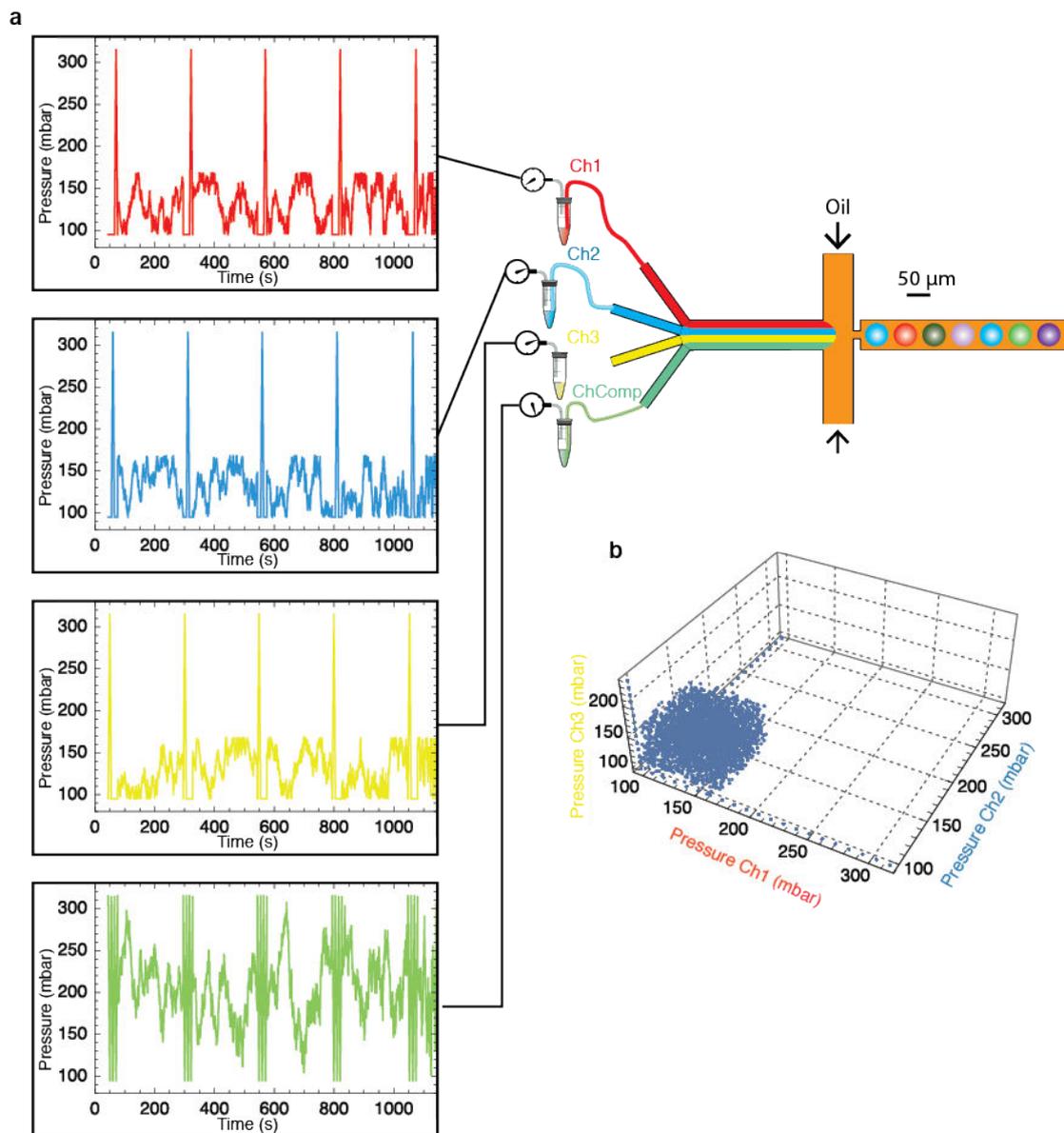


Figure 2.14: Pressure profiles (a) and resulting volume in pressure space (b) for random-walk cube mapping.

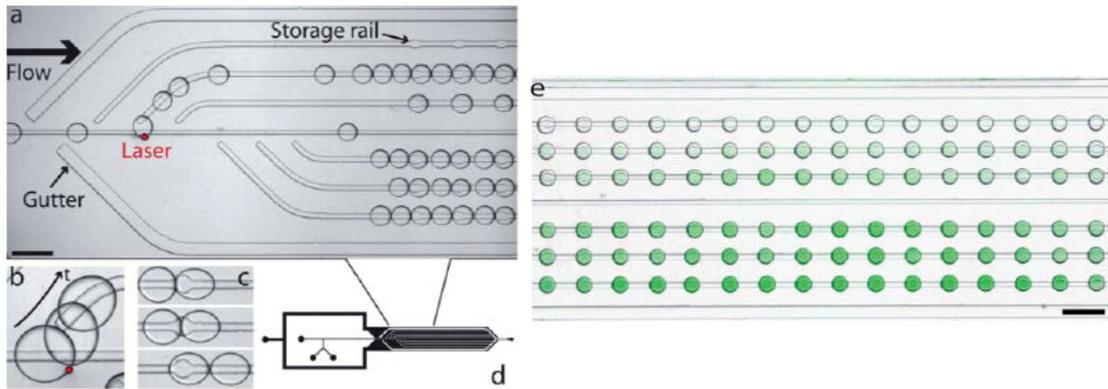


Figure 2.15: Droplet indexing: rail and anchors from Baroud *et al.* [77]. Droplets flow on a central rail (a) until they are deflected to a storage rail by the laser heat (b). Droplets are trapped on the anchor points once the flow is reduced (c), forming a big cluster (d). The flowrate is $40 \mu\text{L}\cdot\text{min}^{-1}$. Droplets can be sorted along a gradient of concentration of green fluorescent beads(e). Scale bars are $600 \mu\text{m}$.

mapping), which reduces the probability of unwanted chemistry before encapsulation. The other advantage of this method is the ease with the implementation of another varying parameter (i.e. an hypercube mapping in parameter space). An alternative deterministic way would use space filling curves to screen efficiently the space.

2.4.2 Sorting droplets

We now have a microfluidic tool allowing the generation of droplets with different concentrations of chemicals by changing pressures applied to aqueous channels. The immediate question that comes to mind is: when observing a particular droplet in array, how can we find the concentration of parameters it contains ?

One solution is to construct a ordered array of droplets: the content of droplet is indexed by its spatial location in the array. This method has been developed by Baroud *et al.*: by patterning “rails” in the PDMS chip and with the help of laser, they can guide, sort and store droplets at precise location [77]. Roughly speaking, the rail is simply a groove etched from the PDMS layer of the chip. The size of droplets is chosen to be larger than the heights of the device;

droplets are flattened like “pancakes” which cause a large buildup of surface energy. Droplets anchor to grooves because they offer an additional volume to locally relieve the pancake shape into a sphere, which decreases surface tension. If the external convective flow of the continuous phase is not too strong, droplets remain tucked and smoothly follow the contours of the grooves. The device shown in the article is made of a central rail from which emerge 6 storage rails with anchors (Fig.2.15, a). To move a droplet from the central rail to its storage area, a laser is activated to deflect a droplet to its destined storage rail. The force coming from the local heat produced is enough to eject the droplet from the rail (Fig.2.15, b). Anchor points along the storage rail ensure that a droplet is trapped at each anchor site. If the laser is synchronized with the droplet generation, it becomes possible to temporally sort droplets. The authors demonstrate an example of controlled filling along droplets with a gradient of green fluorescent beads (Fig.2.15, e). Depending on the number of anchor points and rails, the storage capacity is the main limitation. Indeed, in our system with the usual ratio of water/oil pressure, the frequency of droplet generation is higher than 1 kHz. Knowing that the scanning time for the simplest mapping (i.e the line) is around 10 seconds, it means that the device should be able to store at least 10000 droplets. The cube mapping will require more than 1 million anchor points for one scan. Another limitation of spatial sorting is that it is sensitive to sorting errors: slight desynchronization between generation and sorting will build up over time and result in the incorrect indexing of a large swath of droplets.

Another technical solution is a direct, online, readout within the microfluidic chip. After generation, a droplet flows inside a channel until it crosses a detector which reads its content (Fig.2.16). This technique is used in digital PCR and protein engineering, and allows high throughput analysis. Ismagilov and colleagues developed it to investigate the kinetics of an enzymatic reaction with a resolution lower than the millisecond [78]. Droplets are generated from a

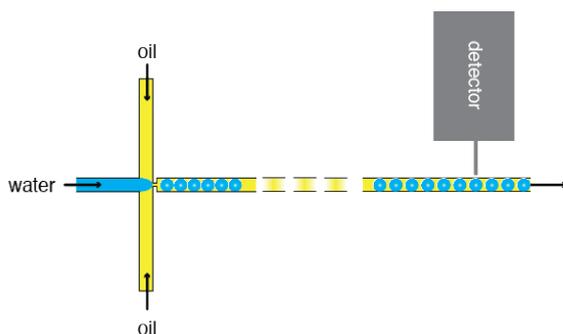


Figure 2.16: Droplet indexing: direct readout. After generation, the droplets flow in the outlet channel until they cross the detector

T junction from the combination of 3 aqueous flows containing different proportions of reagents. As droplets travel along the outlet channel, the reaction residence-time increases. Fluorescence images (acquisition time scaled on the outlet channel length) deliver each kinetic profile [78]. However, the reaction starts immediately after droplet generation. So the channel's length has to be scaled to fit the timescale of the reaction. Since the classical timescale for simple PEN-DNA toolbox networks is around 3 hours, this method does not fit our needs.

2.4.3 Barcoding

The approach we chose was to forgo spatial indexing and online reading. Instead, we barcode the content of a droplet with a fluorescent, chemical inert, species which tracks the concentration of a parameter species. By reading, off chip, the barcodes' fluorescence of a droplet, we can find -after proper calibration- the concentration of parameter it contains. After some attempts with labelled DNA or BSA, and following the work of Ji and al. [79], we barcoded aqueous channel with fluorescent Dextran. Dextran conjugated to a wide choice of fluorophores are now commercially available. We chose fluorophore according to our available laser and to minimize spectral overlap. Each channel contains a specific Dextran that encodes the concen-

tration of its parameter species. When the channel gets mixed to other channels, the Dextran is diluted accordingly, and thus keeps track of the concentration of the parameter species.

We estimated the relative errors associated with using labelled dextran as barcodes. In our barcoding strategy, the fluorescence intensities of dextran are taken as a proxy for the concentrations of a non-fluorescent parameter species (DNA template or enzyme) that has been co-encapsulated. Yet dextran are branched polysaccharides whose physicochemical properties (charge, diffusion coefficient, . . .) differ from their surrogate species. We performed the following experiment to assess the reliability of labelled dextran as barcodes. We prepared two mixtures, each containing a labelled dextran and a labelled DNA strand (Fig.2.17, a). We then generated droplets containing r % of the mixture 1 and $(100 - r)$ % of mixture 2, with r taken between 0 and 100, using our microfluidic chip and pressure scanning strategy. This experiment recapitulates various sources of microfluidic noise that could affect the barcoding precision. The mixture 1 contains Cascade-labelled dextran (5 nM), Alexa 647-labelled dextran (100 nM), the FAM-labelled $\alpha\text{toi}\beta$ (100 nM), and the TAMRA-labelled $\beta\text{toi}\alpha$ (10 nM). The mixture 2 contains Cascade Blue-labelled dextran (200 nM), Alexa 647-labelled dextran (20 nM), the FAM-labelled $\alpha\text{toi}\beta$ (10 nM), and the TAMRA-labelled $\beta\text{toi}\alpha$ (100 nM). All mixtures contain the same buffer. The droplets are scanned at room temperature after generation (spec: 512x512 pixels per field of view, $20 \mu\text{s}\cdot\text{pixel}^{-1}$, CA $120 \mu\text{m}$, zoom 1x, objective 20x (NA0.75)). Fig.2.17 c shows the scatter plot of fluorescence intensities. As expected, the intensities of dyes in the same microfluidic channel are strongly correlated, while dyes in different channels are anti-correlated. More precisely, we can estimate the (relative) reading errors associated with measuring the fluorescence of dextran markers instead of their surrogate DNA strand. Using Mathematica we perform a linear regression to extract the coefficients of variation. For (Cascade Blue Dextran, TAMRA-labelled DNA) this error

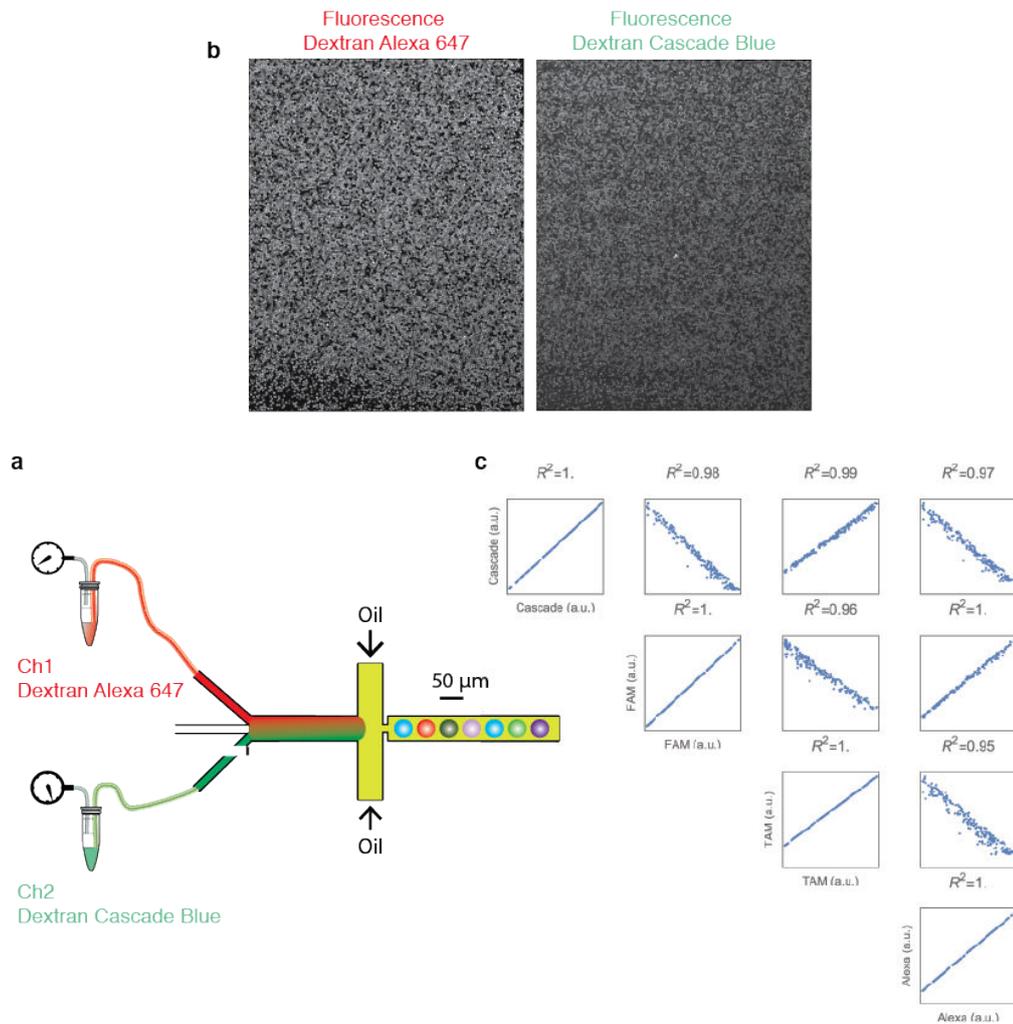


Figure 2.17: Molecular barcoding. Droplets are formed from two aqueous channels containing different fluorophores following a line mapping script (a). The pressure ratio in between the two channels (left) leads to a gradient of concentration, thus fluorescence in between the droplets (b). Scatter plots (in arbitrary fluorescence units) showing the correlation of the fluorescence dyes within a single field of view ($N=218$ droplets).

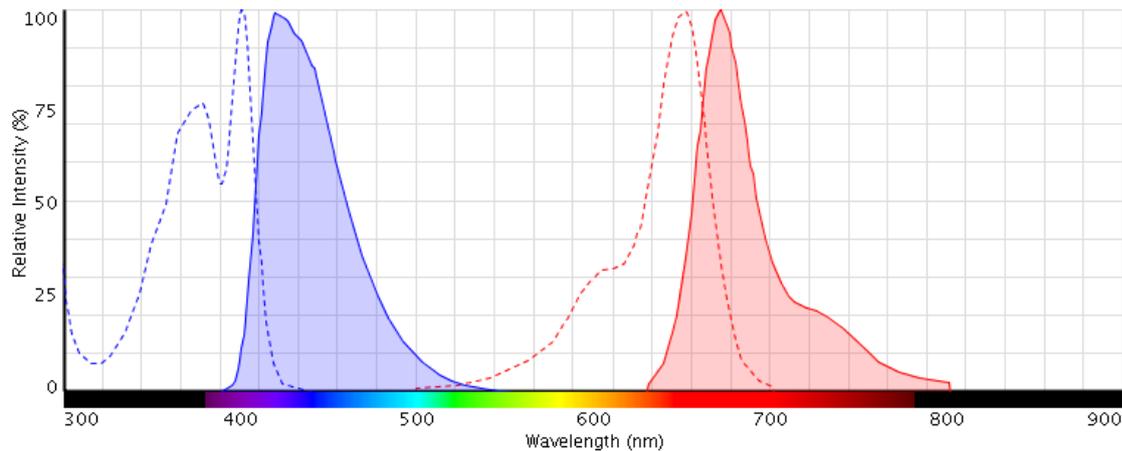


Figure 2.18: Excitation (dashed) and emission (plain) spectra for the barcodes Dextran Cascade Blue (blue) and Alexa 647 (red), from Life Spectra Viewer

is $\sim 3\%$. For (Alexa Dextran, FAM-labelled DNA) this error is $\sim 4\%$. This experiment therefore confirms that labelled dextrans can reliably barcode the concentration of DNA templates with a relative precision of a few percents. The experiment also shows that leakage during droplets generation does not significantly add noise to barcoding. Upon confocal imaging, we find that the fluorescence of Alexa 647 and Cascade Blue are anticorrelated, which shows that Dextran can act as a surrogate barcode.

2.5 Observation chamber

The following step after generation of the droplets is their time-resolved monitoring in a dedicated chamber. Indeed, for a given experiment, we need to record the intensity of reporter fluorophores (reporting on the state of the system) as well as barcoding fluorophore (which keep track of experimental conditions). As we saw previously, we chose to monitor the droplets off chip, in a separate chamber. This method requires an observation chamber in addition to the microfluidic chip. Here are some important properties that must be met by the chamber (besides being easy to fabricate) :

- **monolayer:** for an optical readout, droplets needs to be stored in a monolayer. Multi-layers complicates he measurement of fluorescence.
- **physical stability and absence of coalescence.** We saw previously that the use of surfactant prevents the coalescence of droplets, which is why we add surfactants both to the water and oil phases. But the surface of chamber should be rendered hydrophobic to stabilize emulsion.
- **absence of evaporation:** it should be possible to incubate an emulsion at a relatively high temperature (the working range of working temperatures of the PEN-DNA toolbox: 42-49 řC) without evaporation of oil, water or penetration of air.
- **compactness:** the density of droplets inside the device should be maximal, meaning that the packing of the emulsion within the chamber has to be tight. Tight packing increases the throughput of fluorescence reading, in addition to preventing movement of droplets.
- **immobility:** since the timescale of an experiment with the toolbox is ~8 h, it is utterly important that droplets do not move noticeably. Slight displacement of droplets complicates data analysis since it requires tracking and may introduce indexing errors.

The current chamber design and protocol is an upgrade from the original device described by Hasatani *et al.* [80]. The emulsion was originally sandwiched between two glass slides, a thick bottom slide (Matsunami Micro slide glass 76 mm \times 52 mm thickness 0.8~1.0 mm) and a thin cover slide (Matsunami Micro cover glass 22 mm \times 24 mm thickness 0.12~0.17 mm), both modified with hydrophobic coating (Durasurf, Harves), sealed by Araldite Rapid (Araldite Professional Adhesives) glue. Although the system was robust enough

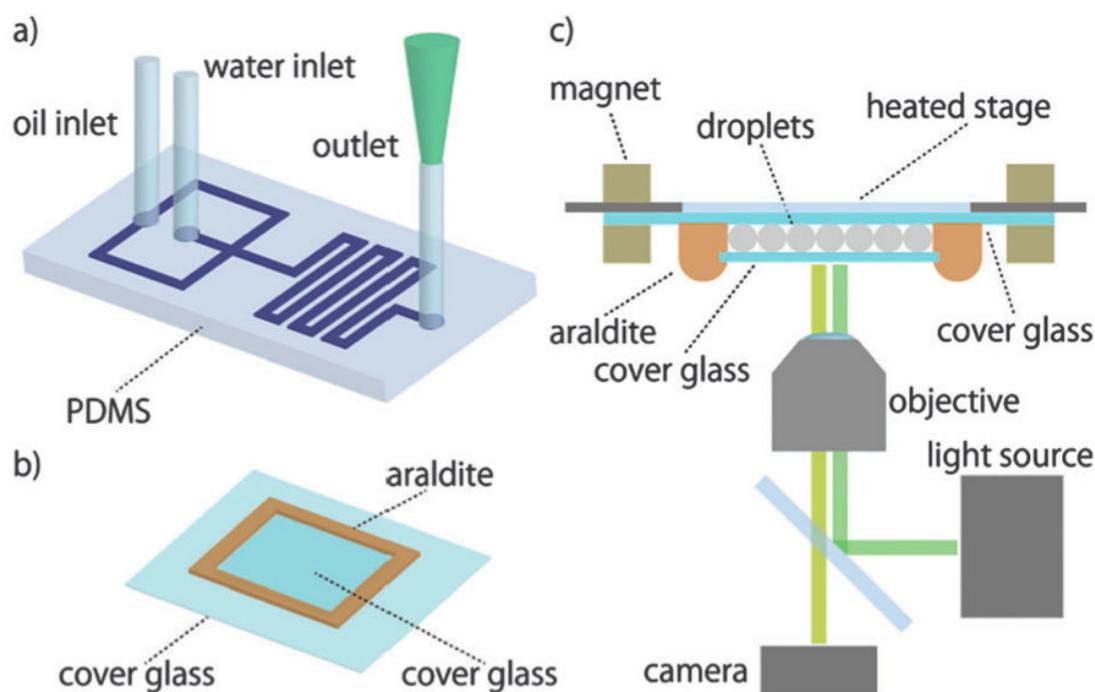


Figure 2.19: Observation chamber developed by Hasatani *et al.*: after generation the droplet emulsion is sandwiched between two glass slides coated with hydrophobic coating (Durasurf) and the device is sealed by Araldite glue. First, droplet are generated using a microfluidic device (a). The resulting emulsion is sandwiched between two hydrophobic-coated cover glasses and sealed with Araldite (b). Finally, the chamber is reverted on the heat plate to enhance imaging conditions (c). Reproduced from [80].

to undergo full day of timelapse, several aspects needed to be fixed: merging of droplets was not uncommon, packing was not tight and droplets were not immobile (actually the droplets moved so much that we initiated a collaboration with a specialist of colloidal tracking, Mathieu Leocmach, to help us with that problem).

First, we had to prevent the emulsion from stacking into a bilayer (which worsens the quality of image acquisition and makes image processing difficult). We introduce spacers to physically constrain the array in a monolayer and reduce coalescence. We first used polystyrene beads (Polybeads, polystyrene microspheres $50 \mu\text{m}$), spread on the bottom slide before deposition of the emulsion. The formulation of the polystyrene beads (in water) was found to be incompatible with

our system (hydrophobic glass coating, water-in-oil droplets) so we turned to glass beads instead (Polysciences Inc.) filtered on 50 μm CellTrics filter (Partec). The spacers' diameter controls the depth of device and was chosen to match the targeted size of droplets: $\sim 50 \mu\text{m}$. The distribution of the spacers all over the surface ensures that the bottom slide and cover slide remain parallel. The spacers also support the cover slide and reduce the pressure it exerts on the droplets, which was a factor contributing to coalescence. Finally, the spacers are also useful to avoid that the droplets arrange in bilayers (Fig.2.20, left), this conformation making the optical readout much more complicated. However, droplets in contact with the spacers tend to merge because spacers are not treated with a hydrophobic coating.

Having eliminated one cause of coalescence (the pressure exerted by the coverslide on the droplets) we turned to other culprits. The main interaction between the emulsion and the chamber is through a thin hydrophobic coating. We changed the Durasurf coating for a Cytop 809M (10% diluted in the provided solvent, Asahi glass), an amorphous fluoropolymer widely used in microfluidic for hydrophobic coating (water contact angle of 112° , against 53° for hexadecane), and compatible with optical measurements [81]. This coating is applied by spincoating at 500 rpm for 5 seconds then 2000 rpm for 30 seconds. The freshly coated glass slides are then baked at 180°C for 60 minutes. The emulsion was found to be much more stable, and the residual merging events were often due to dust speckles coming from the continuous or dispersed phase.

The last issue to solve was the air tightness over the long term. We often found that air bubbles appeared, grew and thereby disrupted the droplet array. The presence of air bubbles is not a problem *per se*, but their disruptive movement seriously complicated the tracking of droplets. In fact sudden jiggles were such a serious problems that it made unusable a large number of droplets (typically a few dozens percents of the population).

Where do the bubbles come from ? We suspect that our major source of air bubble is the sealing of the chamber, when we (carefully) deposit the cover slide on the emulsion and spacer, and then seal the assembly with an adhesive. Also, the spacers possibly enhance the growth of bubbles by nucleation. So we investigated both the filling and the sealing of the device. The choice of Araldite as a glue seal was discussed. This highly viscous two-components glue is uneasy to handle, and its optical properties (autofluorescence in green) make it difficult to scan droplet near the edge. But its main advantage, compared to other tested glues, that it is relatively inert regarding the chemistry happening inside the droplets. To ensure a tight junction between the edge of the cover slide and the bottom slide, we decided to supplement Araldite with a photocurable glue on the whole perimeter of the chamber. We focused on one glue in particular: NOA81 (Norland Products Inc.). The advantage is that this glue has already been used for cell management [82], so we guessed it would not interfere with our biochemistry. As it turns out experimentally, NOA spacers do interfere with the chemistry in the neighboring droplets, as we can see on (Fig.2.20, right).

We suspect that NOA monomers remain after curing and that their diffusion through the oil phase induces side-reactions within the droplets. As a consequence, we extended the exposure time to one minute to thoroughly crosslink the monomers. We also avoid as much as possible using data from droplets next to NOA materials (border or spacers), and removed the central spacer that served as a pillar.

The other front was the improvement of the filling technique. Instead of simply “dropping” the cover slide on the emulsion (which had to be deposited perfectly parallel and without shudder to prevent the emulsion from moving or air bubbles from being introduced), we thought about a way to directly prepare the chamber before filling it with the emulsion. This pre-assembly would solve both the bubble and packing problems. After some work, we managed to do it thank

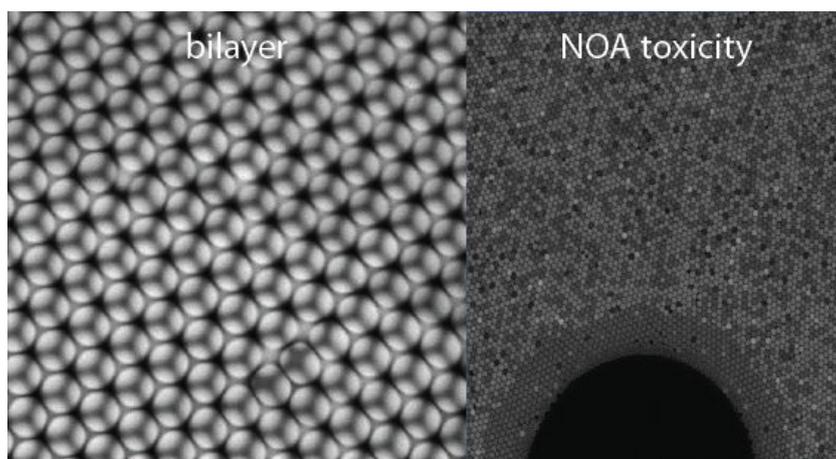


Figure 2.20: Microscopy bloopers: droplets form bilayer (left). Toxicity on the molecular program of photocurable glue NOA81(right).

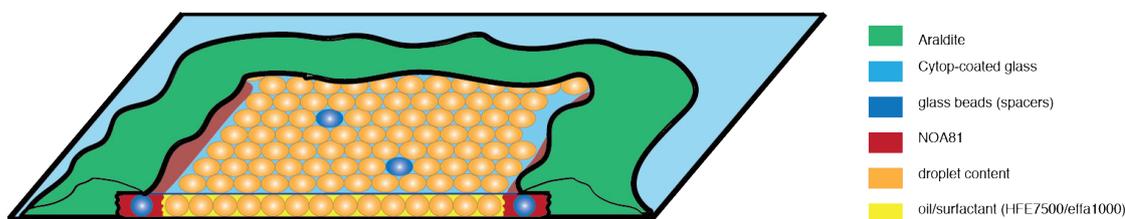


Figure 2.21: Observation chamber filled with emulsion.

to the NOA81 and the spacers. First, a small amount of spacers are spread on the bottom slide. Then, four drops of NOA are dispatched on the corners so as to fit with the four corners of the top slide. The assembly is then photocured, resulting in a rigid chamber with 4 openings (one per side of the cover slide). We sealed with NOA81 two parallel edges: the chamber becomes a tunnel, in which the emulsion can spread from one end to the other by capillarity and without introducing bubbles. After filling, the closure was ensured by NOA81 on the two remaining edges. Unfortunately, NOA81 itself does not provide enough airtightness. We improved long-term airtightness with a copious pouring of Araldite to form a “casing”.

To summarize, the current incubation/observation chamber consists of a thick glass slide for the bottom, and a thin cover slide for the top. Bottom and cover slides are coated with Cytop 809M

to make them hydrophobic and reduce merging. We use glass beads -filtered at $50\ \mu\text{m}$ to match the size of droplets- spread on the bottom slide with a tip. To rigidify the chamber and avoid bending of the thin top slide (which would compromise the spatial homogeneity of the imaging), we pour NOA81 on the spacers. Lastly, to enhance droplet filling and prevent air bubbles, we apply $\sim 10\ \mu\text{L}$ of NOA on two parallel edges of the chamber. We then cure the construct with an UV trans illuminator for 1 minute. The resulting chamber is filled by capillarity with $\sim 25\text{-}30\ \mu\text{L}$ of droplets emulsion, the 2 open edges are closed with NOA and UV cured again (the emulsion is protected from UV exposure by an aluminum foil). Finally, we seal the chamber with Araldite Rapid to improve long-term air tightness. The final result is shown in Fig.2.21.

2.5.1 Confocal microscope setup

The observation chamber is tightly held on a heating plate (Tokai Hit MAT-1002ROG-KX) thanks to two magnets. Immersion oil (Olympus, type F) ensures a good thermal contact between the heating plate and the chamber. The setup sits on an inverted confocal microscope (Olympus Fluoview FV1000) mounted on a IX-81 chassis and equipped with a xy stage Optosigma Bios-206T, 4 lasers (405, 473, 559 and 635 nm) and a 20x (NA=0.75) or 10x (NA=0.40) objective. As for Hasatani's device, because of the to the thickness of the heating plate and bottom slide of the chamber, interferences appear while imaging. Those interferences are lowered as soon as we revert the setup so that imaging of the droplet monolayer is done through a single thin cover slide.

2.6 Imaging

The last part of this chapter deals with the acquisition of fluorescence from the observation chamber. As mentioned previously, we use a

confocal microscope to record the fluorescence levels for each droplet, rather than a classical wide-field epifluorescence. The main advantage of confocal microscopy is its ability to generate small optical sections (above the micrometer size) that excludes light from out-of-focus objects (such as dust). Because it is a scanning technique, there are also less spatial inhomogeneities, as described in Fig.2.22.

Why did we choose confocal imaging ? The information of our system is encoded in fluorescence levels which can be very close (the indirect reporting of N-quenching shifts fluorescence by only a few dozens percent). Our need of a very high signal-to-noise ratio is a clincher that decided the use of confocal microscopy.

However, confocal scanning has a few drawbacks. The limited number of laser wavelength available (four in our microscope) is one, since we need one wavelength per barcode and observable. (See Table 2.4 for a non exhaustive list of available fluorophores.) This problem would be less significant with an epifluorescence microscope, which has a much wider range of excitation sources (mercury lamp or LED) and filters. The second limitation is that confocal scanning is slow, as the laser raster the sample to form an image. Acquisition of a single field of view typically takes ~ 10 seconds, which must be compared to the 0.1-1 s required for wide-field fluorescence microscopy. Because this time-limit is a serious bottleneck, we have spent considerable time optimizing the acquisition protocol to maximize the number of droplets scanned per unit of time.

2.6.1 Time-lapse

The chamber is heated at a given temperature by the hotplate, and fluorescence data is recorded for each droplet. For each droplet, the fluorescence levels of the barcodes indicate the initial parameter conditions (i.e. the coordinates in the parameter space), while the fluorescence levels of the input and output give the final state of the system (the color in the bifurcation diagram).

Laser wavelength (nm)	405	473	559	635
Fluorophores	Cascade Blue Pacific Blue	FAM Oregon Green	TAMRA JOE BMN3 Alexa Fluor 555 Alexa Fluor 546 DY-547 NED Cy3	Alexa Fluor 633 dy-630 dy-633 dy-631 BMN5 Cy5

Table 2.4: Available fluorophores according to the laser wavelengths of the confocal microscope

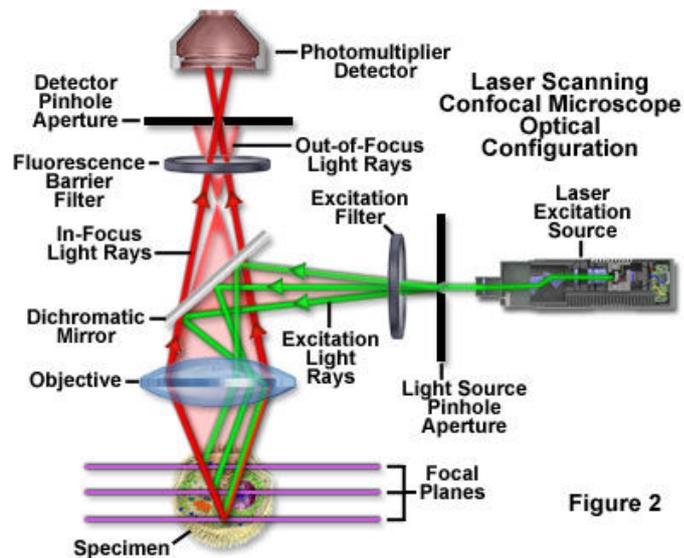


Figure 2

Figure 2.22: Confocal principle in epi-fluorescence laser scanning microscopy, from Olympus Microscopy Resource Center

2.6.1.1 Fast settings

We have to set the confocal properties for the experiment. The detector pinhole value reflects the number of photons reaching the detector. A smaller aperture leads to less background, but also less signal. On the contrary, a high pinhole aperture emphasizes background. Usually, we set the value of the pinhole to $\sim 100 \mu m$ to get the maximum amount of light and avoiding interferences (see above). The PMT collects light from the sample, and gives an intensity for each pixel. Depending on the laser intensity, the PMT needs to be adjusted so that weak signals are detected (high laser intensity, low PMT) without too much photobleaching (avoided with lower laser intensity, and higher PMT). The ideal values (maximum number of photons collected for minimal laser intensity) are empirically set for each experiment.

We also use a simple formula to approximate the number of photons hitting the sample: $zoom^2 \cdot N.A.^2 \cdot laserpower \cdot time \cdot pixel^{-1} \cdot C.A.^2 \cdot mag^2 \cdot npix$, where $npix$ is the number of pixel and mag the magnificence.

2.6.1.2 Time constraint

Another dimension that one should take care about is how long can a scan be. The dwell time is defined as the amount of time that the laser beam lingers in an unit of space corresponding to one pixel in the image. Long dwell time provide more photons per pixels. But it also leads to more photobleaching and longer acquisition time. The global timescale of the studied network also constrains the dwell time. In the case of a endpoint measurement (like the bistable switch where we only need the final state of the system), a long dwell time, and thus a long acquisition time, can be tolerated. In the case of a timelapse experiment, time-resolved measurements are required and long dwell time cannot be used. These factors, combined with the number of droplets required defines the time spent to acquire one

image.

2.6.1.3 Colors

The limited number of laser wavelengths available (405 nm, 473 nm, 559 nm, 635 nm, which is a respectable number for a laser confocal) makes some mapping quite delicate. In the case of a network with two different fluorescence observables, explored along 3 parameters, we need 5 colors to map the diagram of the network.

One solution to maps more than 4 dimensions is to simultaneously use one of the reporter dye as a barcode. This is possible if the dynamic of the observable is easily deconvolved. For example, initially reading the fluorescence of dye attached to a predator/prey template will provide the initial concentration of this template. Reading the variations of intensity of this dye over time will inform us about the oscillations of concentrations of preys binding to the template. Another option would be to use a XLdye: a fluorophore which exhibits a big gap between excitation and emission wavelengths (Stoke shift). Imaging the sample with one laser gives access to two parameter informations by using two different monochromator value as emission filters.

2.6.1.4 Optical defects

One of the biggest challenge in the confocal imaging of droplets arrays was the suppression of optical interferences. The laser goes through several layers of materials (oil, water, glass, air...) with different refractive indexes. Reflections occurring at interfaces whose refractive indexes are mismatched were found to severely compromise the quality of fluorescence readings, especially near the green wavelength. To smoothen the transition between materials with different refractive indexes, Booth and Wilson [83] provide some advices, such as reducing the pupil size to improve the focal distribution, or tuning the refractive index of the immersion medium to reduce mismatch. The

first improvement we undertook was to revert the heating plate. The «classical» position, where the heating plate sits at the bottom, leads to 5 interfaces (air/heating plate, heating plate/immersion oil, immersion oil/bottom slide, bottom slide/droplet oil and finally droplet oil/droplet content). By reverting the hotplate, we reeducate stacking to 3 interfaces (air/cover slide, cover slide/oil, and oil/droplets). An added benefit is that glass layers are now much thinner, going from a few millimeters (heating plate+ bottom slide) to hundreds of microns (coverslide). Glass thickness is important because most objectives are optimized for a standard glass thickness of ~ 170 microns.

The second improvement was to upgrade objectives to the best numerical aperture available. Numerical Aperture is the most important parameter of the most important component of the microscope: the objective. NA characterizes the «opening» of an objective lens, that is its ability to collect light from off-axis rays. In epifluorescence, NA is doubly important because the objective acts also as a condenser: it captures and condenses light from the illumination source. As a rule of thumb, the number of photons collected from a sample scales like NA^4 ! Even minor improvements to the NA can improve the imaging of dim samples. Importantly, the performances of an objective depend on the wavelength. Thus the choice of colors and working objectives should not be neglected when planning an experiment. In our case, the worst spherical aberrations were seen at the lowest wavelength: 405 nm and 473 nm. The best settings were then the use of a 20x objective (UPlanSApo 20x, N.A.: 0.75 nm, Olympus).

2.6.2 Endpoint method

This method suits experiments for which a time-lapse is not necessary and the final steady-state of the system contains the information.

Droplets are collected after generation in a PCR tube, in an incubator so that the network can compute at a constant temperature.

The incubation time and temperatures depend on the system, typically between 42 and 49 °C, from 3 to 24 hours.

One classical issue during the heating process of an emulsion is the coalescence of the droplets. This phenomenon is consecutive to the drainage of the oil phase, thinning the liquid phase in between the droplets. V. Taly and coworkers proposed a solution to avoid the merging of water-in-oil emulsion happening during a digital PCR experiment [56]. First, a cylindrical PDMS plug is thrust in the PCR tube. The plug shows with two thin holes (one inlet, one outlet). The emulsion is then injected inside the tube through the inlet, making sure that no air or oil is trapped in the device. After tubing removal, the PCR experiment can start.

An other solution, less efficient though, would be to cover the cream by another -lighter- oil (typically mineral oil) to prevent the evaporation. Surprisingly, we observe very little merging without using a plug or different oil film covering. This could be explained by the fact that, in the case of a PCR experiment, the temperatures reached are much higher than ours (moreover we do not cycle temperature). The emulsion turns out to be quite stable, no or few coalescence was observed.

Once the steady state is presumed to be reached, we fill the observing chamber with the incubated emulsion (as described before), and record its fluorescence over the whole array using confocal scanning.

2.7 Conclusion

In this chapter we detailed the development of a microfluidic tool for the experimental mapping of a bifurcation diagram. Our device generates millions of water-in-oil droplets with different composition by focusing aqueous flows. The composition is controlled by individually adjusting in real-time the pressure of each aqueous channel. The script executed by the pressure controller sample the parameter

space in a deterministic fashion in the 2D case and in a random fashion in the 3D case. The extent or resolution of the map are easily adjusted by changing the maximum concentration of parameter or the number of iterations in the script. Further improvements can be achieved in the script edition by the use of space-filling curves to optimize the screening time.

We managed to upgrade the observation chamber to allow for long term observation (up to 3 days without noticeable merging, evaporation or displacement of droplets) by optimizing the surface treatment, chamber structure, evaporation control. In the end, the chamber consists of two hydrophobic glass slides with 50 μm diameter glass spacers, and sealed by both NOA 81 photocurable glue and Araldite. Fluorescent monitoring is done with a confocal microscope, and for each droplets $n + v$ fluorescence levels are recorded, n being the number of barcoded parameters and v standing for the number of fluorescent observables reporting on the global output of the network. The biggest limitation to our present setup was the number of available laser wavelengths. Indeed, the confocal microscope we are using is equipped with 4 lasers, leading to a 3D mapping of only one fluorescent observable. The analysis of more complex network will require new strategies in terms of barcoding and/or monitoring, such as the use of XL barcoding fluorophores (i.e. long Stokes shift between excitation and emission) meaning that one laser gives two informations depending on the monochromator settings. Another option is the deconvolution of fluorescent signals [84]. In the next chapters we will use that setup on two different networks already studied in the laboratory: the bistable and the predator-prey systems.

Chapter 3

Mapping a bistable circuit

3.1 Introduction

3.1.1 bistability: two examples in biology

Bistability can be intuitively understood as the existence of two stable steady states within a dynamic chemical system. This means that the long term dynamic of a bistable system depends not only on its parameter, but also its history, a phenomenon known as hysteresis. Additionally, the system can often be perturbed by a chemical stimuli to jump from a stable state to another. Such switching in bistable systems can be seen as a storage function, or memory and has found uses in many biological processes.

As an example, Bhalla *et al.* showed that the Mitogen-Activated Protein Kinase signaling network (MAPK) -a major cellular pathway dealing with cell cycle and the whole secretory function- exhibits a bistable activity [85]. The simplified network (Fig.3.1, green panel) is a positive feedback loop in which the activity of MAPK is linked to the activity of Protein Kinase C (PKC) through the activation of multiple intermediates. The activation of MAPK or PKC induces a rise of the concentrations, which are then sustained. The system is therefore bistable, with a basal and activated state. The authors showed a simulated bistability plot by running the simulation with a fixed concentration of MAPK and reading the value of active

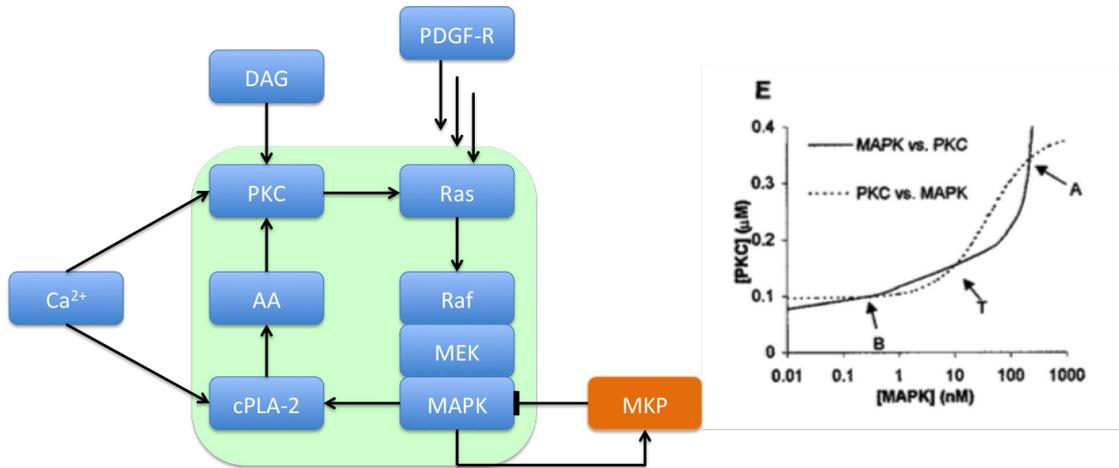


Figure 3.1: Positive feedback loop in MAPK pathway and bistability plot, from [85].

PKC. Then, they fixed the level of PKC and monitored the resultant MAPK levels. The intersection points are the stable steady states, either basal (B) or activated (A). The third point T is the threshold point that the system has to reach to flip its state. Together with hysteresis, i.e. the ability to maintain the steady state after removal of trigger, this network becomes a good example of information storage, since the system is in a defined state in presence of a triggering molecule.

Positive-feedback loops are not the only path to bistability : other connections such as double negative feedback or autocatalysis can also do the job [23, 86], but a system requires non-linearity coupled to the feedback circuit [87].

Another biological example is detailed by Robert *et al.* [88] about the switch in lactose operon regulation in *E.Coli*. Under certain conditions of initial growth rate and the concentration of LacI repressor protein, *E. Coli* may or may not turn on the transcription of lactose operon's genes.

The authors investigated the effect of these two epigenetic predisposing factors on the expression of the lactose operon in *E.Coli*. They vary the two parameters for the expression of lactose operon: the initial growth rate and the concentration of Lac1 repressor pro-

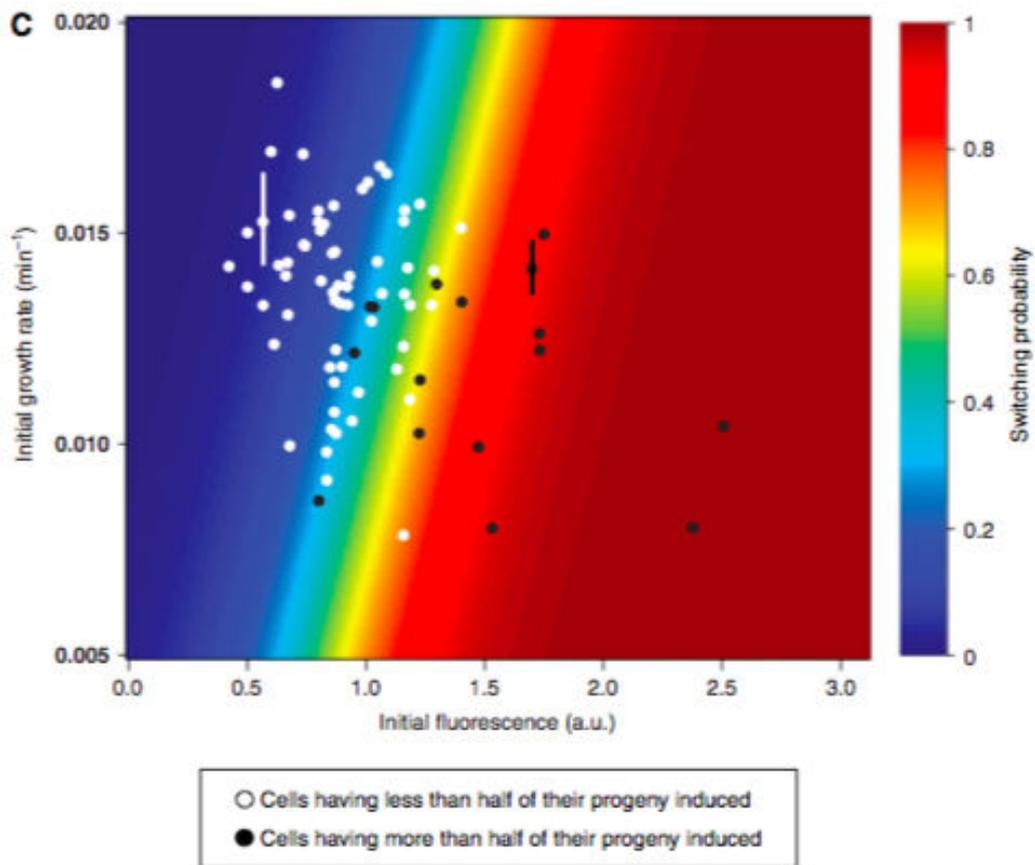


Figure 3.2: Switching probability of a cell as a function of its initial growth rate and initial fluorescence, from [88].

tein -inhibiting the production of lac operon- and monitor the putative expression of lactose operon in E.Coli. They finally map a diagram showing the result (expression or inhibition of the lactose operon) in function of both factors and the resulting switch probability (Fig.3.2).

3.1.2 Introduction to bifurcation theory

Many levels of modeling are available to represent a dynamic system like a gene regulatory network. At the higher level, Boolean or Bayesian networks capture the overall function of a network, but their abstraction omits -almost by definition- important lower levels mechanisms (such as kinetics and thermodynamic aspects). To faithfully encompass kinetics effects and understand how changing kinetic parameters will affect the system, a lower level model made of parametrized Ordinary Differential Equations (ODE) is more appropriate. (See [89] for an extensive comparison between modelization methods). The limitation of the modelization by ODE is the increasing number of parameters as soon as the network topology expands. For a given network, each behavior (*e.g.* bistability, monostability or oscillations) can be characterized as a family of solutions of the ODE, each family corresponding to a set of parameters, that are different from the variables. In other words, the goal of the molecular engineer is to find the parameters that give rise to a desired dynamic.

Let us consider a dynamical system $\dot{x} = f(x, \mu)$, where x are the dynamic variable in the phase space \mathbb{R}^n , μ are the parameters in the parameter space \mathbb{R}^m and $f : \mathbb{R}^{m,n} \rightarrow \mathbb{R}^n$ is the function defining the dynamical system. We define the phase portrait as the partition of the phase space in orbits with their direction in respect to time. These orbits can be equilibria, periodics... depending on the solutions they represent. The phase portrait holds all the states the system tends to (*i.e.* all orbits). It can be seen as «a flow of some fluid, where the orbits show the paths of liquid particles as they follow the

current» [90].

Two systems whose phase portraits are homeomorphic (i.e. one shape can be continuously deformed to become the second one) are called topologically equivalent. A bifurcation occurs when -for a given dynamical system- a change of the parameter leads to a topologically non-equivalent phase portrait. A bifurcation diagram is then a lamination of the parameter space of a dynamical system, representing the phase portraits for each layer.

Many types of bifurcation exist, depending on their conditions of occurrence and the kind of change they characterize in the phase portrait. In this manuscript we consider in particular three types of bifurcations. First, the Andronov-Hopf bifurcation, which is characterized by the apparition or disappearance of a periodic orbit when changing the value of the stable point. The second type is the saddle-node bifurcation, which corresponds to the collision and disappearance of two equilibrium points in the phase portraits, and delimits here the area of monostable and bistable behaviors, and finally a particular kind of bifurcation for which the eigenvalues of the equilibrium are null.

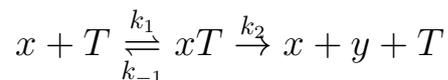
3.1.3 Two PEN-DNA-toolbox examples

We now explain in more details how bistability can be implemented in the PEN-DNA toolbox. We showed that bistability is a very common feature in biological networks and data mining has exposed many simple motifs that can support bistability in the biological world [91, 92]. A common point is that bistability requires at least one positive feedback loop. The modularity of the PEN-DNA toolbox (which allows the connection of positive and negative feedback loops) makes it possible to design a bistable circuit, by connecting two inversion modules in an antagonistic way therefore creating a bistable loop [45]. Each autocatalyst stimulates its own production and the production of the other catalyst's inhibitor. This system

a priori leads to four possible final states (no species surviving, only one species surviving, both species surviving), depending on the initial conditions (Fig.3.3, top). The implementation in DNA strands compatible with the PEN-DNA toolbox was conducted as follows (Fig.3.3, bottom): trigger α binds to the template $\alpha to \alpha$ (named the input region) that is designed to generate autocatalysis (its sequence is a repeat of the complementary strand of the trigger). The polymerase then elongates the trigger into an adduct trigger-template, which exhibits a recognition site for the nicking enzyme. After cutting, two α strands are released. In the same way, the autocatalyst $\beta to \beta$ produces β . The trigger β can also interact with the template $\beta to i \alpha$ to generate the $i \alpha$, the inhibitor of the autocatalytic template $\alpha to \alpha$. Likewise, $\alpha to i \beta$ produces $i \beta$, the inhibitor of $\beta to \beta$.

Adrien Padiac successfully implemented this network and used it as a footboard for the implementation of more complex networks, such as a binary counter [45]. An alternative design of a bistable system is in principle possible, which does not rely on cross inhibition and requires only one autocatalytic module [93] (Fig.3.4). The difference with the previous bistable system lies in an undegradable pool of triggering agent γ that ensures a constant production of a strand β , which in turn produces $i \alpha$, the inhibitor of α . On the other hand, α is produced in an autocatalytic way, thus generating $i \beta \gamma$ the inhibitor of β production. Such a system is compatible with the DNA toolbox and the network is quite easy to design (Fig.3.4, bottom), the trigger γ being protected from the exonuclease by addition of phosphorothioates modifications at the 5' end of the strand).

Despite being rudimentary, the one-autocatalyst bistable circuit is less robust than the two autocatalysts one. Indeed, let us consider a simple mathematical model based on Michaelis-Menten kinetics on the amplification of a template T (xtoy):



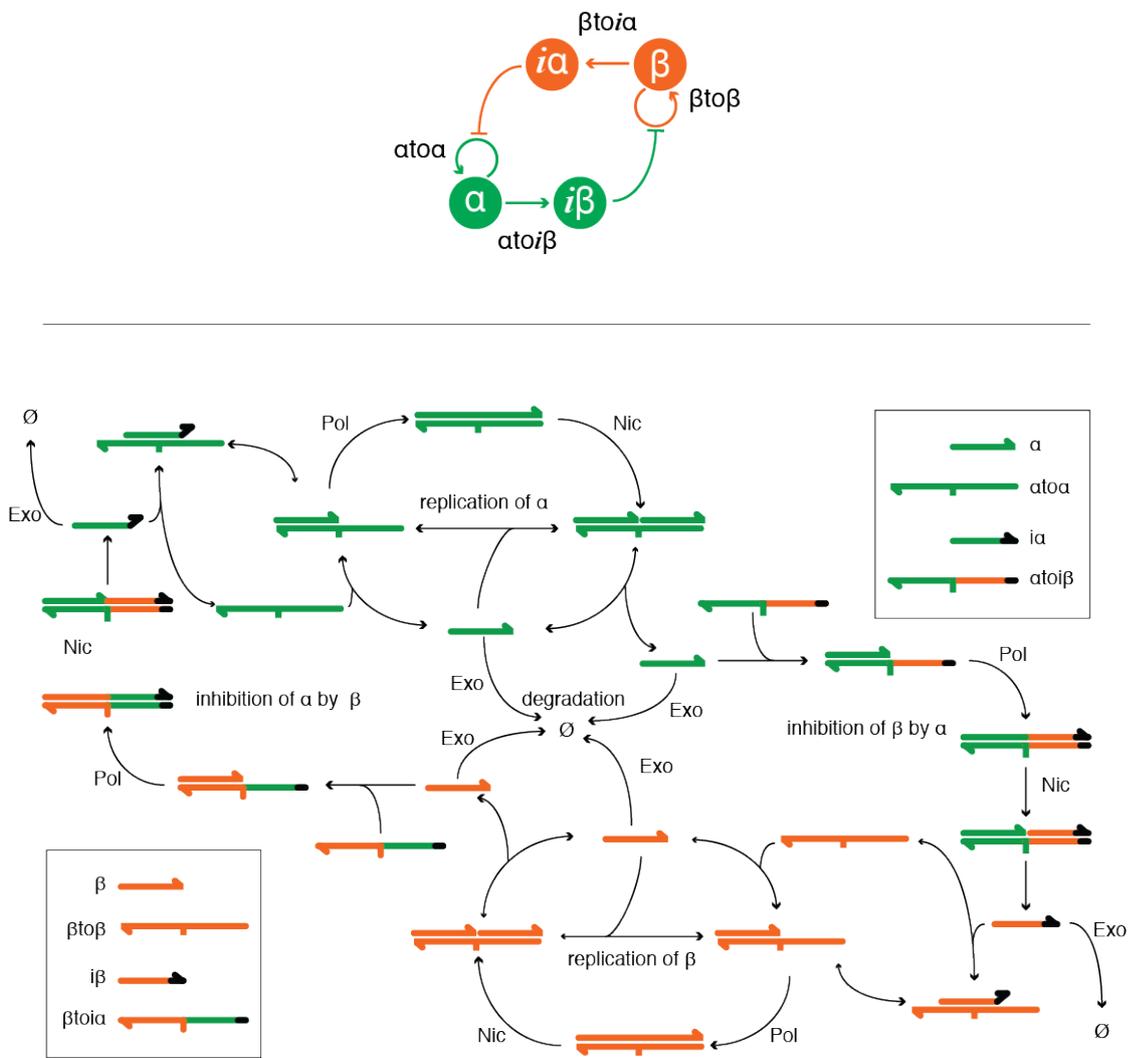


Figure 3.3: Bistable circuit with two autocatalysts : reactions network and implementation with PEN-DNA toolbox

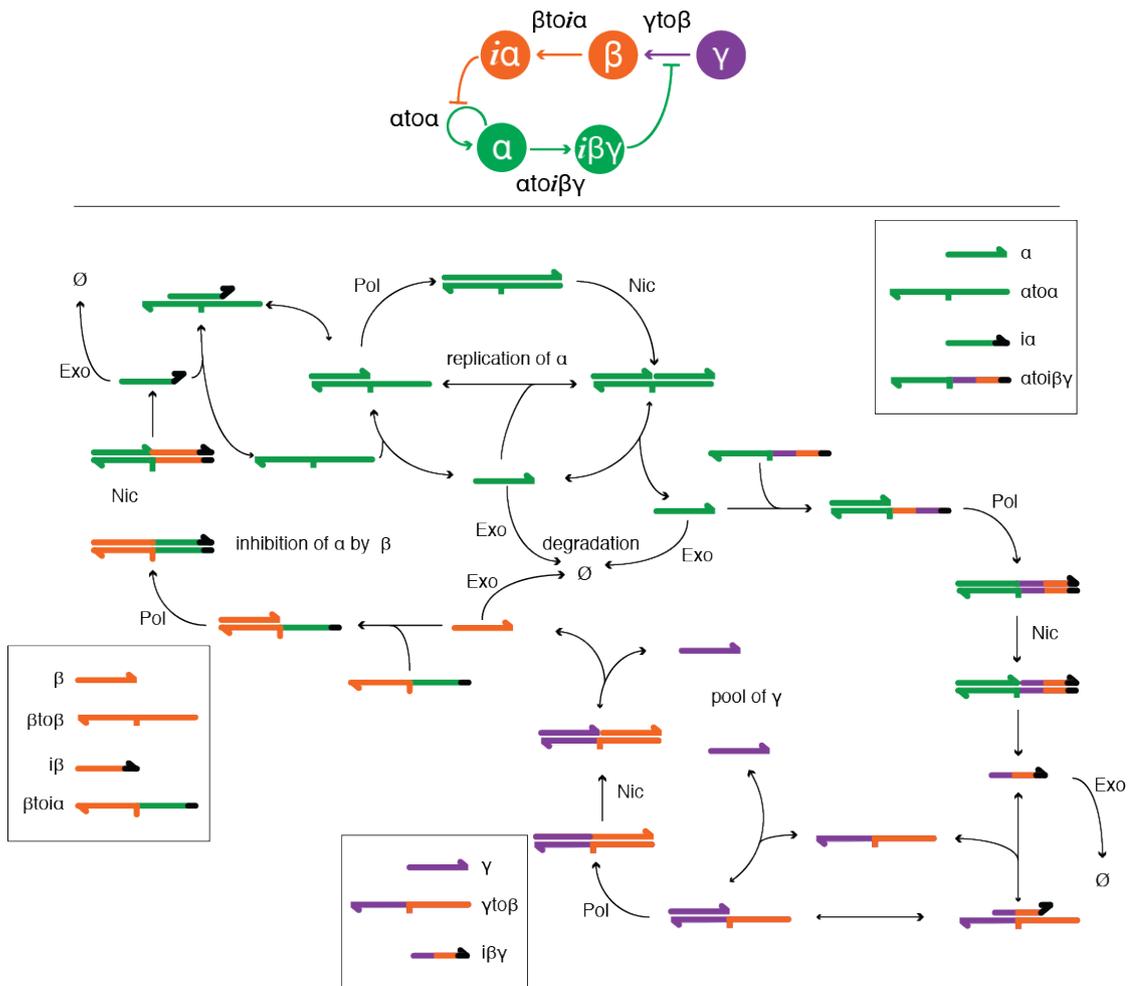


Figure 3.4: Alternative bistable circuit : reactions network and corresponding implementation in PEN-DNA toolbox framework

The second reaction involves complex transformations (two enzymatic reactions plus multiple de-hybridizations) but here, we assume a first order transformation:

$$\frac{dy}{dt} = \frac{k_2 \cdot T \cdot x}{K_x + x}; K_x = \frac{k_{-1} + k_2}{k_1}$$

Considering now the reactions involved in the bistable system as in Fig.3.3, thus assuming that $k_2 \lll k_{-1}$, K_x is equivalent to the dissociation constant of x on T, the inhibiting strand i_{xy} is a competitive inhibitor and the degradation is a first order kinetics, we can write the equations for the species x (which are α , β , i_α and i_β for the bistable) with the form:

$$\frac{dx}{dt} = \frac{k_x \cdot T_x \cdot x}{K_x(1 + \frac{x}{K_x} + \frac{i_x}{K_{ix}})} - D \cdot x$$

After adimensionalisation and Linear Stability Analysis (LSA), it appears that the bistable circuit with two autocatalysts (Fig.3.5, a) displays a larger bistability area than the mono-autocatalyst architecture (Fig.3.5, b). Therefore the former is likely to more robust experimentally than the latter.

As a consequence, we chose the double autocatalysts network as the first circuit to be mapped using our droplet-based scanning platform. First, this is a proof of concept for the device, since an experimental analysis has already been conducted in bulk. Secondly, a detailed mapping of bifurcation would bring about new information about the inner working of the circuit and reveal the optimal regime of bistability.

The experimental protocol can be split into three phases: first, the preparation of the fluidic circuit (device, tubing, connections), microscopy platform and all submixes. Then, the droplet generation (setting the invasion pressures, running the script and collection). Finally, crafting the observation chamber and follow-up of the system. This protocol has been optimized -timewise- but still requires some

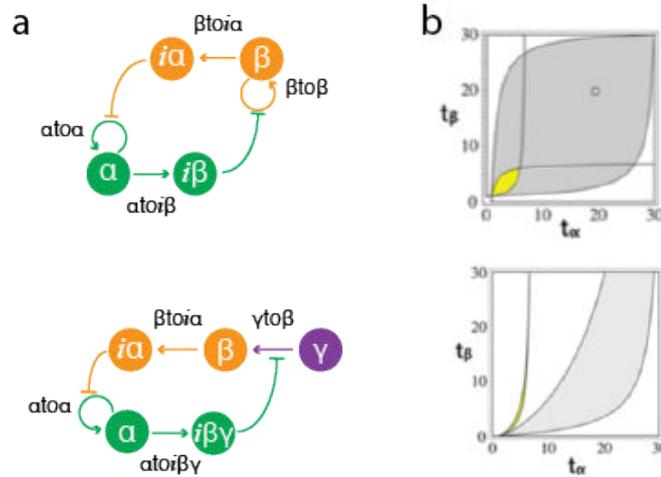


Figure 3.5: Stability of bistable networks : comparison of computed bifurcation diagrams. Networks (a) and bifurcation diagram showing bistability domains in the $\{t_\alpha, t_\beta\}$ space (b), for a ratio activator/inhibitor binding constant of 20 (yellow) or 100 (grey). Charts from Adrien Padirac.

time. During the moment the submixes are done and the moment the droplets flows out of the device in rapid succession, the enzymes used by Padirac *et al.* can process DNA strands before encapsulation. This effect gives imprecisions -or errors- in the corresponding bifurcation diagram. Fortunately, we use thermophilic exonuclease ttRecJ (expressed and purified in the laboratory using the protocol of Yamagata *et al.* [94]), so that at room temperature its activity is lowered . However, to mitigate pre-incubation activity as much as possible, we decided to change the polymerase from the original setup for Bst 2.0 WarmStart DNA Polymerase. This polymerase is an engineered version of Bst DNA Polymerase with additional warm-start functionality (this function are obtained through mutagenesis, or the use of inhibitors such as aptamers or antibodies [95, 96])

DNA strands were ordered from Biomers (Germany). The sequences are indicated in Table4.3. For fluorescent reporting, we use nucleobase quenching [97] to report on the levels of α and β . It uses the fact the fluorescence of a dye attached to the 3' or 5' end of a ssDNA shifts up or down when this strand becomes double-stranded.

	Sequence 5' → 3'	3' modification
$\alpha to \alpha$	C*C*A*AGACTCAGCCAAGACTCAG	Phosphate
$\beta to \beta$	A*A*C*AGACUCGAAACAGACTCGA	Phosphate
$\beta to \alpha$	A*C*T*CAGCCAAGACAACAGACTCGA	TAMRA
$\alpha to \beta$	A*C*T*CGAAACAGACCCAAGACTCAG	FAM
α	CTGAGTCTTGG	
β	TCGAGTCTGTT	

Table 3.1: Sequences of DNA strands used in the bistable switch. Stars correspond to a phosphorothioate modification, to prevent degradation by the exonuclease ttRecJ.

We attach a TAMRA dye (positive shift upon binding of β) to the template $\beta to \alpha$ and a FAM dye (negative shift upon binding of α) to the 3' of the template $\alpha to \beta$. It must be kept in mind that fluorescence shifts do not vary linearly with concentration of free triggers, although their relation is expected to be monotonic.

3.2 Results

3.2.1 Reaction assembly

For generating the diagram of the bistable switch (Fig.3.9), the master mix contains the following reagents: Tris-HCl (45 mM, pH~8.4), NaCl (50 mM), KCl (10 mM), Mg^{2+} (from $MgCl_2 + MgSO_4$, 8 mM in total), $(NH_4)_2SO_4$ (10 mM), dNTPs (400 μM each, New England Biolabs, NEB), Synperonic F108 (0,1% v/w, Sigma-Aldrich), Netropsin (2 μM , Sigma-Aldrich), Dithiothreitol (3 mM, Sigma-Aldrich), BSA 9000S (500 $\mu g.mL^{-1}$, NEB); the enzymes: polymerase Bst 2.0 WarmStartTM (4 units.mL⁻¹, NEB), exonuclease ttRecJ (20 nM), and nickase Nt.BstNBI (100 units.mL⁻¹, NEB), the inhibitor-producer templates ($\beta to \alpha$ and $\alpha to \beta$, 20 nM each). We use a Warmstart version of the polymerase to mitigate production delays associated with the microfluidic protocol.

This mix is then split into 2 sets of 3 tubes and completed with parametric species and their fluorescent barcodes (summarized in

	Channel 1:		Channel 2: Buffer adjust		Channel 3	
	$[\alpha] > [\beta]$	$[\beta] > [\alpha]$	$[\alpha] > [\beta]$	$[\beta] > [\alpha]$	$[\alpha] > [\beta]$	$[\beta] > [\alpha]$
Barcode 1: Alexa 647	200 nM	200 nM	20 nM	20 nM	20 nM	20 nM
Barcode 2: Cascade Blue	5 nM	5 nM	5 nM	5 nM	200 nM	200 nM
$[\alpha]$	10 nM	0,1 nM	10 nM	0,1 nM	10 nM	0,1 nM
$[\beta]$	0,1 nM	10 nM	0,1 nM	10 nM	0,1 nM	10 nM
$[\alpha to \alpha]$	500 nM	500 nM	-	-	-	-
$[\beta to \beta]$	-	-	-	-	500 nM	500 nM

Table 3.2: Submixes composition for diagram mapping

Table3.2):

- $[\alpha to \alpha] = 500$ nM, $[\beta to \beta] = 0$ nM barcoded with 200 nM of Dextran Cascade Blue (10,000 MW, Life Technologies)
- $[\beta to \beta] = 0$ nM, $[\beta to \alpha] = 500$ nM barcoded with 200 nM of Dextran Alexa Fluor 647 (10,000 MW, Life Technologies)
- $[\alpha to \alpha] = 0$ nM, $[\beta to \beta] = 0$ nM, 5 nM of Dextran Cascade Blue, 20 nM of Dextran Alexa Fluor 647 (this base level of Dextran raises their minimal fluorescence over the limit of detection of the microscope and thus improves quantification for low concentrations of template) and water for volume compensation.

The sets differ by their $\frac{[\alpha]}{[\beta]}$ ratio, corresponding to an initial state $\alpha\beta=01$ (0.1 nM of α and 10.1 nM of β) or $\alpha\beta=10$ (10.1 nM of α and 0.1 nM of β). For the independent 1D scans we use only 2 two tubes ($[\beta to \beta] = 0$ or 500 nM).

For droplet encapsulation, the surfactant was prepared from a perfluoropolyether carboxy-terminated polymer (Krytox, DuPont) and a polyetheramine (Jeffamine M1000, Huntsmann) based on the synthesis scheme described in [75]. The continuous phase is a mix of HFE 7500 oil (3M) with 2% (w/w) of surfactant.

Experiment		BD 01: $[\beta] > [\alpha]$	BD 10: $[\alpha] > [\beta]$	1D scan	
Resolution (pixel)		512 x 512	512 x 512	320 x 320	
Objective mag. (N.A.)		20x (0.75)	20x (0.75)	20x (0.75)	
C.A. (μm)		120	120	120	
Lasers	405 nm	%	40	40	-
		PMT (V)	708	708	-
		VBF	425-460 nm	425-460 nm	-
	473 nm	%	30	30	55
		PMT (V)	740	755	720
		VBF	495-555 nm	495-555 nm	495-555 nm
	559 nm	%	30	30	30
		PMT (V)	644	644	659
		VBF	BA575-620	BA575-620	BA575-620
	635 nm	%	20	20	10
		PMT (V)	668	596	595
		VBF	BA655-755	BA655-755	BA655-755

Table 3.3: Acquisition parameters for confocal microscopy. The percentages indicate the laser power, “C.A” refers to the confocal aperture and “VBF” to variable barrier filter.

3.2.2 Data analysis

3.2.2.1 Data collection

Droplets were imaged with a confocal laser microscope (Olympus Fluoview FV1000) mounted on a IX-81 chassis and equipped with a xy stage Optosigma Bios-206T, 4 lasers (405, 473, 559 and 635 nm) and a 20x (NA=0.75, Fig.3.9) or 10x (NA=0.40, Fig. 3) objective. The droplet chamber was incubated on stage with a heat plate Tokai Hit MAT-1002ROG-KX. Droplets were imaged at room temperature after incubation at 42°C. The observation chamber was filled as previously described and was turned upside down to reduce the number of glass layers in the optical path and the total thickness [80]. After scanning, images were stitched by the Olympus software. Table 3.3 displays the conditions of acquisition for confocal microscopy for the four fluorescences recorded: Cascade Blue, FAM, TAMRA and Alexa Fluor 647.

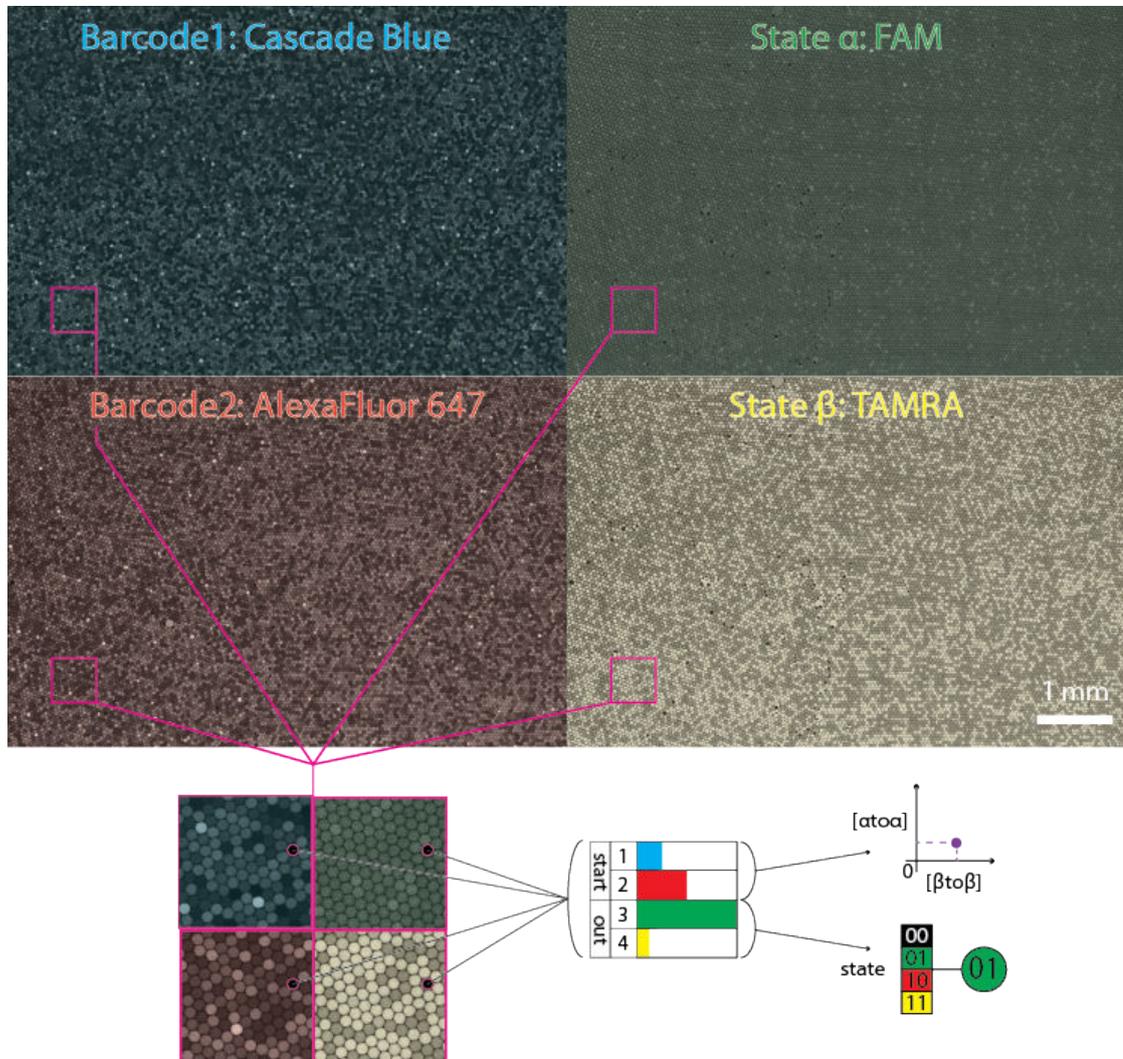


Figure 3.6: Collection of fluorescence levels. For each droplet, we extract the intensities for the barcodes (Cascade Blue and Alexa Fluor 647) and the reporters (FAM and TAMRA). The barcode level gives the initial conditions i.e. the coordinates in the phase space, while the reporter levels render the steady-state of the system. Images show the array after 11h at 42°C.

3.2.2.2 Image processing

All images were processed with Mathematica software. We scanned a large portion of the droplet chamber (typically 15x15 fields of view) in order to compute a flat-field correction and reduce vignetting. To that end, we scanned, averaged and gaussian filtered ~200 fields of view, which yielded a correction mask whose maximum intensity was set to 1 in each channel. We then divided pixels-by-pixels and channel-by-channel the scanned fields of view by the correction mask to compensate vignetting.

Droplet detection The detection of the droplets requires an image with a good contrast between droplets. For this we used the TAMRA channel (Fig.3.7), showing a high signal-to-noise ratio and bimodal distribution of intensities. To this raw image we applied a Gaussian filter with a radius of 70% the droplet radius. We then applied the *MaxDetect* function, which gives an image consisting of the extended maxima (sets of connected pixels brighter than their surrounding). We then applied a Gaussian filter of a few pixels and detected morphological components, which yields the centers of the droplets. We removed centers which were too close to their neighbors or located near the image borders. We then extracted the intensities of the 4 fluorescence channels by integrating over the whole droplet.

Processing of fluorescence The next step is the manual rescaling of fluorescence values. To do so, we used the apexes of the calibration axes (Fig.3.8, a) to convert the coordinates in the barcode space ($\text{fluO}_{\text{Alexa}}$, $\text{fluO}_{\text{Cascade}}$) into positive coordinates in the concentration space ($[\alpha\text{to}\alpha]$, $[\beta\text{to}\beta]$). The same way, we normalized the values of $\text{fluO}_{\text{TAMRA}}$ by manually defining the minimum intensity to 0 and the maximum intensities to 1, ruling out droplets lying outside those bonds. Finally, the values of fluO_{FAM} were normalized by a maximum intensity to 0 and the minimum intensity to 1 (because of the negative

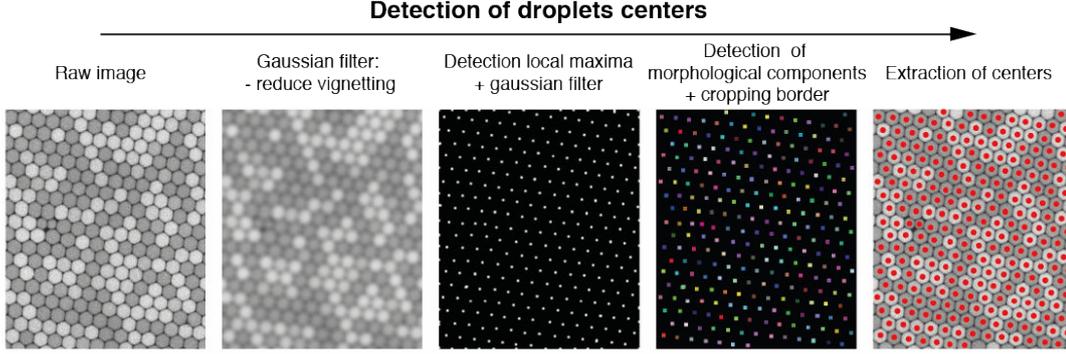


Figure 3.7: Image processing workflow

fluorescence shift in this case).

The direct correlation between the fluorescence values and the concentrations of α and β is tempting, but the relationship between them is monotonic but not linear. Indeed, the saturation of the fluorescence shift is reached when the triggers exceed their binding K_m (100 nM, the threshold being probably lower because of the stabilizing effect of stacking interactions, or elongation by the polymerase). This indirect and saturating reporting is sufficient enough to distinguish a high state from a low state as showed by the bimodality of the fluorescence histogram in Fig.3.8, c.

This whole process leads to a set of points in a 4D space ($[\alpha to \alpha]$, $[\beta to \beta]$, $fluo_{FAM}$, $fluo_{TAMRA}$) which we projected into the 2D concentration space e ($[\alpha to \alpha]$, $[\beta to \beta]$) to access the raw bifurcation diagram (Fig.3.9, a). The color of a point indicating the level of $fluo_{FAM}$ and $fluo_{TAMRA}$. We then smoothed those raw diagrams by taking for an arbitrary point ($[\alpha to \alpha]$, $[\beta to \beta]$) the local median ($fluo_{FAM}$, $fluo_{TAMRA}$) of its 20 nearest neighbors in parametric space (Fig.3.9, b).

A rapid look at the raw diagram is enough to clearly identify the bifurcations of steady states. The relative positions and areas of the regions follow the expected scheme: the region 10 of high α and low β corresponds to circuits with high $\alpha to \alpha$ and low $\beta to \beta$, expanding when the initial α is increased. The superimposition of

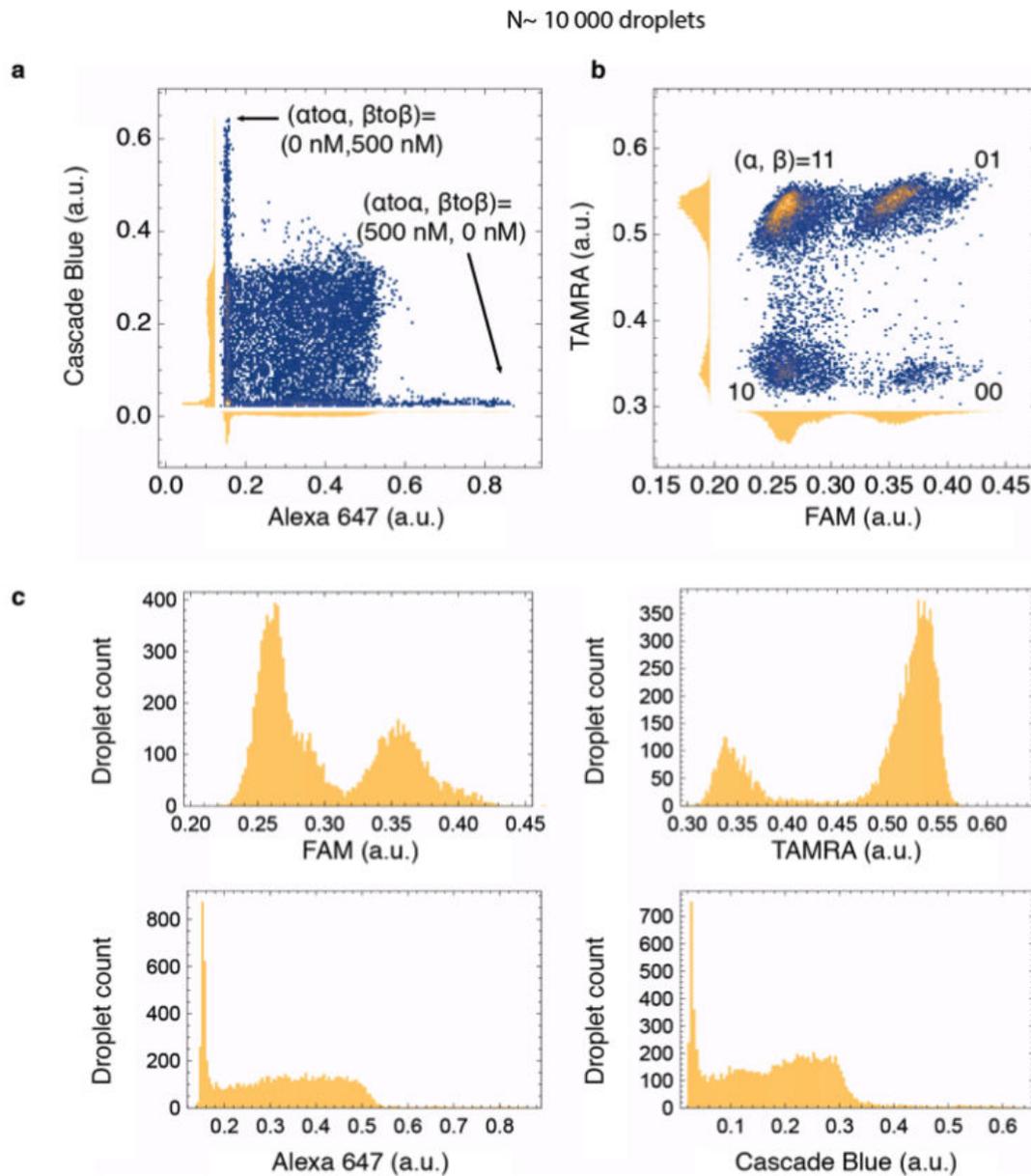


Figure 3.8: Distribution of raw fluorescence in droplets for a bistable circuit started in $\alpha\beta = 01$. (a) Coverage of the parameter space. The density histogram shows the fluorescence of droplets in the (Alexa647, Cascade Blue) space. The microfluidic pressure script also explores vertical and horizontal arms along the axis in addition to the square mapping. Their apexes serve for calibration, indicating the fluorescence of droplets with a maximal amount of template $\alpha\text{to}\alpha$ or $\beta\text{to}\beta$. (b) Density histogram of (FAM, TAMRA), which are proxies of (α, β) . Low levels of α translate into high levels of FAM, while high levels of TAMRA translate to high levels of β . The histogram of the projected distribution is shown on the side. (c) 1D histograms of the individual fluorescence channels.

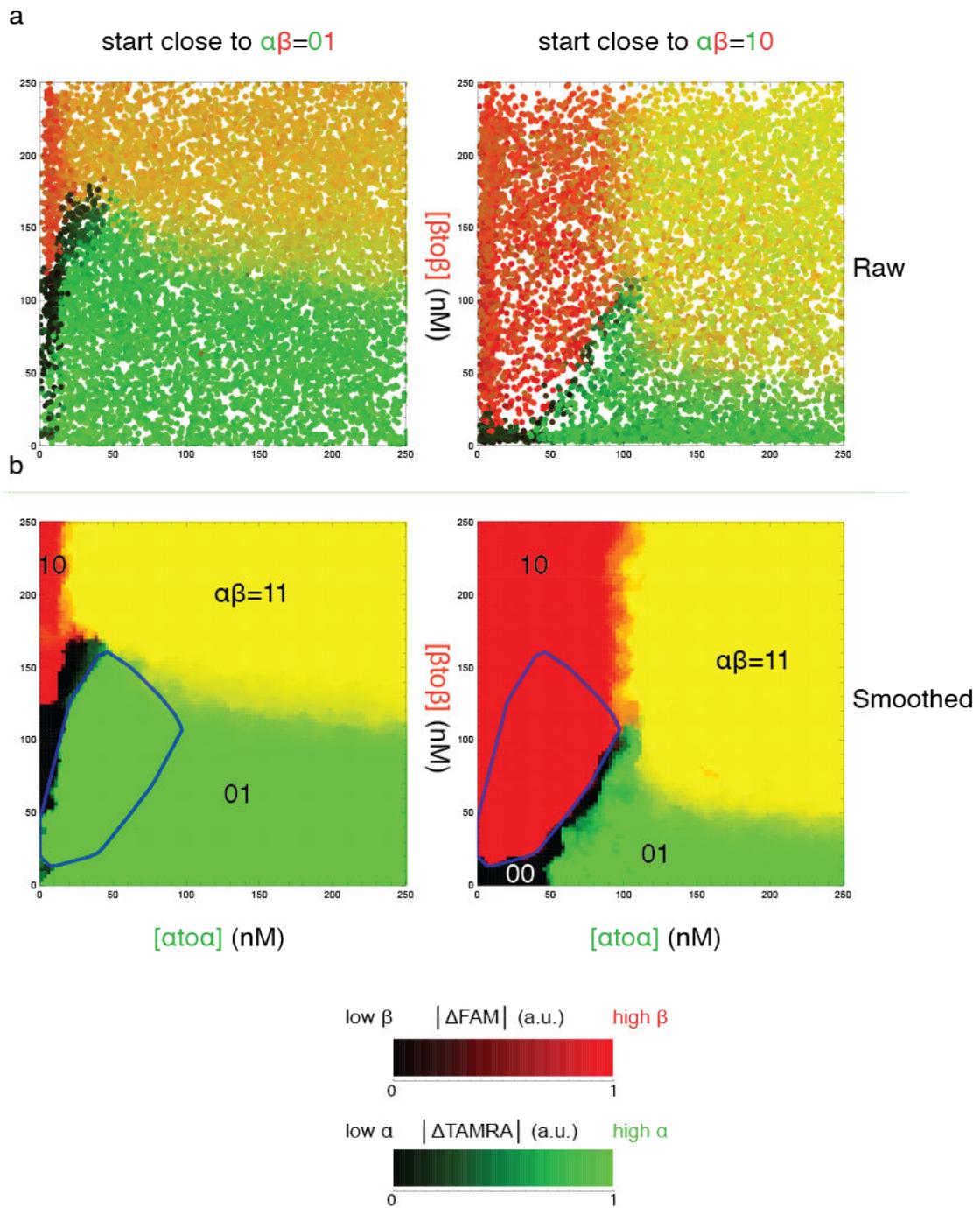


Figure 3.9: Experimental bifurcations in the bistable circuit. (a) Steady state reached in the (α, β) space vs. the parameters, after 11 hours at 42°C . The colors encode fluorescence shifts of reporters and reflect the levels of α and β (see color bar at the bottom). The initial condition $\alpha\beta = 10$ corresponds to $[\alpha]_0 = 10$ nM and $[\beta]_0 = 0.1$ nM (and accordingly for $\alpha\beta = 01$). (b) Diagrams obtained by locally smoothing and thresholding the fluorescence shifts of a.

the two diagrams displays a bistable region where the circuit shows hysteresis (Fig.3.9,b). The frontiers' sharpness (as assessed by the magnitude of the gradient of α or β) also agrees with qualitative predictions from bifurcation theory [98]: the fold (or saddle-node) bifurcations connecting the bistable to monostable areas (sudden loss of stability of one state) appear as abrupt frontiers. By contrast, transcritical bifurcations (collision and exchange of stability between two fixed points, e.g. green to yellow transitions in Fig.3.9, a and b) materialize as a smooth gradient.

We performed an independent 1D scanning along two vertical lines to confirm the trends seen on the diagram, thus eliminating about potential artifacts from barcoding accuracy (Fig.3.10).

3.3 Interpretation

3.3.1 Simulations

A realistic prediction of the evolution of concentrations requires the integration of many side reactions and intermediate species, whose rates and concentrations are often not directly accessible experimentally. Even when the underlying chemistry of individual reactions is faithfully understood, connecting them in a network can result in counter-intuitive outcomes. As we saw in the introduction, mathematical tools such as Linear stability analysis can link the dynamics of the network with its topology and parameters. But analytical approaches often fail for large and complex systems and their scope is even more limited when the underlying chemistry is not precisely known. In well-mixed systems, the potential number of interactions grows like the square of the number of species; keeping track of all these interactions and their kinetic consequences quickly becomes impossible for the human brain. It gets even worse when global effects like enzymatic saturation and competition are taken into account.

In order to model *in silico* the PEN-DNA, the DACCAD software

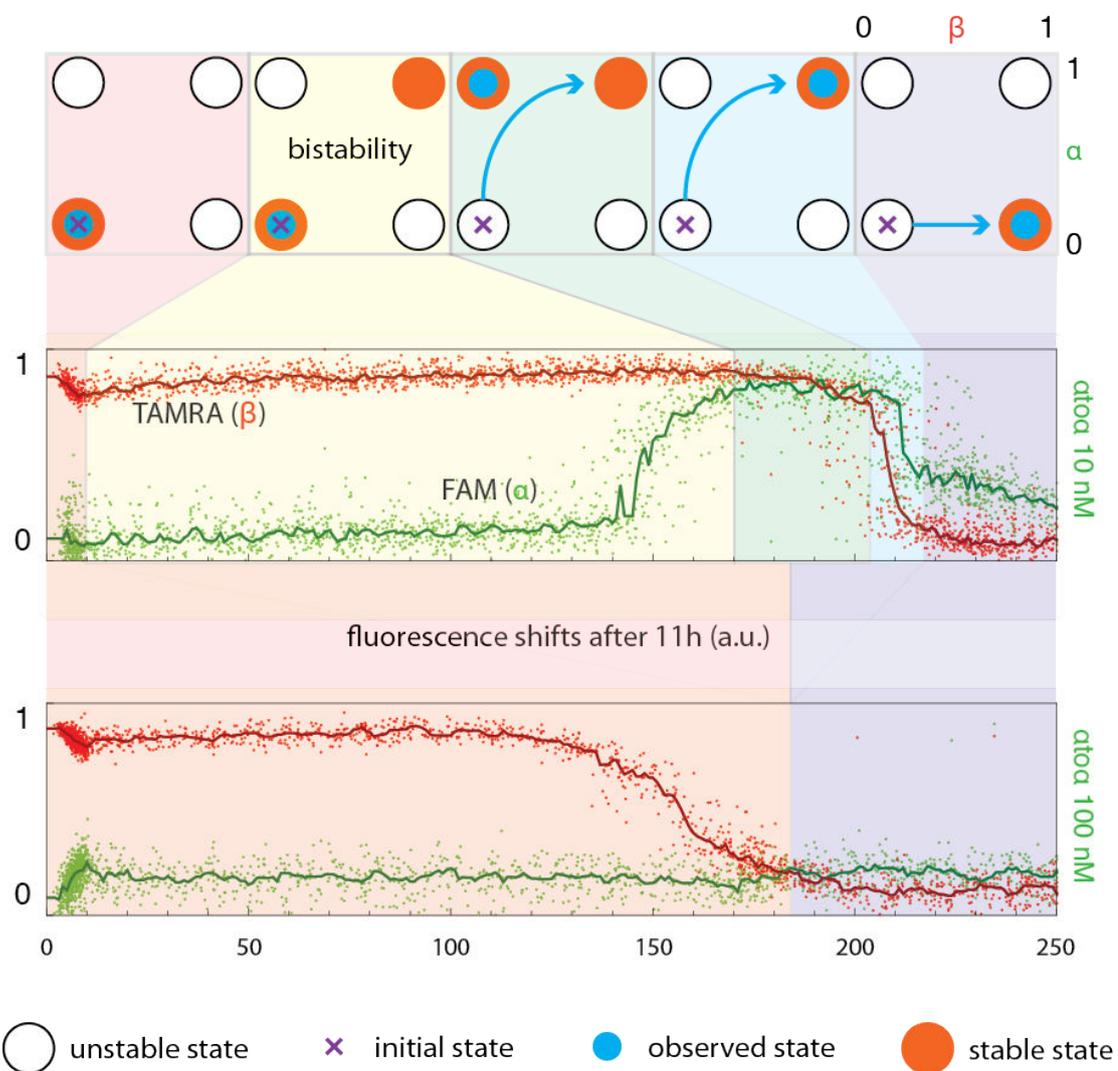


Figure 3.10: 1D experiment : line across the diagram, started from $\alpha\beta = 10$, along the axes in parameter space : $[atoa] = 10$ nM (~ 3700 droplets) or 100 nM (~ 3000 droplets). The continuous line is the local median.

was developed [99] to (i) provide a straightforward graphical interface allowing to compile the dynamic of a network from its topology and parameters, (ii) eliminate routine and error prone tasks such as ODE writing and solving, (iii) and assist optimization by allowing easy tuning of parameters to match a desired dynamic (although the function is still limited).

We modeled -using the DACCAD framework- at the domain level the interactions between signal strands (short, produced and degraded endogenously) and template strands (longer, exogenous strands protected against exonuclease degradation). A given signal strand is either completely free or completely double-stranded to its complementary domain. Mismatch, partial hybridization and signal/signal or template/template interactions are ignored. Fig.3.3 shows the main reactions.

The full model comprises 26 separate species interacting through 60 chemical reactions modulated by 39 parameters. Association/dissociation kinetics and enzymatic parameters are taken from Padirac *et al.* [45]. Strand-displacement kinetics, dangle slowdown and stacking are derived as in Aubert *et al.* [99]. From this basis, we upgraded the models by addition of other factors, going closer to a full, complete description of all interactions within the network. The 3 models only differ by their enzymatic kinetics. For an enzyme *enz* acting on substrate *si* the kinetics is defined as:

$$\begin{aligned}
 M1 : & \quad \frac{V_{enz}}{K_{enz,s_i}} \cdot [s_i] \\
 M2 : & \quad \frac{V_{enz}}{K_{enz,s_i} + [s_i]} \cdot [s_i] \\
 M3 : & \quad \frac{V_{enz}}{K_{enz,s_i} \cdot (1 + \sum_{s_j \in \text{substrates of } enz} [s_j]/K_{enz,s_j})} \cdot [s_i]
 \end{aligned}$$

with V_{enz} denoting the maximum rate for enzyme *enz* and $K_{enz,si}$ the Michaelis constant for *si* with respect to enzyme *enz*.

In other words, M1 is first order, M2 follows the classical Michaelis-Menten law while ignoring competition effects, and M3 generalizes it by including the load of all substrates of an enzyme. For each model, we iteratively optimize its parameters to fit the experimental bifurcation diagrams of Fig.3.9 (the target function), the set of optimized parameters depending on the model (e.g. M1 does not have enzymatic saturation, thus does not optimize Michaelis parameters).

We assign a fitness to a set of parameters by comparing the bifurcation diagram it predicts to the experimental diagram. Evolutionary algorithms are known to perform particularly well in such context as they require few assumptions on the search space. We use CMA-ES [100, 101] which locally approximates the fitness gradients through stochastic sampling and evaluates the fitness sensitivity in order to avoid the pitfalls of premature convergence and flat fitness landscape. We further constrain the search to a bounded region by assigning the worst fitness to individuals whose parameters exceed specified bounds (typically 10-fold variations for enzymatic parameters and 100-fold for other parameters). Penalizing such individuals allows us to skip unnecessary ODE evaluation (which gets prohibitively slow for unrealistic parameters). We perform 10 runs for each model, one run typically comprising 2000 evaluations (i.e. diagram evaluation) over 40 generations of 50 individuals. A shared set of 6 parameters (stabilization effect of stacking on all 4 templates, as well as the ratios *pol/exo* and *nic/exo*) is optimized for all models. Five additional parameters (Michaelis constants for all enzyme/substrate combinations) are optimized for M2 and M3 (Fig.3.11).

Optimized parameters for each model, along with their default values, are shown in Table3.4. We accelerate the evaluation of bifurcation diagrams -a major computational bottleneck- by sparsifying. First, we coarse-grain the experimental diagrams of Fig.3.9, b (going from a $\sim 100 \times 100$ binning to a 20×20 binning). Secondly, we sparse these coarse diagrams by removing large and homogenous regions

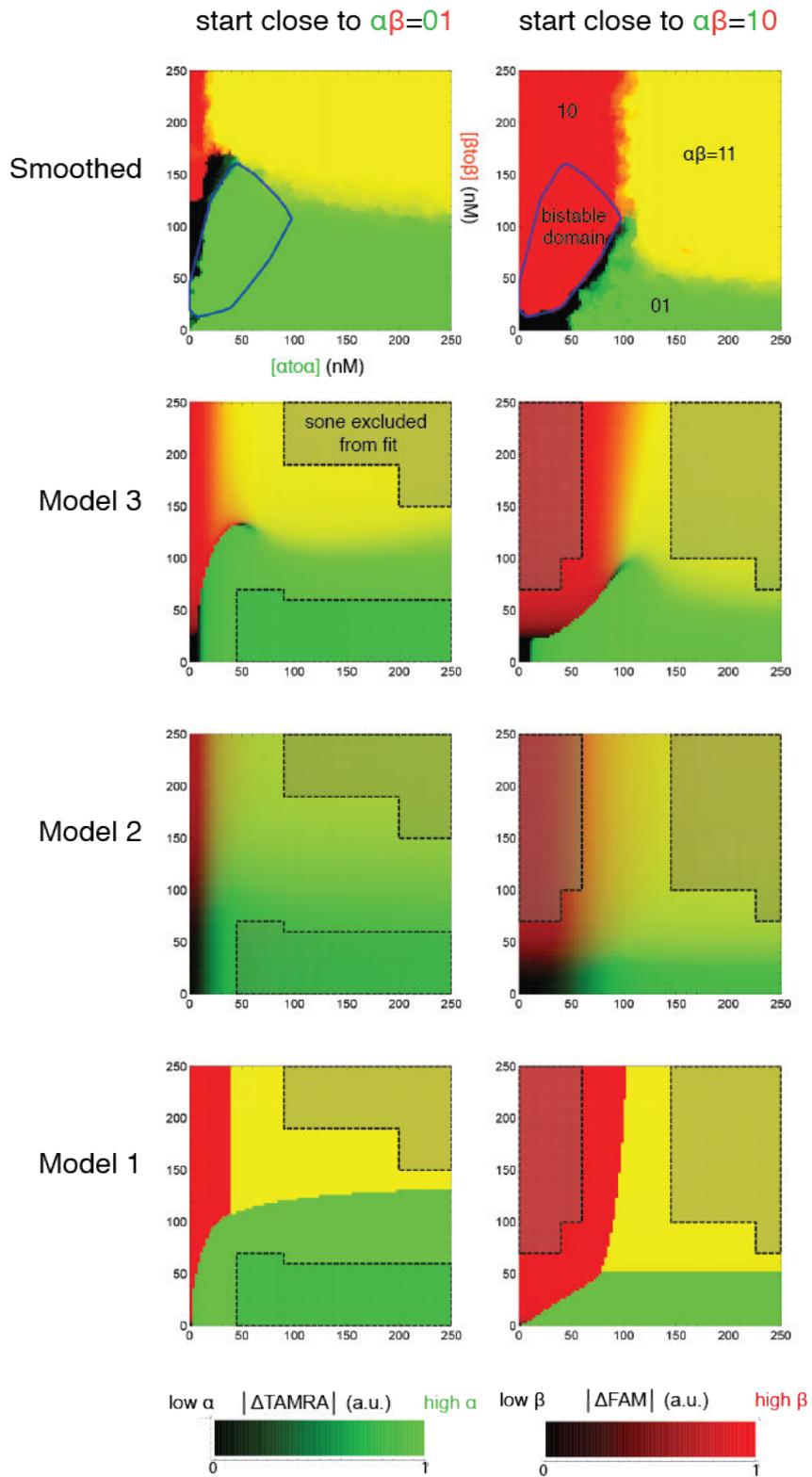


Figure 3.11: Results of simulated bifurcation diagrams by three models

around the frontiers (indicated in Fig.. 2f, g, h), going from 400 points to ~ 200 points. Sparsing speeds up diagram computation and penalizes models that poorly approximate the positions of bifurcations. The sparse diagram is simulated for a model time of 11 hours. We then evaluate the diagram fitness as follows:

$$fit = \sqrt{\frac{\sum_{([\alpha to \alpha], [\beta to \beta]) \in conditions} dist([\alpha to \alpha], [\beta to \beta])^2}{\#conditions}}$$

with

$$\begin{aligned} dist([\alpha to \alpha], [\beta to \beta])^2 = & ([\alpha to i \beta]_{l,exp,\alpha} - [\alpha to i \beta]_{l,sim,\alpha})^2 \\ & + ([\beta to i \alpha]_{l,exp,\alpha} - [\beta to i \alpha]_{l,sim,\alpha})^2 \\ & + ([\alpha to i \beta]_{l,exp,\beta} - [\alpha to i \beta]_{l,sim,\beta})^2 \\ & + ([\beta to i \alpha]_{l,exp,\beta} - [\beta to i \alpha]_{l,sim,\beta})^2 \end{aligned}$$

where $[x]_l$ represents the concentration of loaded template (rescaled between 0 and 1 to match the fluorescence shift) and $\#conditions$ is the total number of points in the 2 sparsed bifurcation diagrams. We thus use the square distance between the expected concentration of loaded inhibiting template (*i.e.* those with trigger, which are contributing to most of the fluorescence) and their simulated concentrations for the given initial concentration of templates $\alpha to \alpha$ and $\beta to \beta$. Note that we sum over two different initial conditions ($\alpha\beta = 10$ and $\alpha\beta = 01$), which accounts for the four terms. The deviation of the parameters (likelihood) is computed as follows:

$$likelihood = \frac{\sum_{p \in param} |\log_{10}(\frac{current_value_p}{initial_value_p})|}{\#param}$$

with *param* the set of optimized parameters for a given model. After each optimization run, we keep the best fitting solutions found at every generations (*i.e.* one candidate solution per generation). We then map them in the 2D space (*likelihood, fitness*) and draw the

Pareto fronts of models, *i.e* individuals such that no other individual has simultaneously a better fitness and likelihood. Pareto fronts are commonly used in multi-objective optimization [102, 103] for identifying groups of candidate solutions that are optimal with respect to multiple criteria. A model is said to dominate another if none of the potential solution on its Pareto front are dominated by solutions from the other models. Note that a model can dominate over a particular range of parameters and be dominated over another.

All models share the same 39 parameters: initial conditions (4 signal species initial concentrations, 4 template initial concentrations, 3 enzyme concentrations), hybridization/denaturation parameters (hybridization rate, stability of all signal species, stability penalty for inhibitor on their target template compared to their generating template, left and right dangle stability increase per template, stacking denaturation slowdown for every template, invasion rate of an inhibited template by the input and output) and Michaelis-Menten parameters (activity for each enzyme, a total of five K_m because the polymerase has a different K_m for its input-templates-output substrates, and the exonuclease has a different K_m for activators *vs* inhibitors).

The resulting plot of performance (Fig.3.12) quantitatively suggests that the model with saturation and competition (M3) is the most plausible and that its corresponding fits satisfyingly reproduce the positions, shapes and smoothness of the experimental bifurcations (Fig.3.12). This confirms predictions [104, 105, 106] that the sharing of limited resources -such as enzymes- introduces global and non-linear couplings between substrates, which qualitatively impacts the dynamics of biochemical regulation networks, the shape of their bifurcations diagrams and their properties such as robustness and performance. Yet, none of the models captures the thick 00 buffer lining the bistable area (Fig.3.9, b), suggesting that additional mechanisms or stochastic effects get magnified close to bifurcation frontiers.

Enzymes Kms (M2, M3)	Michaelis-Menten parameters (nM)
Km_{pol}	80
$Km_{pol,displ}$	5.5
Km_{nick}	30
$Km_{exo,signal}$	440
$Km_{exo,inhib}$	150
Stacking slowdown (M1, M2, M3)	Modify the stability value of input and output strands when both are present on template, separated by a nick
$\alpha\alpha$	1
$\beta\beta$	1
$\alpha\beta$	1
$\beta\alpha$	1
Enzyme concentration (M1, M2, M3)	Normalized compared to their concentrations in [45]
Polymerase	1
Nickase	1

Table 3.4: Description and default values of optimized parameters. Candidate solutions always contain the parameters indicated for their model, in the order presented in this Table. Except for Michaelis-Menten parameters, those parameters were not measured directly and should be considered best guesses.

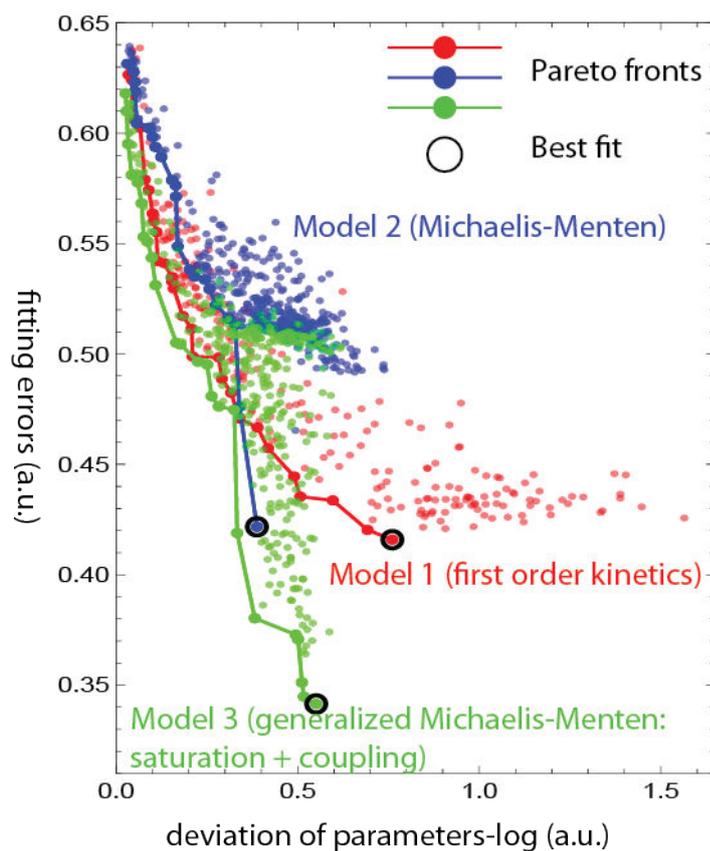


Figure 3.12: Fitting performances of 3 kinetic models. The horizontal axis shows the deviation of the fitted parameters from their initial values. The vertical axis display the distance between the experimental and calculated diagram. Lines represent the pareto fronts for each model, and the best fit is circled in black.

3.3.2 Discussion

A close inspection of the diagrams highlights some unexpected features. First, we expect that all steady states should yield at least one surviving trigger - except in the small 00 area close to the origin. Indeed, assuming continuous dynamic and first-order amplification, we expect the 00 zone (the set of template concentrations that yield a null steady state) to be a rectangle independent of the initial concentrations of triggers. Because we assume that the dynamic is continuous and deterministic, the concentration of a species that is initially strictly positive can never become identically null. In other words, both species cannot lose simultaneously, provided there is enough autocatalytic template.

Let us look at this assertion in more detail. We consider a simplified model that was validated experimentally by Padirac *et al.* [45]. Triggers are amplified according to a first-order Michaelis-Menten amplification whose rate is modulated by the concentrations of autocatalytic template, polymerase and the nickase. The triggers are degraded according to linear kinetics by the exonuclease:

$$\alpha' = \frac{p_\alpha \cdot [\alpha to \alpha] \cdot \alpha}{K_\alpha + \alpha + i_\alpha} - k_\alpha \cdot \alpha$$

$$\beta' = \frac{p_\beta \cdot [\beta to \beta] \cdot \beta}{K_\beta + \beta + i_\beta} - k_\beta \beta$$

The constants p_α and p_β capture the overall replication rates of the triggers by the polymerase and nickase. The variables i_α and i_β are the concentrations of species inhibiting α and β respectively (whose kinetics need not be written explicitly here). The constant k_α and k_β encompass the degradation rates of the exonuclease. For amplification of α to occur ($\alpha' > 0$), it is necessary that the level of autocatalytic template $[\alpha to \alpha]$ exceeds a threshold concentration $\frac{K_\alpha \cdot k_\alpha}{p_\alpha}$ (and similarly for β). This contrasts with second-order autocatalysis, where the concentrations of the initial catalyst and template

must both exceed a critical threshold to ignite replication. In other words, the steady states $\alpha = 0$ or $\beta = 0$ are never stable in first-order amplification once the threshold is crossed because any infinitesimal level of trigger will be amplified. Let us define the rectangle R_{00} in the parameter space (α, β) by

$$R_{00} = \left\{ 0 \leq [\alpha] \leq \frac{K_\alpha \cdot k_\alpha}{p_\alpha}, 0 \leq [\beta] \leq \frac{K_\beta \cdot k_\beta}{p_\beta} \right\}$$

This rectangle is thus necessarily included in the 00 zone. Now we prove the reverse inclusion, namely that the 00 zone is included in the rectangle. Suppose that we start outside R_{00} (at least one template exceeds its threshold level) but that α and β both asymptotically decay to zero. Then the inhibitors also eventually decay to zero. After a (possibly long) relaxation time T_∞ , we have $\beta, i_\beta \ll K_\beta$ and $\alpha, i_\alpha \ll K_\alpha$ and thus

$$\alpha' = \left(\frac{p_\alpha \cdot [\alpha]}{K_\alpha} - k_\alpha \right) \cdot \alpha$$

$$\beta' = \left(\frac{p_\beta \cdot [\beta]}{K_\beta} - k_\beta \right) \cdot \beta$$

Given α and β are both non null (because of their deterministic and continuous evolution), at least one species is amplifying (the species whose template exceeds its threshold concentration), which contradicts the assumptions that the system is converging toward the null state. This simple model predicts that the 00 zone should be a rectangle whose shape is independent of the initial concentrations of triggers. This is clearly not the case in Fig.3.9 where the black areas are highly asymmetrical and extend around the bistable zone, meaning that at least one of our hypothesis must be false.

One argument is that the actual mechanism of replication may be higher order, which would provide the necessary hysteresis to stabilize the steady state $\alpha = 0$ or $\beta = 0$. The second argument is that low volume effect may be at play. The concentrations of α and

β cannot be arbitrarily low without being null ($40 \text{ fM} \times 65 \text{ pL} \approx 1$ molecule). In their race against degradation and cavitating the stochasticity of chemical reactions, the autocatalytic templates may sometime lose and the copy numbers of α and β both reach and remain at 0 (something impossible in a well-posed, deterministic ODE because of the uniqueness of its solution). Since a template cannot replicate a strand that is absent, the 00 zone in our diagram may therefore represent circuits that are “stochastically trapped” in the null state 00. Lastly, when the levels of templates are close (but superior) to their thresholds, the relaxation may drastically slow down and exceed our experimental time scale (10 hours). This phenomenon would be reminiscent of critical slowdown, where some dynamic systems have been shown to take a very long time to recover from quasi-lethal perturbations [107].

proves the failure of our prediction and signals that either convergence to the local steady state(s) is drastically slowed down like in a catastrophic slowdown [107].

Also, some frontiers show a pronounced non-monotonicity. For instance in Fig. 3.9, up-left, it is counter-intuitive that $\beta to \beta$ initially represses α for $\beta to \beta$ between 150 and 200 nM, but eventually permits its survival for $\beta to \beta$ over 200 nM. (Similar non-monotonicity is seen for the 01 frontier in Fig. 3.9). This suggests a hidden mechanism perhaps indirectly cross-activating α and β , whereas the designed circuit topology enforces cross-inhibition between them and DNA sequences were specifically engineered to avoid such undesired interactions.

Note that this non-monotous shape is correctly reproduced only in model 3, suggesting that it is an effect of global coupling through competitive enzyme saturation.

3.4 Conclusion

In conclusion we successfully implemented the bistable circuit designed by Adrien Padirac after modification to fit the device (warm

start to avoid early reaction starting before encapsulation). Then, we generated droplets with different ratios of $\frac{[\beta]_0}{[\alpha]_0}$ for two initial conditions of $\frac{[\beta]}{[\alpha]}$, recorded the fluorescence levels after 11h of evolution, thanks to a home-made data-extraction protocol. The resulting diagrams, once superimposed, display the expected features: an area where β survives, a distinct region where α survives, corresponding to the monostability areas, a region where both triggers survive and the overlap of the two monostable domains defining the bistable area (Fig. 3.9).

More surprisingly, the diagrams display features that were not predicted by the mathematical model. A null area, where none of the triggers survives sits next to the origin of the graph and its shape is different according to the initial conditions, meaning that initial conditions play a role in the putative stochastic trapping of these droplets and provide another level of bistability. at least on the timescale of the experiment, within this system. The shape of the 00 area either signals that either convergence to the local steady state(s) is drastically slowed down (e.g. by catastrophic slowdown [107]) or even arrested (e.g. by small number effects [108]), or that an alternative stable state emerges from an unsuspected mechanism [109]. In any case, these trapped states could provide alternative options for memories on practical time scales. Another feature is the non monotonicity of some frontiers of the diagram, suggesting global coupling in competitive saturation of enzymes. Finally, we upgraded the mathematical model by addition of parameters and dependences between the variables, and evolved the parameters for them to fit with the experimental diagram.

In this process, we established the proof-of-concept for this microfluidic platform for fast and thorough analysis of a network. The next step will be to implement and map another system, with the addition of a new dimension in the parameter space.

Chapter 4

Mapping the predator-prey system

4.1 Introduction

The concept of molecular ecosystem have been widely used to describe the interactions in between proteins, genes, or cells in biology [110, 111, 112, 113]. Imported from the lexical field of ecology, this concept was convenient to describe the density and complexity of interactions happening during cellular process such as mitosis, differentiation, protein expression... But would it be possible to give that expression back to its original field ? In other words, would it be possible to use our knowledge of the gene regulation networks to emulate an ecosystem as it works in ecology? Inside this concept of molecular ecosystem, we isolate the predator-prey systems, referring to two (or more) interacting species for which the increasing growth rate of one species is decreasing the growth rate of the second species. Food webs are the canonical form of predator-prey interactions, and the widely studied Lotka and Volterra equations provide good theoretical descriptions [42, 43]. The equations can be written as:

$$\dot{x} = (b - py)x$$

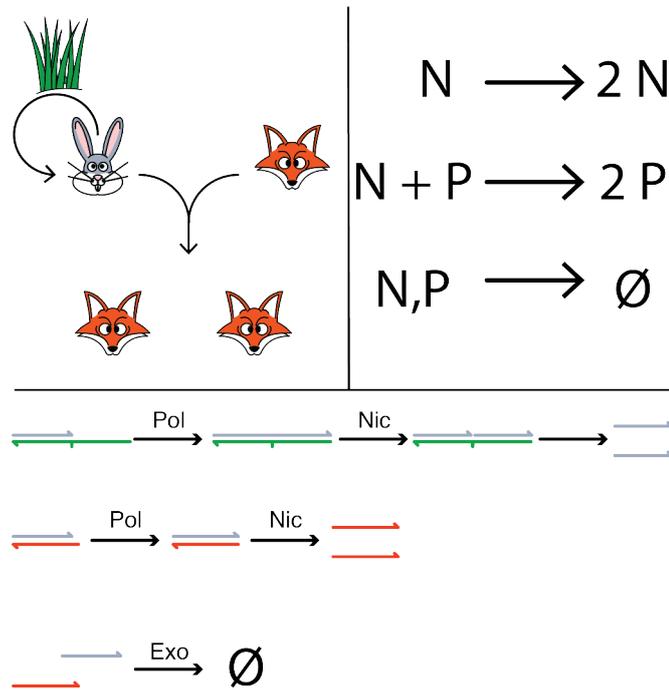


Figure 4.1: Predator-prey system(left) described by Lotka and Volterra equations, and three main rules (right) with P for predator, N for prey (left). PEN-DNA toolbox compatibility and the resulting interacting strands (bottom).

$$\dot{y} = (rx - d)y$$

where b is the growth rate of prey x , p reflects the impact of predation on prey, r is the growth rate of the predator population and d is their death rate.

Although they have been studied and modeled for one century, the fine behaviors of such systems is not yet fully understood, and many mathematical models have been developed to try to deepen our understanding. As an example, Williams and Flaxman studied the effect of the quality of resources on the dynamics and use for that lady beetles as predators, pea aphids as preys and bean plants as a preys' growing resource [114].

Since the predator-prey systems are nothing but species interacting among each others, would it be possible to model such architec-

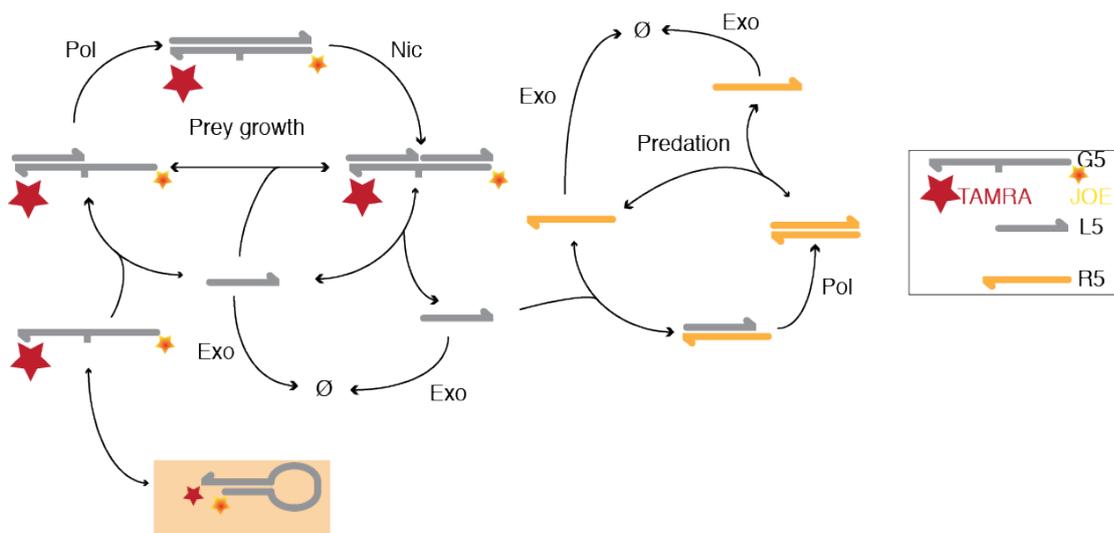


Figure 4.2: Reaction network of the predator-prey system. A prey (L5) replicates by an autocatalytic mechanism identical to the bistable switch involving a grass template (G5). In addition, a predation reaction converts a prey into a predator (R5), catalyzed by a predator and pol: binding of a prey to a predator forms a partial duplex that is elongated by pol to yield two bound predators. Those melt readily at experimental temperature, yielding two free predators.

tures with the PEN-DNA toolbox? Considering the three rules coming from the equations: (1) prey growth is autocatalytic, (2) predation is also autocatalytic and (3) both predator and preys populations are decaying, it is possible to draw a compatible network (Fig.4.1). This network is suited for PEN-DNA toolbox and was implemented since 2012 by Rondelez et al [41], as described in Fig.4.3. The dynamics are intuitively understood when one considers the amount of prey over time: at first, in the absence of predators, preys grow exponentially fast, and so are growing predators because of the number of available preys. At one point, the growth of predator reaches the level of the growth of preys and overcomes it. The number of preys dramatically decreases and the lack of preys prevent the production of predators. Predators finally die, the system is back to its original conditions -low level of both preys and predators- thus ready for another cycle. The inner life cycle of each species is described in Fig.4.2. Experiments confirm this prediction as shown in introduction section.

Among real animal populations, the dynamics of predator-prey systems seem much richer than simple limit cycles oscillations, driven by the spatialization and trophic complexity of the system [115]. The next step was then to spatialize the molecular ecosystem to explore the pattern formation and global dynamics [116].

Our approach will be slightly different: instead of looking for spatialization and out-of-system effects on the dynamics, we will focus on finding-out how the inner parameters of the network can trigger exotic behavior. To do so, we will map the bifurcation diagram of the system in function of 3 parameters.

4.2 Mapping the bifurcation diagram

Using the same approach than for the bistable switch, we aim at mapping the predator-prey system in a three dimensional parameter space. The nature of each parameter is given by the mechanism of this molecular ecosystem. It entails 3 catalytic steps whose rates can be parameterized by the concentration of their exogenous catalysts: 1) the autocatalytic growth of DNA preys, 2) predation: autocatalytic conversion of preys into predators (catalyzed by endogenous predators and pol), 3) degradation of predators and preys (catalyzed by an exonuclease exo).

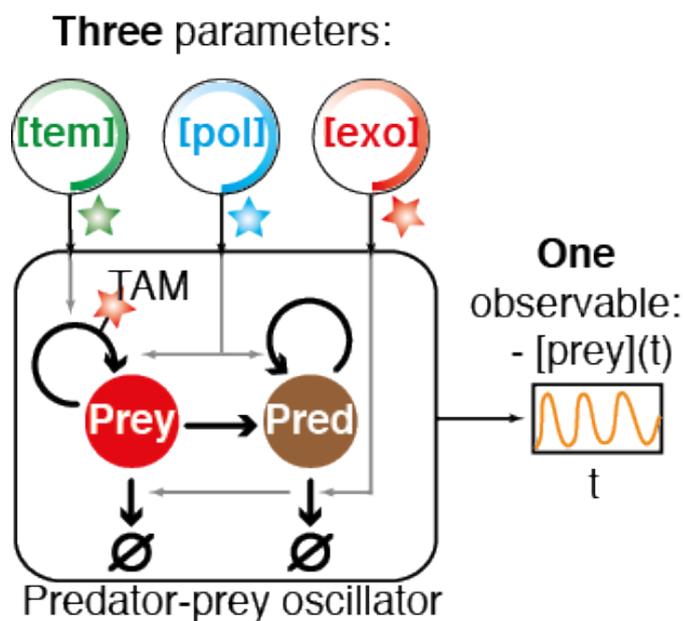


Figure 4.3: Principle of the PP mapping: the DNA-programmed circuit involves two autocatalytic loops where the product of the first one acts as a substrate of the second, as well as an enzymatic sink. The system is naturally parameterized by the concentrations of the polymerase and exonuclease enzymes as well as the DNA replication template. The fluorescence shift of a dye attached to this template exposes the dynamics of preys. Droplet generation is set to sample a cuboid in the parameter space (exo, tem, pol). Approximately $\sim 10^4$ droplets are scanned in 4 fluorescence channels every 6 minutes.

4.2.1 Experimental implementation

We ordered the strands from Biomers (Germany). The predator-prey network is made of three strands representing the actors of the system: a prey strand with, a predator strand and a Grass template, detailed in Table 4.3. The nature of the predation reaction forces the sequence of the predator to be palindromic.

As a consequence, the sequence of the prey will display a pseudo-palindromic form, and this pattern is reproduced on the grass template, leading to a secondary structure, as shown in Fig.4.2. The palindrome should be as short as possible to maintain a high K_d and avoid product inhibition. Surprisingly, dealing with unstable partial

duplexes does not seem to be a problem for the Bst polymerase [41].

length of prey sequence (nb of bases)	sequence (5' → 3')	T _{melt} prey (°C)
9	CATTC ³ CGG	40
10	CATTCGGCCG	43.6
	CATTC ³ CGGG	47.2
11	CATTCAAATTTG	33.1

Table 4.1: Melting temperatures for grass/prey adducts in function of the sequence length. Data calculated with mfold application of DINAMelt server[117, 118], with $[Na^+] = 50$ mM and $[Mg^{2+}] = 8$ mM

Also, sequence and length should be tuned to give a T_m for prey-grass adduct close to the working temperatures [35]. Indeed, the amount of GC pairing in a structure has a strong effect on the T_m of the Prey-Template and Prey-Predator adducts. Several lengths and sequence were tested (example in Table 4.1) in order to set the working temperature between 40 °C and 50 °C. For instance, considering a 9 bases-long prey, sustained oscillations were not obtained below 49 °C (Fig.4.4). In the rest of this chapter, we focused on 10 bases for the prey, 14 bases for the predator and a 20 bases-long grass template.

Finally, preys embed the recognition sequence for the nicking enzyme, but not within a single prey sequence. If it was the case, then the sequence is found also on the predator, thus becoming a target for the nickase. This motivated the replacement of the classical Nt.bstNBI by Nb.BsmI (NEB): the nicking position of this enzyme lays within the recognition site. That way, the prey contains a discontinued recognition site, as shown in Fig.4.5.

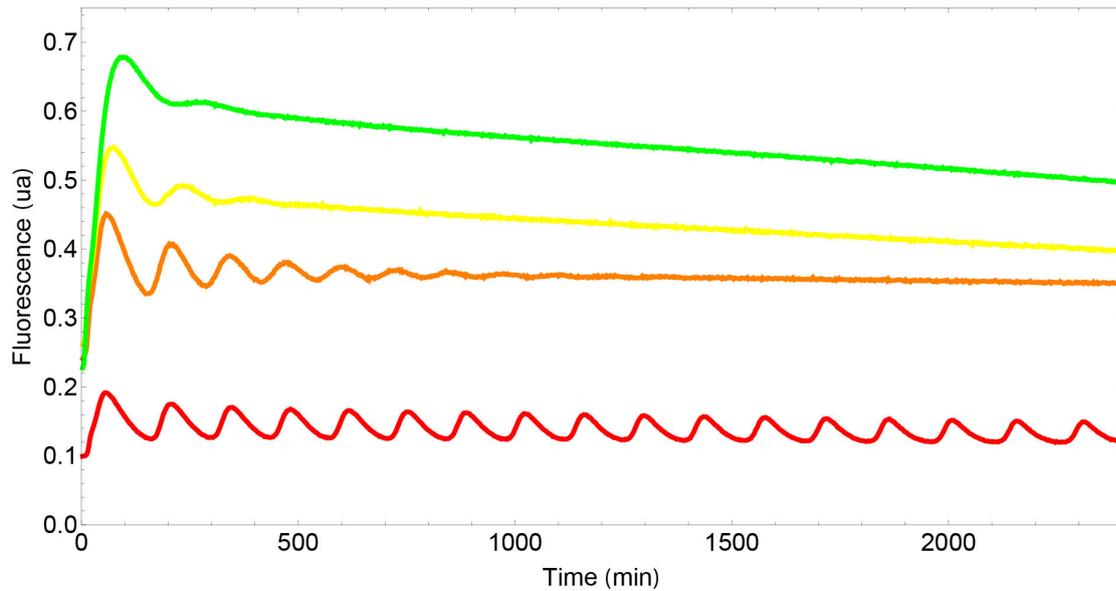


Figure 4.4: Timetraces of a 9bp-long predator-prey system ran at different temperatures. The network quickly reaches a steady-state at 45°C (green), and shows some damped oscillation at 46°C (yellow) and 47.5°C (orange). Sustained oscillations are observed at 48°C (red). All reactions were carried out in a CFX96 Touch thermocycler (BioRAD) with 140nM of template (5'-CGCGGAA TGCGCGGAA TG-3'), 10nM of prey (5'-CATTCCGCG-3'), and 20nM of predators (5'-CATTCCGCGGAATG-3') in a buffer containing 20 mM TrisHCl (pH=8.8), 10mM (NH₄)₂SO₄, 10 mM KCl, 50 mM NaCl, 8 mM MgSO₄, 400 μM dNTPs (NEB), 0,1 % Synperonic F108 (Sigma Aldrich), 2 μM Netropsin (Sigma Aldrich), 4 mM dithiothreitol (Sigma Aldrich), 100 μg/mL BSA (NEB), 1x EvaGreen® (Biotium), 30 pg/μL Extremely Thermostable Single-Strand Binding protein (ETSSB, NEB), 400units/mL Nb.BsmI (NEB), 40units/mL Bst DNA Polymerase large fragment (NEB), and 32.5 nM ttRecJ exonuclease.

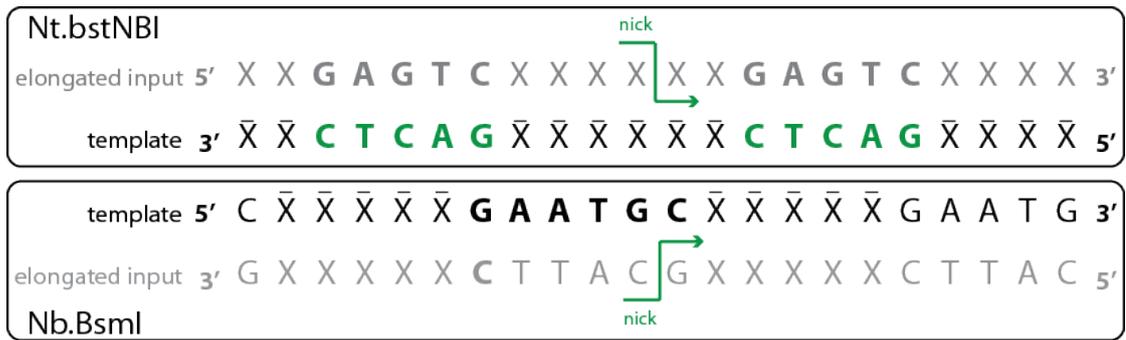


Figure 4.5: Two nicking enzyme restriction site for two strategies

4.2.2 Barcodes and reporters

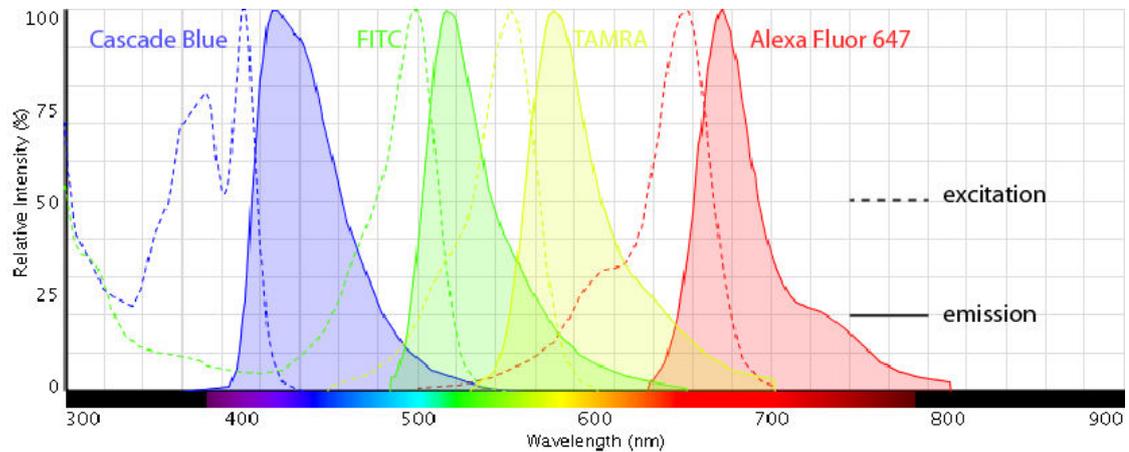


Figure 4.6: Four fluorophores used to map the diagram: Cascade Blue, FITC and Alexa Fluor 647 barcode the parameters, and TAMRA to record the evolution of the system.

We wish to extend the microfluidic mapping approach to higher dimensional spaces and unsteady dynamics. This requires some modification compared to the experiments in the previous chapter.

First of all, we need an additional barcode for the third parameter. Since our limit comes from the specifications of the confocal microscope (i.e. 4 lasers so 4 fluorophores), we aimed for fluorophores whose excitation spectra are compatible with lasers at 473 nm and 559 nm. From the oligothèque we found three candidates: a grass template coupled with JOE in 5' (G5JOE), another with a TAMRA

grass template	fluorophore(s)	λ absorption (nm)	λ emission (nm)	barcode
G5JOE	JOE (3')	520	548	self
G5TAMRA	TAMRA (3')	555	580	self / FITC
G5TAMJOE	TAMRA (3')	520	548	self / FITC
	JOE (5')	555	580	

Table 4.2: Candidates for predator-prey mapping and corresponding barcodes

in 5' (G5TAM) and finally one with TAMRA in 5' and JOE in 3' (G5TAMJOE). Concerning the barcode, one option is to consider self-barcoding for the grass template: an image taken at room temperature before any chemical reaction start can render the gradient of grass template by itself. The problem of this method is that this barcode cannot be measured after the reaction started (i.e. the working temperature is reached) and it could be a problem if droplets are moving during the timelapse experiment. So the strategy of adding a proper barcode is safer.

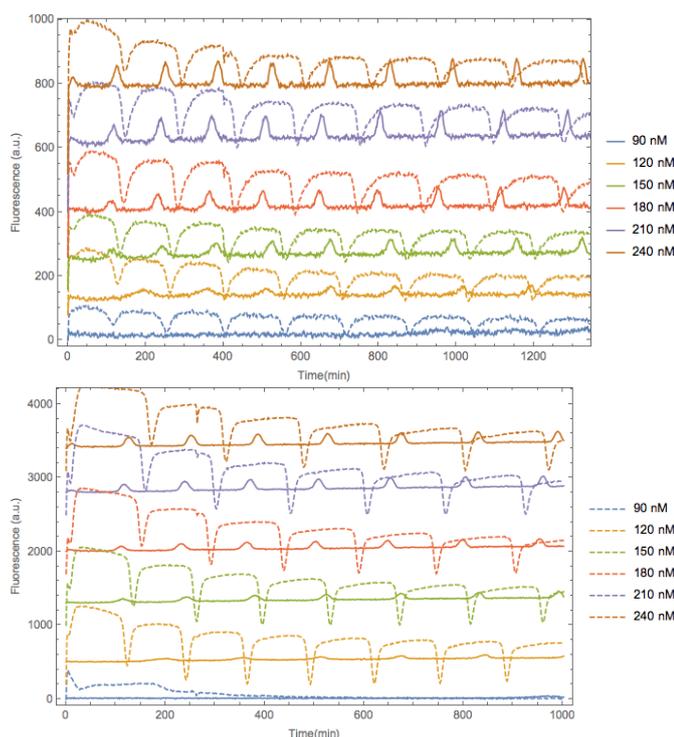


Figure 4.7: Gradient of concentration of grass template in the predator-prey network. Two grass templates are compared (same sequence, different fluorophores). Up: TAMRA (dashed) in G5TAM versus TAMRA in G5TAMJOE (plain). Down: JOE in G5JOE (dashed) versus JOE in G5TAMJOE(plain). Experiments conducted at 45 °C in 20 mM Tris-HCl, 10 mM (NH₄)₂SO₄, 10 mM KCl, 50 mM NaCl, 8 mM MgSO₄, 0.1 % Synperonic F108, dNTP (400 μM each), 2 μM Netropsin, 3 mM DTT, EvaGreen® 1x, 500 μg/ml BSA, 400 Units.mL⁻¹ Nb.bst.BsmI (NEB) and 5 μg.mL⁻¹ of ETSSB (NEB), 1 % (v/v) of Bst Full Length® DNA Polymerase (NEB, stock solution at 800 units.mL⁻¹), 1.3 % (v/v) of ttRecJ exonuclease (stock solution at 3.84 μM) . The prey strand is initially present at 20 nM, the predator strand at 10 nM. Fluorescence baselines have been drifted for clarity.

Considering the absorption spectra listed in Table 4.2, it appears that TAMRA has a closer λ_{max} than JOE to the 559 nm laser. The full system was first tested in a thermocycler, to estimate the boundaries of the future cube in the parameter space, and the results are shown in Fig.4.7.

G5TAMJOE has a shorter period for relatively high concentrations, and even if the amplitude is dramatically decreased, the peaks are well defined and the noise/ratio signal seems good . For this two reasons we decided to select the G5TAMJOE, and barcode it using

FITC-labelled dextran. Another curious feature is that, compared to G5JOE and G5TAM, the signal of G5TAMJOE is reverted, positive shifts are observed for each cycle of prey population.

4.2.3 Timelapse

Unlike the bistable, the predator-prey network has a dynamic behavior, it is thus necessary to record the fluorescence signal overtime. We had to adapt the setup to this new technical challenge, in which the scanning area is a crucial parameter.

4.2.3.1 Scanning time

Scanning time should be shorter than the timescale of the reaction. Indeed, if the scanning area is too large, or the scanning too slow, the first and last images of the first frame will not be chemically equivalent (Fig.4.8), meaning that the reactions started prior to the scanning for the last images of the frame. Even if it is not a big issue in general, it can become critical when one of the parameter is self-barcoded. In this case, we lose the coordinates of all droplets that started earlier. As a rule of thumb, we limited the scanning time around 7min.

The scanning area is defined by two criteria. First, the resolution of the final diagram. Indeed, a good coverage of the parameter space requires the acquisition of fluorescence levels of many droplets (each droplet leads to one point in the diagram). Empirically, we consider that having a precision $\sim 3\%$ for each parameter is good enough. It means that we need $\sim 10^4$ droplets to map in three dimensions. On the other hand, the scanning time has to be in the range of 7 min, keeping in mind that the stage movement (most time consuming) and the optical settings (scanning time per pixel, image resolution, number of image in the frame) cannot be reduced.

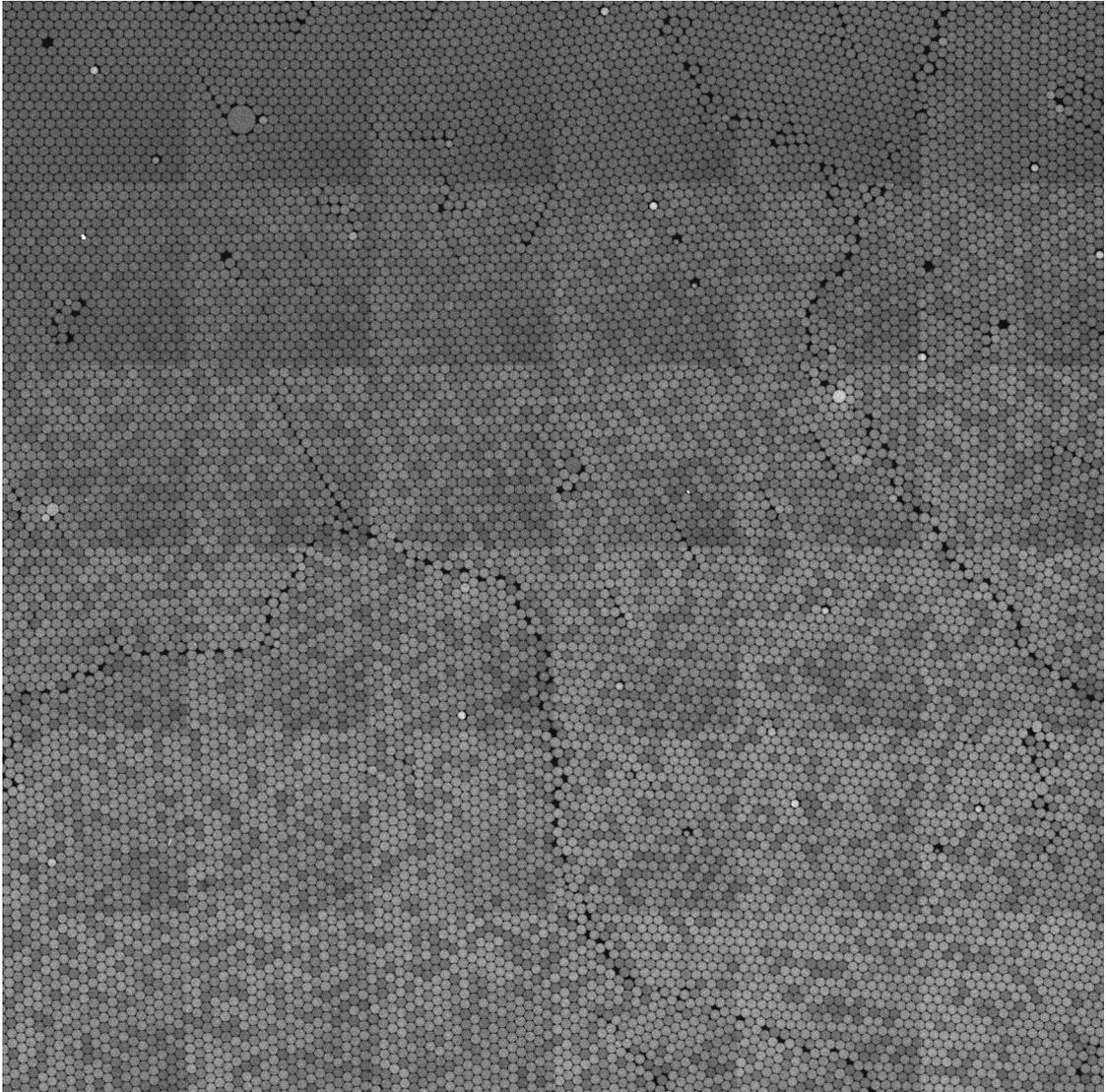


Figure 4.8: Scanning area limitation: first frame of a timelapse. The first images (top left) are homogenous -meaning chemistry did not occurred yet- while the late images show contrast amongst the droplets.

4.2.3.2 Cost and storage

When monitoring a large area for a long time, one has to consider the amount of data produced. Ideally, we want to map the maximum amount of droplets as long as possible. As an example, a 24 hours timelapse of a 2D experiment (3 fluorescence levels to monitor) with reasonable scanning area -around 15 images- generates ~5 Go of movie file. The predator-prey experiment -3 dimensions, 4 fluorescence levels- generates more than 30 Go of data. These heavy microscopy images are first processed by the microscope itself, and it was not rare that the software crashed because of the amount of data. They are then processed for signal extraction, and the heavier they are, the longer is the treatment. This is becoming an important factor in the development of new microscopy techniques [119]. That is why we worked to optimize the settings of the timelapse (resolution, number of image per frame) to reduce the size of the data. Moreover, such long timelapse (up to 5 days) increase the cost of the experiment. We estimated that a 24 hours timelapse costs around 250 euros, with 95% of the budget for the microscope.

4.3 Results and analysis

4.3.1 Reaction assembly

	Sequence 5' → 3'	3' mod.	5' mod.
Grass Template	C*G*G*CCGAATGCGGCCGAATG	JOE	TAMRA
Prey	CATTCGGCCG		
Predator	CATTCGGGCCGAATG		

Table 4.3: Sequences of DNA strands used in the predator-prey network. Stars correspond to a phosphothioate modification (which prevents degradation from the exonuclease).

We prepare a master mix containing 20 mM Tris-HCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 10 mM KCl, 50 mM NaCl, 8 mM MgSO_4 , 0.1 % Synperonic F108, dNTP (400 μM each), 2 μM Netropsin, 3 mM DTT, 500 $\mu\text{g}/\text{ml}$ BSA,

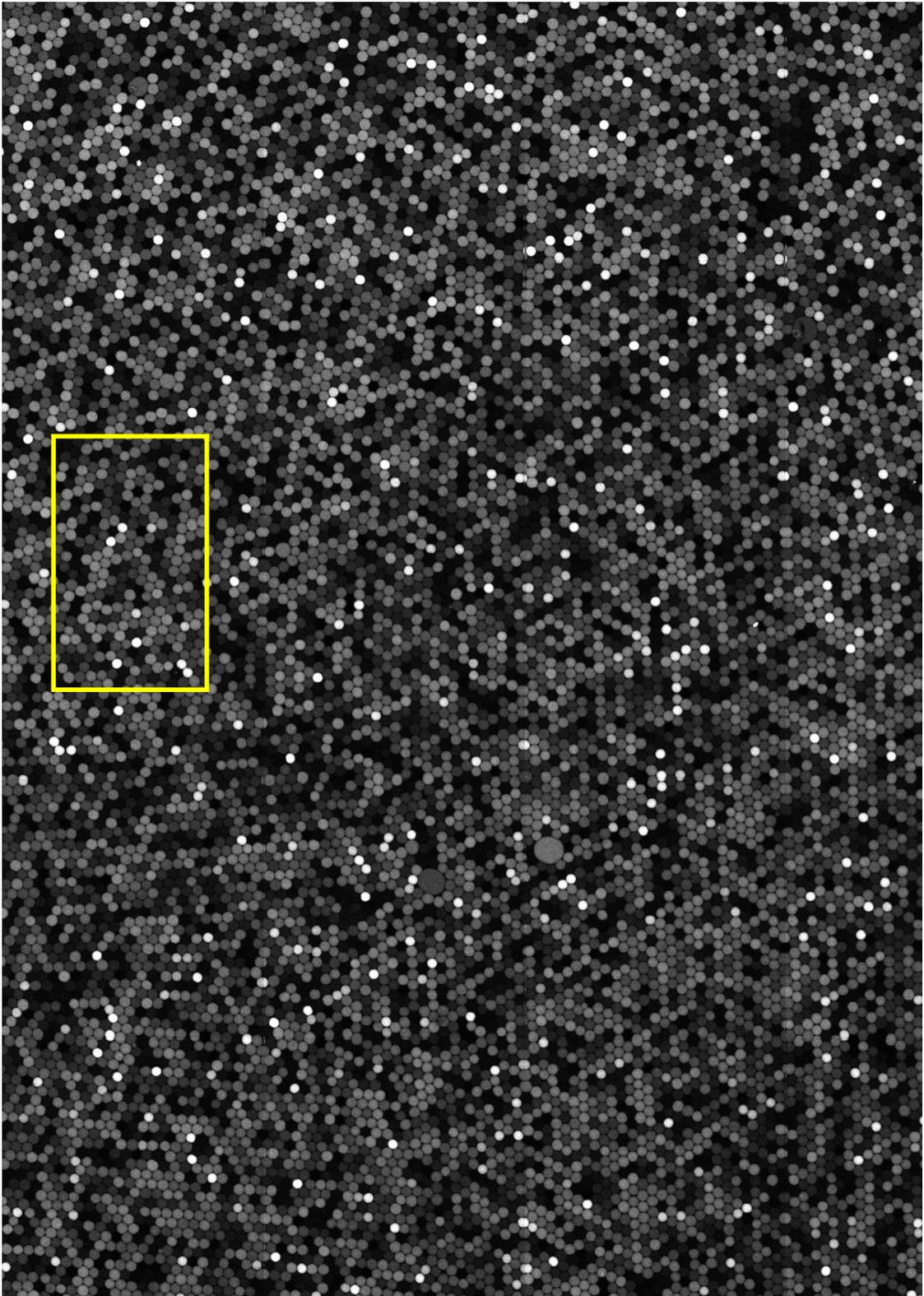


Figure 4.9: Droplet array for the predator-prey system with Bst2.0 WarmStart polymerase. The first (out of 798) image in the observation (TAMRA) channel.

400 Units.mL⁻¹ Nb.bst.BsmI (NEB) and 5 μ g.mL⁻¹ of ETSSB (NEB). The prey strand is initially present at 20 nM, the predator strand at 10 nM. The master mix is split in 4 tubes:

- Template: 1.2 μ M of grass template, and 200 nM Dextran Fluorescein (40,000 MW, Anionic, Lysine Fixable, Life Technologies)
- Polymerase: 7.5 % (v/v) of Bst 2.0 WarmStart® DNA Polymerase (NEB, stock solution at 800 units.mL⁻¹) with 200 nM of Dextran Cascade Blue
- Exonuclease: 18 % (v/v) of ttRecJ exonuclease (stock solution at 3.84 μ M) with 200 nM of Dextran Alexa Fluor 647
- Buffer compensation: with 20 nM of Alexa Fluor 647, 5 nM of Dextran Cascade Blue and 20 nM of Dextran FITC.

We used the Monte-Carlo-based space sampling script to (see) to generate the droplets. Timelapse experiments required no change in the crafting of the observation chamber, the array being stable for more than one week.

4.3.2 Image processing

For the predator-prey system, by contrast with the bistable system, no fluorophore would be present at the same concentration in all droplets: in each four fluorescence channels some droplets had a very small signal. We therefore used a two steps procedure to extract the coordinate of droplet centers:

- First, the localization method described in the previous chapter was applied to all four channel images of a given time point. The corresponding clouds of points were combined, clustered, and for each cluster, only the point associated with the brightest pixel was kept (this implies that for each droplet we used the brightest channel to detect the center). However, some droplets

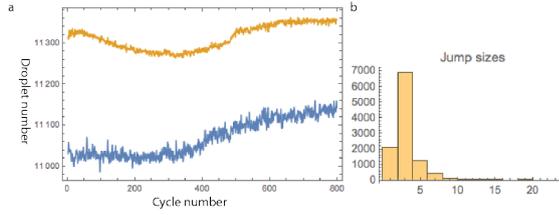


Figure 4.10: Number of droplets detected in each frame and max of jump lengths for each trajectory

lying very close to the origin in barcode space (hence dark in all channels) were missed by this procedure.

- To avoid the possible bias introduced by systematically missing a particular area of parameter space, we used a reconstruction step that leveraged the highly crystalline arrangement of droplets in the image: all detected droplet centers were converted to a disk with a radius equal to 90% of the average droplet diameter (computed using all already discovered droplets). This created a black and white image, which was subsequently blurred using a Gaussian filter. We could then apply the *MaxDetect* function to detect the center of the white areas, which was taken as an approximation of the position of the missing droplets. This graphical procedure was fast enough, and we manually checked that it indeed discovered most “dark” droplets (adding approximately 3% to the total number of detected droplets in a set of images and reducing the discrepancy from image to image, see 4.10). The complete process was then repeated for each 798 four-colors frames of the timelapse movie.

We then tracked the droplets to reconstruct their spatial trajectories during the long incubation on the microscope stage (Fig.4.11). We adapted classical tracking algorithms to cope with the large size of our data set. For each frame n , we detect the droplets centers (x_i, y_i) as indicated above. After that, we supplemented this 2D vector with another 3D vector of barcode intensity, which gives for each frame a set of 5D points $p_i(n) = (x_i, y_i, barcodes_i)$ (the barcodes

are rescaled to be of the same magnitude than the spatial positions). We built the droplet trajectories by recursively linking the $p_i(n)$: the successor of point $p_i(n)$ in frame n is simply defined as the nearest point $p_j(n+1)$ at frame $n+1$. In other words we do not attempt to minimize the total deviation of centers between frames, which allows tracking to be parallelized and significantly sped up. This choice is justified by the fact that droplets move very little, and that the barcode dimensions further decrease tracking errors. Fig.4.10 shows the distribution of the largest step in droplet trajectories, for each trajectory. We see that only a small fraction of the trajectories contain a significant, possibly suspect, jump. We thus separate the trajectories in two groups, depending on whether or not they contain a jump larger than a given threshold (the threshold is set at 10). By looking at the reconstructed trajectories of these two groups in spatial coordinates (Fig.4.11), we see that the trajectories that do include a large jump tend to cluster in a few areas of the image: those on the edge of the image, or those corresponding to merged droplets, which had not been well detected by our segmentation algorithm. We remove those trajectories with large jumps for further analysis, noting that this should not introduce bias in the rest of the analysis. We have also manually verified on randomly selected droplets the correctness of our tracking process.

Once we had a trajectory associated to each droplet, we can give each trace a unique number for further reference. At this stage, we have 9687 a priori valid trajectories.

We now look at the barcode values for the three barcoding fluorescence channels. For each droplet in one frame, we have a 3D barcoding vector. We plot the projection of three dimensional droplet barcode on 2 dimensional space to assess the shape and efficiency of the scanning script (Fig.4.12). Each projection of two different barcodes gives a square-shape with the two-arms calibration. Thanks to that plot, we can link the fluorescence intensities of a barcode to its concentration: the end of an arm corresponds to a droplet full of

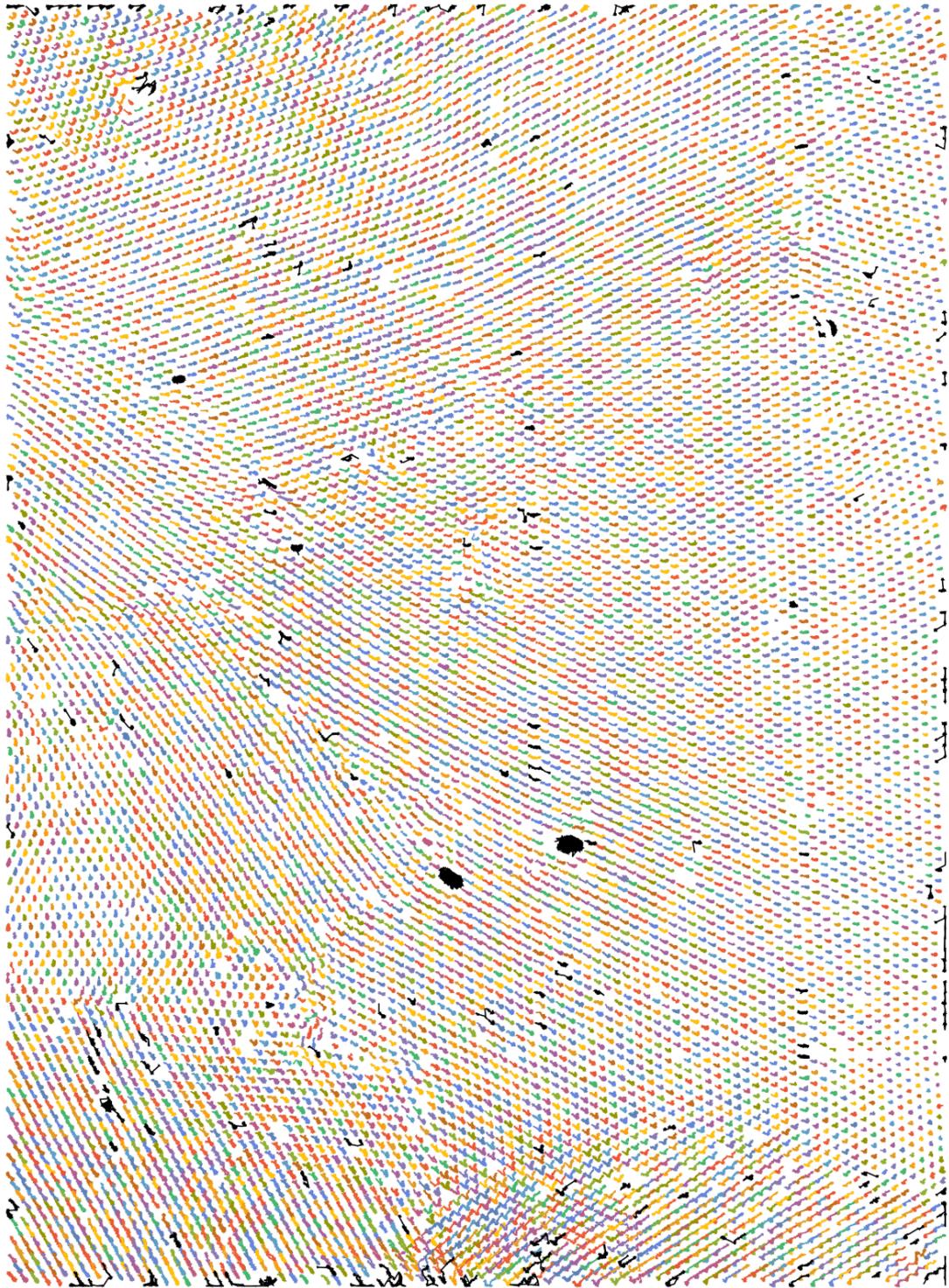


Figure 4.11: Image of all trajectories after tracking and trajectories reconstruction

barcode (whose concentration is known). Knowing the concentration of barcode at each summit of the cube, we can rescale the maximum concentrations reached in the parameter space.

We could extract the fluorescence time-trace by taking droplet-averaged fluorescence values as above. These time traces were smoothed (moving average over 8 frames) and detrended using the built-in Mathematica functions. Peak positions and heights were then detected using the *FindPeak* function and used directly to compute the oscillation score, which is the sum of the absolute height of all detected peaks for one droplet.

4.3.3 About the diagram

During the first day, oscillations in a few percents of the droplets are directly visible. We track the droplets, extract their time traces and score their oscillations (by summing the height of all detected peaks) against their parameters. In the resulting 3D diagram, strong oscillations are confined to an asymmetric, low dimensional manifold (Fig.4.16, a). A scaling argument on the programmed molecular mechanism suggests that the factor $tem \cdot pol/exo$ should shape this oscillatory volume. Indeed, 2D slices in the (pol, exo) plane (constant tem) are bordered by a line $\frac{exo}{pol} \propto tem$ (Fig.4.16, b); production needs to balance degradation (which tends to stabilize fixed points) for oscillations to persist.

The oscillatory region shrinks with increasing tem ; mechanistically tem strengthens the growth of preys compared to their predation, decoupling the decay of preys from the growth of predators, which in turns weakens oscillations. The same scaling is observed in the (tem, pol) slices (constant exo) where the inner boundary approximate a hyperbolic shape $tem \cdot pol \propto exo$. Increasing exo drags this shape to higher $tem \cdot pol$ levels, which by scaling shortens the period of oscillations and improves their score.

Additionally, both sharp and smooth bifurcations are visible: in

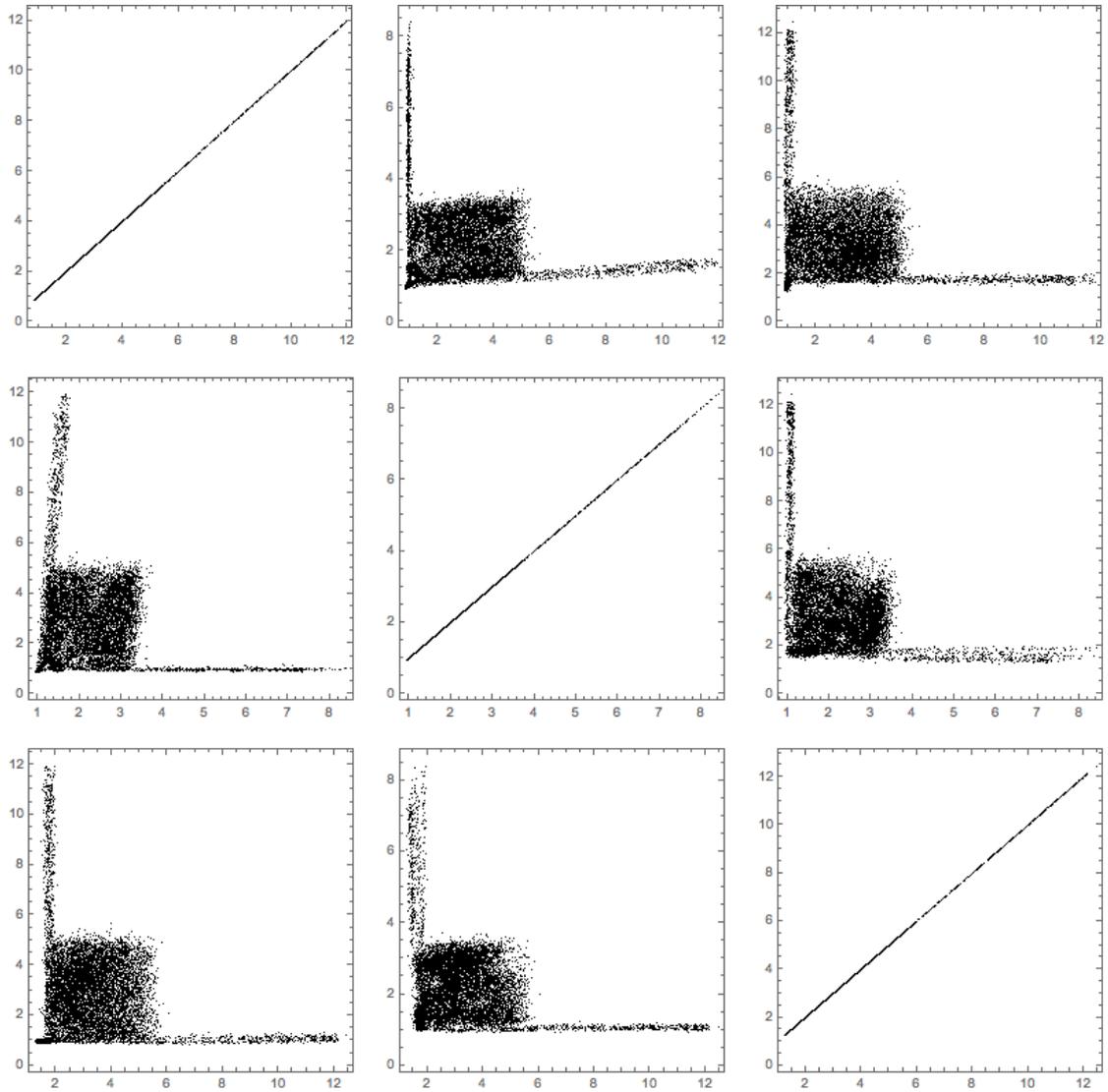


Figure 4.12: Scanning result of the 3D experiment. Each point is a the projection of the three dimensional droplet barcode on 2-dimensionnal space. Each panel x-axis from left to right: blue (pol), Green (template), red (exo) and y-axis top to down blue (pol), Green (template), red (exo). Raw arbitrary fluorescence units are shown, with no corrections.

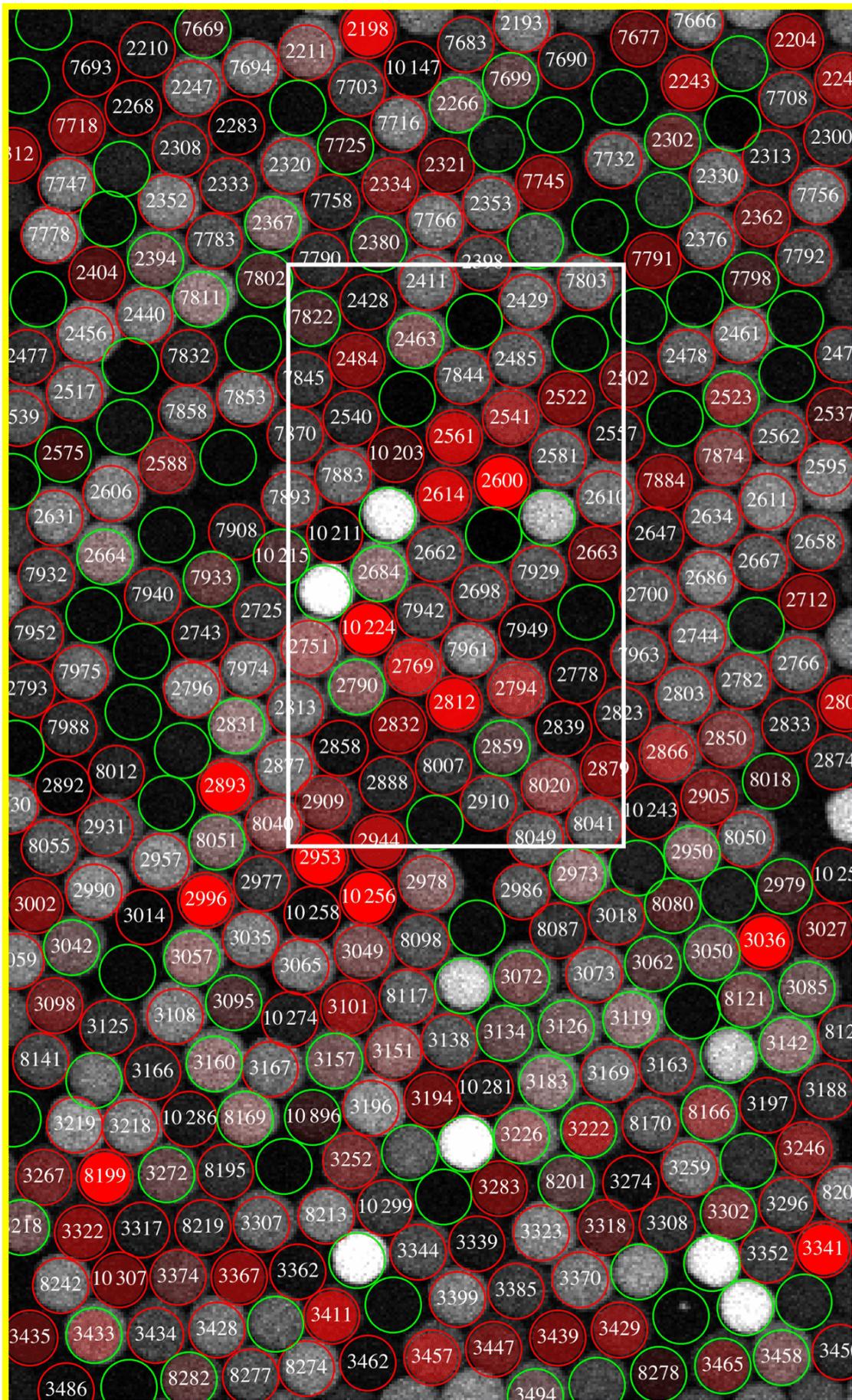


Figure 4.13: Droplet array for the predator-prey system with Bst2.0 WarmStart polymerase. Zoom on the yellow inset in previous image. 368 droplets were detected in this image. 65 of them (circled in green) belong to the self-calibration axes, i.e. they have a ~ 0 value in one of the three fluorescent barcodes. These are used only during the conversion of barcodes intensities into

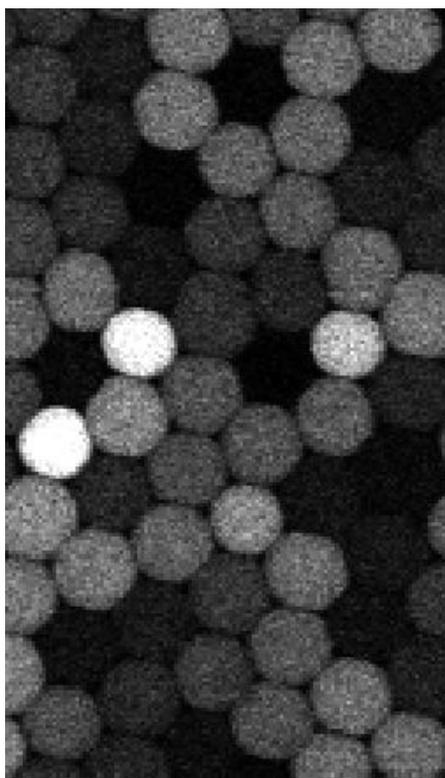


Figure 4.14: Droplet array for the predator-prey system with Bst2.0 WarmStart polymerase. Zoom on the white inset in previous image.

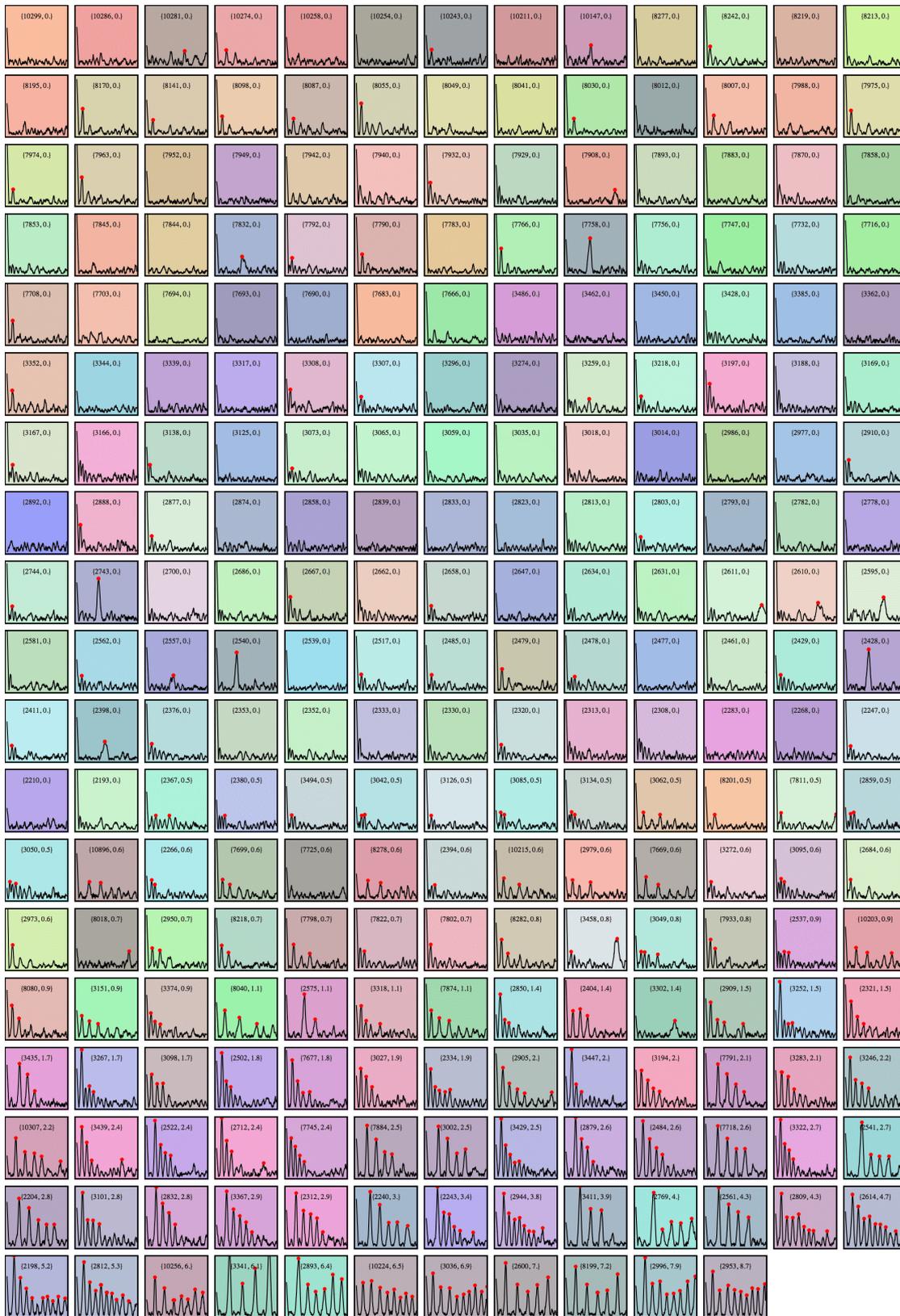


Figure 4.15: 245 time traces extracted from the area shown in Fig.4.14. The plots show the temporal evolution of the fluorescence of droplets (spatially integrated over the whole droplet and after detrending) over 25 hours. Time traces that showed parasitic side reactions ($N=58$), as well as droplets not belonging to the cube ($N=65$ droplets marked green in Fig.4.13), were filtered out. The x-axis indicates time (from 0 to 25 hours) and the y-axis the fluorescence level (arbitrary but identical unit for all traces). The inset shows the droplet identifier, together with the oscillation score. The background color is the RGB combination of the normalized barcodes (i.e. red content = [polymerase]/260; green content = [template]/350; blue content = [exonuclease]/115)

Fig.4.16, b, starting from low pol, oscillations start sharply but gradually weaken into damped oscillations (dashed line and inset). This suggests that the Hopf bifurcation leads to trajectories spiraling down to a stable coexistence focus on one side, but to an extinction state with no oscillatory potential on the other side. These observations agree with a simple analytical model of the predator-prey system (see Introduction).

After the first day, the emergence of opportunistic DNA replicators overcomes the predator-prey system in most droplets [120]. We inspect the dynamics of the remaining droplets looking for traces with low autocorrelation and observed surprising stochastic behaviors (Fig.4.16, c). While most of those surviving droplets either dampen (bottom right in Fig.4.16, c), oscillate sustainably (diagonal in Fig.4.16, c and Fig.4.16, d) or fail to oscillate at all (dark points in bottom left in Fig.4.16, c) - in line with qualitative predictions for deterministic dynamical system - a small fraction of droplets exhibit “bursting” oscillations that spontaneously switch ON or OFF (Fig.4.16, e). Those droplets originated from a small, well-defined region of the parameter space lining the low pol, low temp face of the oscillatory region (Fig.4.16, a). Remarkably, stability analysis of the kinetic model does predict the existence of a small manifold of bimodal steady/oscillatory (i.e. hard-excitable) systems, located precisely at the inner edge of the oscillatory region, and where small amplitude noise could induce a similar bursting behavior. In this case, stochasticity originating from the small size of the compartments would in fact be instrumental to uncover (and possibly exploit) alternative attractors of the molecular system [121].

4.4 Discussion

The mathematical derivation and analysis of a two-variable model for the predator-prey system is presented in the introduction and is only sketched here for reference.

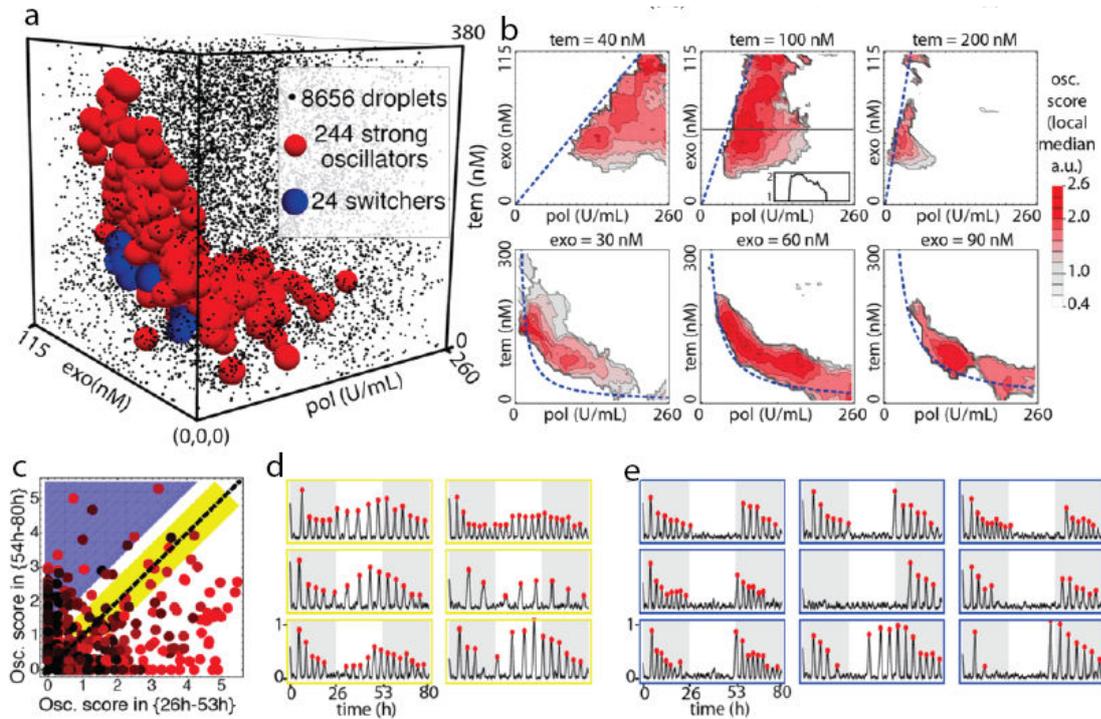


Figure 4.16: Timetraces and bifurcation diagram of the predator-prey oscillator. (a) Classification of the droplets according to their oscillatory score (sum of heights of peaks) over 25 hours in the parameter space. (b) Slices of the diagram along the template (prey growth; top) or exonuclease (decay; bottom) directions. The dashed lines are projections of a unique algebraic manifold $exo = c \cdot tem \cdot pol$ (c, d, e). Long-term behavior reveals stochastic switchers. (c) Comparison of oscillatory scores over day 1, 2 and 3 for the parasite-free droplets. The color encodes the score in the first 25 hours period, while the position encodes the score for the second and third day. The yellow strip corresponds to traces with a strong correlation between day 2 and 3 (see d). The blue strip corresponds to traces with weak oscillations during day 2 but strong oscillations during day 3 (see e). (d) Example of un-interrupted oscillations from the yellow strip of c. (e) The traces from the blue strip in e appear to switch spontaneously between oscillatory and non-oscillatory regime.

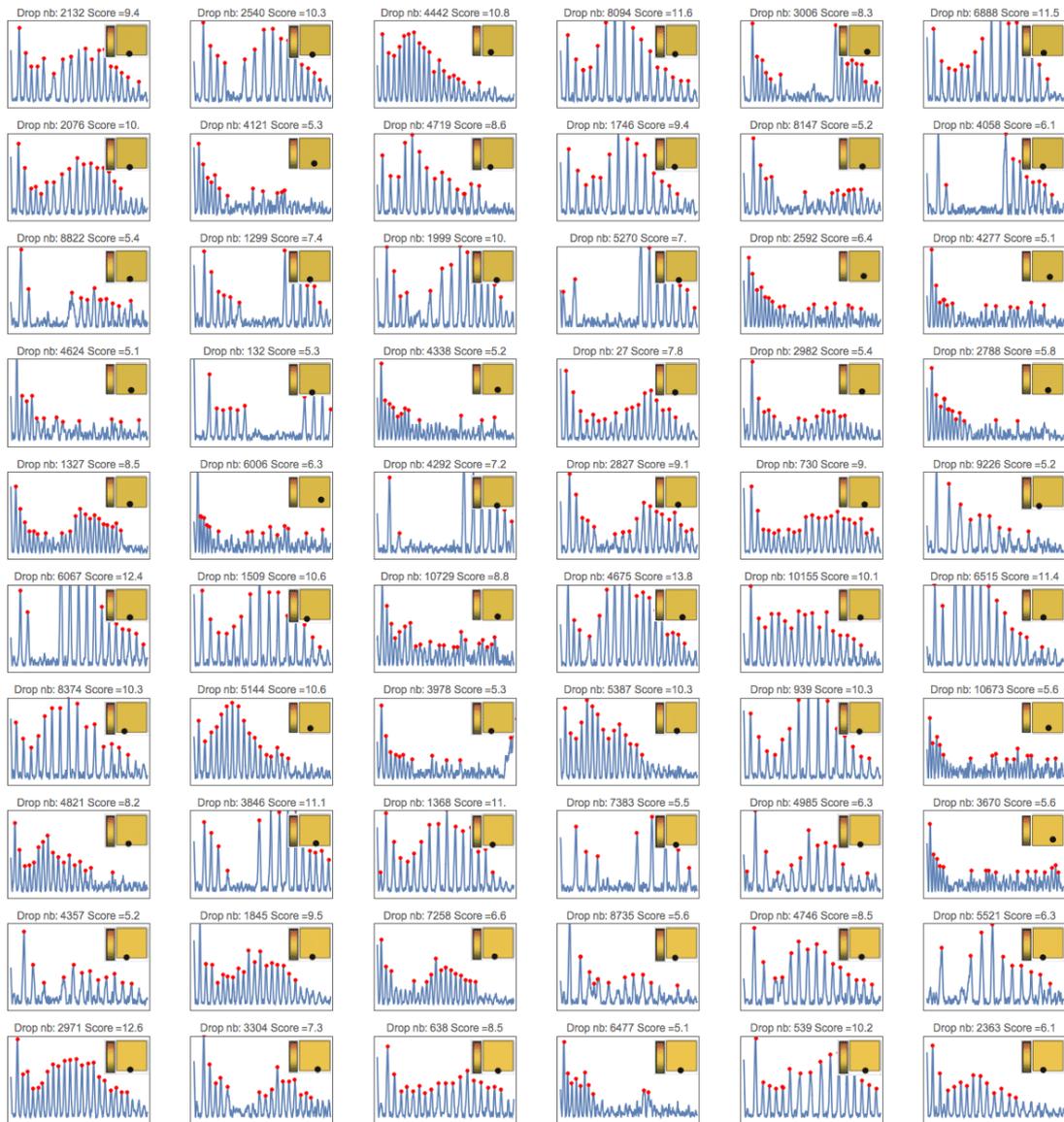


Figure 4.17: Best oscillators: This image and the following show all traces that are detected with an oscillation score above 5. Red dots show the peaks detected by the automatic algorithm. The inset shows the position of the droplet in parameter space: the background color codes for the template concentration following the color scale shown on the left, while the x and y position indicate exonuclease and polymerase concentrations, respectively. The plots have been ordered by increasing template concentration and the droplet number and score are indicated above. The x-axis is time, running from 0 to xx hours and the y-axis is arbitrary fluorescent units for the detrended timetraces.

Noting the three parameters, i.e. the three non-dimensionalized concentrations of the three enzymes pol , tem and exo , we have obtained obtain the following set of differential equations for the evolution of the prey (n) and the predator (p):

$$\frac{dn}{d\tau} = \frac{pol \cdot tem \cdot n}{1 + \beta \cdot pol \cdot tem \cdot n} - pol \cdot p \cdot n - \lambda \cdot exo \frac{n}{1 + p} \quad (4.1)$$

$$\frac{dp}{d\tau} = pol \cdot p \cdot n - exo \frac{p}{1 + p} \quad (4.2)$$

Then we have performed a partial linearization in the case of a mildly saturating system (i.e. when $\beta \cdot pol \cdot tem \cdot n$ is always small compared to 1):

$$\dot{n} = pol \cdot tem \cdot n(1 - \beta \cdot pol \cdot tem \cdot n) - pol \cdot p \cdot n - \lambda \cdot \delta \frac{n}{1 + p} \quad (4.3)$$

This allows us to look analytically at the existence and stability of the steady states.

4.4.1 Bifurcation analysis

By setting \dot{n} and \dot{p} to 0, one finds four equilibrium points:

- (0, 0) i.e extinction of both species,
- $(\frac{tem \cdot pol - exo \cdot \lambda}{\beta \cdot pol^2 \cdot tem^2}, 0)$ i.e. extinction of the predator and stable population of prey, valid for $\lambda \cdot exo < tem \cdot pol$ (i.e. first order production rate of preys overcomes their decay)
- $(\frac{1+tem+\sqrt{\Delta}/pol}{2(pol \cdot tem^2 \cdot \beta + \lambda)}, -1 + tem - \frac{\sqrt{\Delta}}{2pol})$, corresponding to coexistence of both species, valid for $tem > 1 + \frac{\sqrt{\Delta}}{2pol}$ and $\Delta > 0$.
- $(\frac{1+tem-\sqrt{\Delta}/pol}{2(pol \cdot tem^2 \cdot \beta + \lambda)}, -1 + tem + \frac{\sqrt{\Delta}}{2pol})$, also corresponding to coexistence of both species.

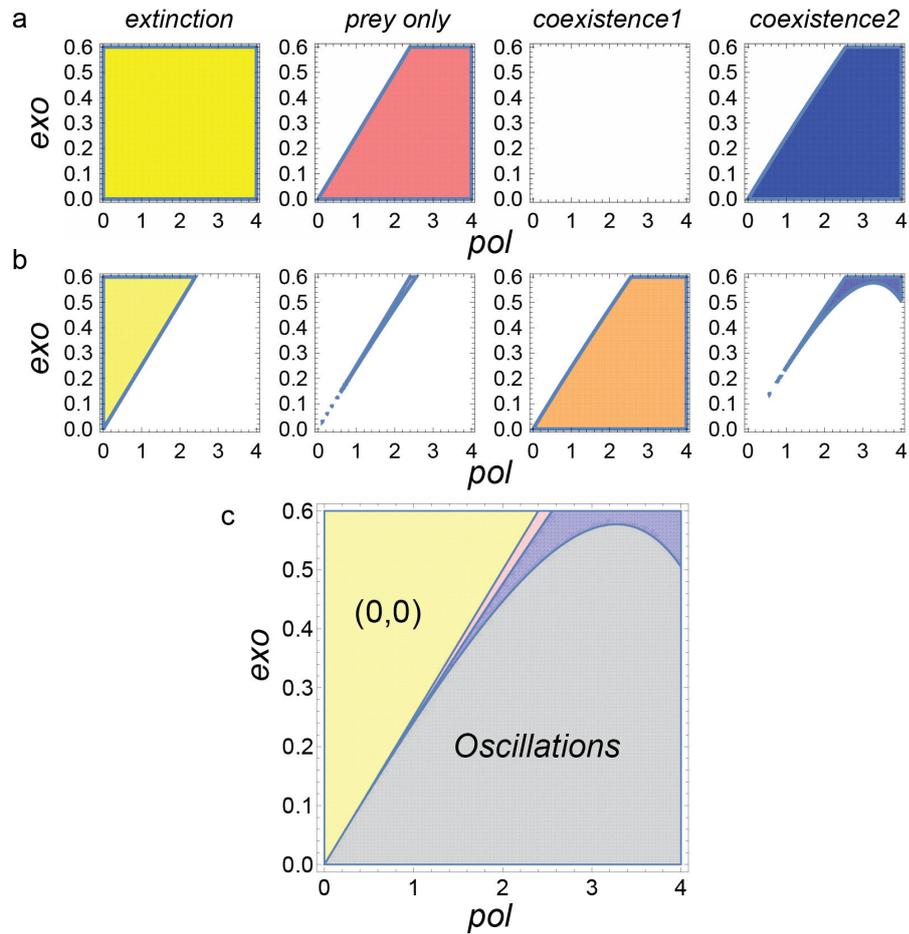


Figure 4.18: Bifurcation analysis of the model defined by equations 4.2 and 4.3. β and λ are set to 0.1 and 4, respectively (see reference [?]). tem is set to 1. **a**, Regions of existence of *positive* solutions. **b**, Stability of solutions. **c**, Bifurcation diagram showing the Hopf bifurcation around the oscillatory region (in gray). Yellow correspond to the extinction region (no prey, no predators), pink to the prey-only area, and blue to the stable coexistence area.

with

$$\Delta = \sqrt{(pol - pol \cdot tem)^2 - 4pol(-pol \cdot tem + exo \cdot pol \cdot tem^2\beta + exo \cdot \lambda)}$$

By looking at the eigenvalues of the community matrix linearized around these points (see [?]) we find that:

- the first point is stable for $(\lambda \cdot exo) > (tem \cdot pol)$.
- The second point is stable for $(\lambda \cdot exo) < (tem \cdot pol)$ and $(\lambda \cdot exo) > [tem \cdot pol(exo \cdot tem \cdot pol - 1)]$.
- The third is never stable over its domain of existence. on trajectories
- The fourth has an unstable region (corresponding to the oscillatory area) and a stable region (corresponding to damping oscillations). These areas are defined by complicated analytical expressions.

We can then use the analytical forms to compute the theoretical bifurcations diagram. The results are schematized in Fig.4.18. Two dimensional slices in $\{pol, exo\}$ and in $\{pol, tem\}$ are shown in Fig.4.19.

The slices obtained in Fig.4.19 reproduce the features of the experimental diagram: the oscillating area is observed below a $exo/pol = cste$ line in the tem slices and beyond a $pol \cdot tem = constant$ hyperbola in the exo slices. The oscillating region decreases in size and migrates toward the axis with increasing tem or increasing exo . For higher pol values, the limit cycle gives way to a fixed stable point, and thus a damping behavior is expected for most initial conditions. For a given set of exo and tem , Fig.4.20 shows the behavior when one increases pol , starting from the extinction area: oscillations appear suddenly, become faster and then evolve to damped oscillations leading to a stable asymptotic state. The “oscillation score” profile can be computed and shows the asymmetrical shape also observed experimentally.

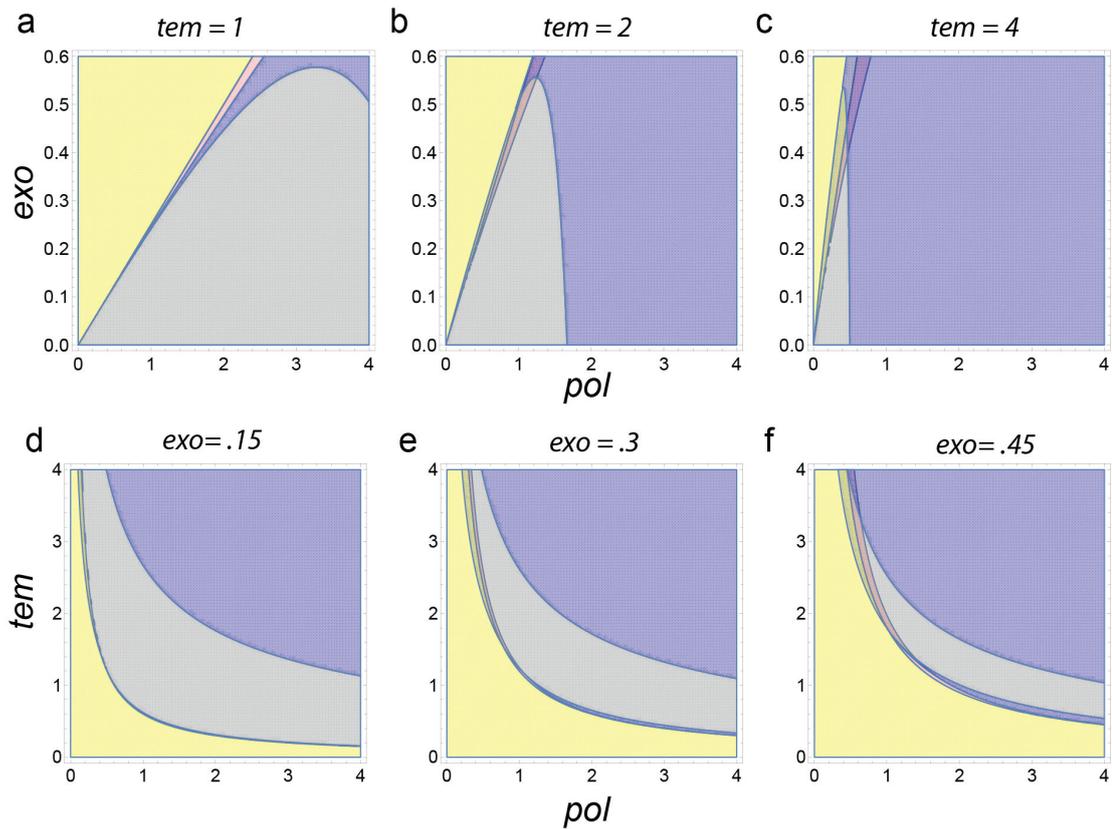


Figure 4.19: 3-dimensional bifurcation diagrams of the 2 variables model along pol , tem , and exo . **a**, **b**, **c**, equi- tem slices of the diagram with non-dimensional tem respectively equals to 1, 2, and 4. **d**, **e**, **f**, equi- exo slices of the bifurcation diagram with non-dimensional exo respectively equals to 0.15, 0.3, and 0.45. The region of stable oscillations is in gray. Yellow corresponds to the extinction region (no prey, no predator), pink to the prey-only area, and blue to the stable coexistence area.

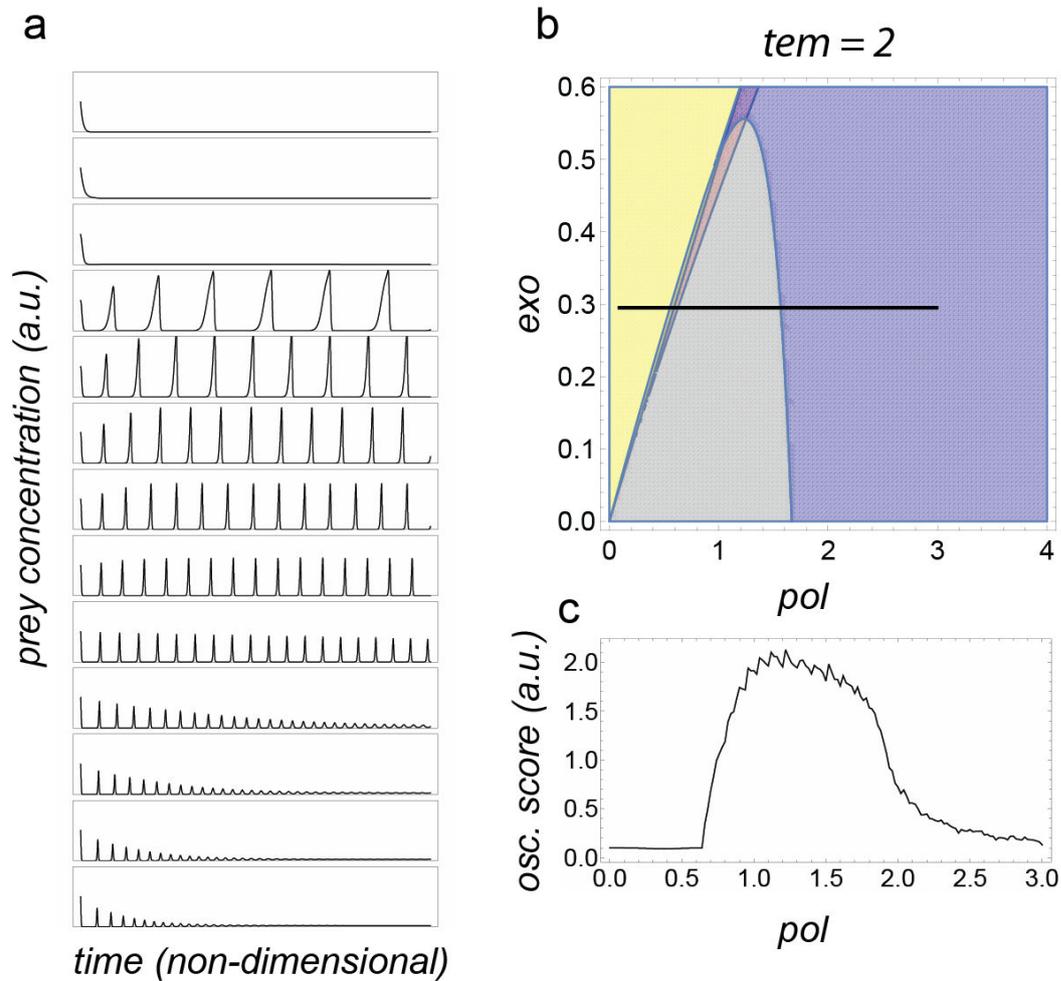


Figure 4.20: **a**, Numerical timetraces for the two-variable model solved with the parameter set $\{tem = 2, exo = 0.3\}$ and pol ranging from .2 (top plot) to 3 (bottom plot) by steps of 0.2 units (corresponding to the black line in the bifurcation diagram shown in **b**). **c**, profile of computed oscillations score for the same segment in parameter space.

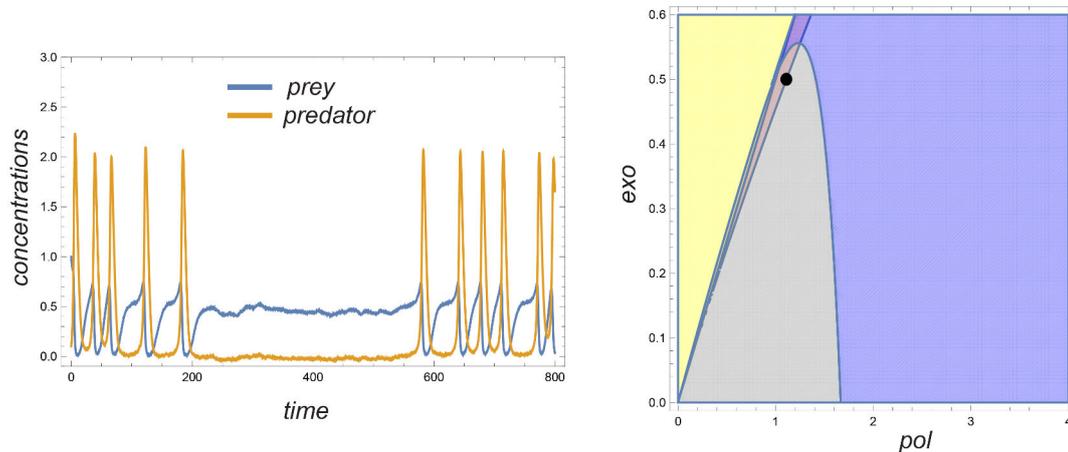


Figure 4.21: Stochastic bursting in the 2 variables model for oscillators close to the extinction area. Left: the ODE system with a noise component (Wiener process of amplitude 0.004, i.e. less than 1% of the prey concentration in the steady state) is solved for the parameter set $\{tem = 2, pol = 1.1, exo = .5\}$ represented by the black point in the bifurcation diagram shown on the right.

Another observation is that, when tem , pol and exo are not too large, the Hopf bifurcation delineating the region of stable limit cycles is almost merged to the boundary of the $(0,0)$ area. As shown above, the equation defining this particular surface is $\lambda \cdot exo = tem \cdot pol$. We thus recover the scaling discussed theoretically above, and obvious from the experimental results. This also suggests that the sharp boundary observed in the experimental diagram on the low (pol, exo, tem) side is indeed a Hopf bifurcation.

Finally, we see that in some cases, the region of stable oscillations can coexist with another stable point, either the extinction state (yellow) or the prey-only (pink) state. This would define a so-called “hard-excitation” system, having two asymptotic attractors, one being a fixed point and the other a limit cycle. Given some noise, such a system would be able to switch spontaneously between cycling and flat behavior. This would happen however on a very limited parameter range, located precisely at the interface between the limit cycle area and the extinction zone (*i.e.* close to the Hopf bifurcation). Indeed, if we numerically solve the two-variable ODE system with a

noise component (Langevin dynamic) within this multimodal region we observe a bursting behavior similar to the experimental results (Fig.4.21).

Globally, linear stability analysis of a very simple two-variable model is in good agreement with the shape of the experimental diagram. It confirms that we do experimentally observe the Hopf bifurcation on the low concentration side, while its exact position on the other side is masked by the existence of damped trajectories. The effect of molecular noise in these small droplets appears to be amplified by the non linear dynamics, especially when the system is close to multiple bifurcations. The spectacular bursting of some oscillating droplets is a good example of this combination.

Chapter 5

Conclusion

Molecular programming is a new, still emergent field which takes inspiration from information processing as it is observed in living organisms. The goal is to create information-managing systems at the molecular level. In the process, we hope to better understand the inner machinery of a cell, but also to exploit the properties of such unconventional computing schemes to bypass issues in other, more classical approaches (e.g. solid-state electronic devices). Maybe the molecular programming approach will one day allow us to design or engineer new and useful molecular architectures displaying unique behaviors.

In this context we developed in the laboratory the PEN-DNA toolbox, which is composed of three cascable reactions inspired by gene regulation networks found in cells, or even in viruses (the lambda phage and its spectacular lytic “switch” was one of the prominent case-study that allowed researchers to better understand the dynamics of regulation of genetic expression and its dramatic effects on phenotypes).

Using the PEN-DNA toolbox, activation, inhibition and degradation are generic reactions that can be connected in arbitrary topologies. These test-tube networks reproduce biological-relevant behaviors such as bistability or oscillations. In this work, we tried to go one step further in the global understanding of how such networks behave

by extracting -experimentally- their bifurcation diagrams. For that, we first developed a microfluidic tool for the experimental mapping of concentration parameter spaces. Our device generates millions of water-in-oil droplets with different composition thanks to a pressure controller for which we developed scripts adapted to the 2D and 3D mapping. In the current implementation the mesh (and therefore the resolution of our maps) is fixed by the iteration number and the number of droplet that we can ultimately image. We have also worked a lot on the observation chamber to make it leak-proof, movement-free, and compatible with long-term timelapse (more than 3 days).

The first system that we tested was the bistable switch from Padirac et al. [45]. We successfully mapped the bifurcation diagram of this network as a function of two parameters depicting the autocatalysis strength of each side. The results showed monostability and bistability areas, concurring with previous theoretical studies, but also revealed some unexpected features. For example, we observed the surprising shape of the 00 domain, which extends asymmetrically on both sides of the bistable area. Therefore, in some area of its parameter space, the system remains in a “forbidden” state (according to simple linear stability analysis (LSA) arguments) even after 11h of evolution. This singularity may come from the network itself, showing the existence of a third stable point because of unaccounted nonlinearities in the interactions between the species. Alternatively, it could be the consequence of small number effects in these tiny compartments. Further investigations should unveil the reasons underlying such particularity. For example, droplet scanning experiments focusing on this particular area and performed for a longer time (up to several days) will help to decide whether there is real stochastic trapping or if the effect is due to a strong slowdown of the system.

We then decided to add a third dimension to the map and to study a predator-prey system [41]. We focused on the effects of growth and decay rate parameters by varying the concentration of

polymerase (a catalytic resource for the growth of predators and preys), the concentration of exonuclease (a catalytic resource for the degradation of both species) and the concentration of grass template (involved in prey growth only). After extraction of the timetraces, it appeared that the strong oscillating area is circumscribed in an asymmetric, low dimensional manifold. We hypothesized a scaling mechanism to shape the oscillatory volume according to the equation: $exo = c \cdot tem \cdot pol$. The diagram showed both sharp and smooth bifurcations, suggesting that the orbits in shape portraits lead to stable coexistence or extinction, while LSA predicted the existence of a narrow, hard-excitations domain. We observed a small fraction of droplets exhibiting “bursting” oscillations that spontaneously switched ON or OFF, all coming from the small, well-defined region of parameter space lining the low *pol*, low *temp* face of the oscillatory region. Such switches may originate from the molecular confinement in low-concentration droplets. This should be confirmed in the future by changing the size of the compartment for a set of parametric conditions lying within this small area.

One may question the technical limits of such an approach. How far can we push this technology? How many parameters can we scan? Which resolution can we achieve in order to reveal rare and exotic behaviors? In the future, more more accurate pressure control scripts using space-filling curves will improve the scanning time. This will allow us to decrease the mesh size and makes the system more user-friendly. The actual generation of droplets -although serial- is not an immediate bottleneck. We routinely generate ~ 1 million of droplets in ~ 45 min, and faster rates of 1 MHz have been achieved in the literature [122] (although without combinatorial preparation). Programming the pressure profiles to mix more than 4 inlets should not be beyond the reach of a skilled operator. Incubation of droplets is not a bottleneck since reactions run in parallel. Fluorescence scanning of the droplets proves to be the most time-consuming operation. We chose confocal laser microscopy because of its high signal-to-noise

ratio. But this scanning method is serial and thus slow. Scanning $\sim 10\,000$ droplets requires about ~ 5 min. Faster rates could be achieved by decreasing the scanning resolution and the time spent on each pixel, pushing up the laser power if needed to compensate for the lower number of photons collected per droplet. Faster methods of scanning could also be considered. Droplet flow cytometry (a sequential method) scans ~ 2000 droplets/s ($\sim 600\,000$ droplets/ 5 min) [72]. Among parallel acquisition methods, a sCMOS camera coupled with epifluorescence microscopy is an attractive option. A 5 megapixel sCMOS camera can nominally scan $\sim 50\,000$ droplets simultaneously at a resolution of 100 pixels/droplet. Assuming an exposure time on the order of 10 seconds per frame (including stage translation, filter switching and acquisition), a sCMOS setup will acquire the fluorescence of 1.5 millions droplets per 5 minutes. Note that parallel scanning methods require more careful calibration than sequential ones to account for the spatial dependence of the illumination and sensors. Fluorescence overlap will also have to be dealt with, but it is realistic to encode the concentration of 6 variables simultaneously [123].

More generally, our microfluidic platform for bifurcation diagram mapping is a great tool in understanding the inner machinery, as much as unveiling the possibilities of any network compatible with the PEN-DNA toolbox. Because of their experimental, assumption-free origin, these maps are more accurate than any model. At the moment, they cannot be used to predict untested conditions, but we expect that they will lead to a next generation of more accurate mathematical models -boosted with feedbacks from the experimental bifurcation diagrams- which in turn will allow more realistic computational predictions.

Finally, our platform will find applications in other molecular contexts where interactions between nodes of a chemical reaction network gives rise to nontrivial dynamics. For example it could be possible to apply our approach to interrogate cellular regulatory net-

works and progress in the understanding of system biology. Ultimately these visual diagrams will offer mechanistic insights that will sharpen our understanding of design principles in molecular systems in a very general way.

Bibliography

- [1] Ralf Dahm. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research., 2008.
- [2] J D Watson and F H C Crick. A structure for deoxyribose nucleic acid. 1953., 2003.
- [3] S N Cohen, A C Chang, H W Boyer, and R B Helling. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 70(11):3240–3244, 1973.
- [4] F Sanger, S Nicklen, and A R Coulson. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12):5463–5467, 1977.
- [5] R K Saiki, D H Gelfand, S Stoffel, S J Scharf, R Higuchi, G T Horn, K B Mullis, and H A Erlich. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839):487–491, 1988.
- [6] A Fire, S Q Xu, M K Montgomery, S A Kostas, and S E Driver. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 1998.
- [7] M Jinek, K Chylinski, I Fonfara, M Hauer, and J A Doudna. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012.
- [8] Leslie E Orgel. Prebiotic chemistry and the origin of the RNA world. *Critical reviews in biochemistry and molecular biology*, 39(2):99–123, 2004.
- [9] John SantaLucia and Donald Hicks. The thermodynamics of DNA structural motifs. *Annual review of biophysics and biomolecular structure*, 33:415–440, 2004.

- [10] B Yurke, A J Turberfield, A P Mills, F C Simmel, and J L Neumann. A DNA-fuelled molecular machine made of DNA. *Nature*, 406(6796):605–608, 2000.
- [11] David Yu Zhang and Erik Winfree. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *Journal of the American Chemical Society*, 131(47):17303–17314, 2009.
- [12] David Yu Zhang and Georg Seelig. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature chemistry*, 3(2):103–113, 2011.
- [13] F JACOB and J MONOD. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3:318–356, June 1961.
- [14] J Chen and N C Seeman. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature*, 1991.
- [15] E Winfree, F Liu, L A Wenzler, and N C Seeman. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394(6693):539–544, 1998.
- [16] Paul W K Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.
- [17] Dongran Han, Suchetan Pal, Jeanette Nangreave, Zhengtao Deng, Yan Liu, and Hao Yan. DNA origami with complex curvatures in three-dimensional space. *Science*, 332(6027):342–346, 2011.
- [18] Dongran Han, Suchetan Pal, Yan Liu, and Hao Yan. Folding and cutting DNA into reconfigurable topological nanostructures. *Nature nanotechnology*, 5(10):712–717, 2010.
- [19] Kyle Lund, Anthony J Manzo, Nadine Dabby, Nicole Michelotti, Alexander Johnson-Buck, Jeanette Nangreave, Steven Taylor, Renjun Pei, Milan N Stojanovic, Nils G Walter, Erik Winfree, and Hao Yan. Molecular robots guided by prescriptive landscapes. *Nature*, 465(7295):206–210, 2010.
- [20] Peng Yin, Harry M T Choi, Colby R Calvert, and Niles A Pierce. Programming biomolecular self-assembly pathways. *Nature*, 451(7176):318–322, 2008.
- [21] Yonggang Ke, Luvena L Ong, William M Shih, and Peng Yin. Three-dimensional structures self-assembled from DNA bricks. *Science*, 338(6111):1177–1183, 2012.

- [22] L M Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021–1024, 1994.
- [23] John J Tyson, Katherine C Chen, and Béla Novák. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2):221–231, 2003.
- [24] Denise M Wolf and Adam P Arkin. Motifs, modules and games in bacteria. *Current Opinion in Microbiology*, 6(2):125–134, 2003.
- [25] Kevin Montagne, Raphael Plasson, Yasuyuki Sakai, Teruo Fujii, and Yannick Rondelez. Programming an in vitro DNA oscillator using a molecular networking strategy. *Molecular Systems Biology*, 7:1–7, 2011.
- [26] Jongmin Kim, Kristin S White, and Erik Winfree. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Molecular Systems Biology*, 2, 2006.
- [27] Shotaro Ayukawa, Masahiro Takinoue, and Daisuke Kiga. RTRACS: a modularized RNA-dependent RNA transcription system with high programmability. *Accounts of Chemical Research*, 44(12):1369–1379, 2011.
- [28] Jongmin Kim and Erik Winfree. Synthetic in vitro transcriptional oscillators. *Molecular Systems Biology*, 7:1–15, 2011.
- [29] Masahiro Takinoue, Daisuke Kiga, Koh-ichiroh Shohda, and Akira Suyama. Experiments and simulation models of a basic computation element of an autonomous molecular computing system. *Physical Review E*, 78(4):041921, 2008.
- [30] Y Sakai, Y Mawatari, K Yamasaki, and K Shohda. Construction of AND Gate for RTRACS with the Capacity of Extension to NAND Gate. *DNA Computing and . . .*, 2009.
- [31] A Kan, Y Sakai, K Shohda, and A Suyama. A DNA based molecular logic gate capable of a variety of logical operations. *Natural Computing*, 2014.
- [32] M Takinoue, D Kiga, K Shohda, and A Suyama. RNA Oscillator: limit cycle oscillations based on artificial biomolecular reactions. *New Generation Computing*, 2009.
- [33] Joseph N Zadeh, Conrad D Steenberg, Justin S Bois, Brian R Wolfe, Marshall B Pierce, Asif R Khan, Robert M Dirks, and Niles A Pierce. NUPACK: analysis and design of nucleic acid systems. *Journal of computational chemistry*, 32(1):170–173, 2011.

- [34] Jeffrey Van Ness, Lori K Van Ness, and David J Galas. Isothermal reactions for the amplification of oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 100(8):4504–4509, 2003.
- [35] Alexandre Baccouche, Kevin Montagne, Adrien Padirac, Teruo Fujii, and Yannick Rondelez. Dynamic DNA-toolbox reaction circuits: a walkthrough. *Methods*, 67(2):234–249, 2014.
- [36] Luca Giorgetti, Trevor Siggers, Guido Tianana, Greta Caprara, Samuele Notarbartolo, Teresa Corona, Manolis Pasparakis, Paolo Milani, Martha L Bulyk, and Gioacchino Natoli. Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Molecular cell*, 37(3):418–428, 2010.
- [37] R Sarpeshkar. Analog versus digital: extrapolating from electronics to neurobiology. *Neural computation*, 10(7):1601–1638, 1998.
- [38] R Sarpeshkar. Analog synthetic biology. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 372(2012):20130110, 2014.
- [39] Lulu Qian and Erik Winfree. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332(6034):1196–1201, 2011.
- [40] Lulu Qian and Erik Winfree. A simple DNA gate motif for synthesizing large-scale circuits. *Journal of The Royal Society Interface*, 8(62):1281–1297, 2011.
- [41] Teruo Fujii and Yannick Rondelez. Predator–Prey Molecular Ecosystems. *ACS Nano*, 7(1):27–34, 2013.
- [42] A J Lotka. Undamped oscillations derived from the law of mass action. *Journal of the American Chemical Society*, 1920.
- [43] V Volterra. Variations and fluctuations of the number of individuals in animal species living together. *J Cons Int Explor Mer*, 1928.
- [44] Xi Chen and Andrew D Ellington. Shaping up nucleic acid computation. *Current Opinion in Biotechnology*, 21(4):392–400, 2010.
- [45] Adrien Padirac, Teruo Fujii, and Yannick Rondelez. Bottom-up construction of in vitro switchable memories. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47):E3212–E3220, 2012.

- [46] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004.
- [47] N Rosenfeld, M B Elowitz, and U Alon. Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology*, 2002.
- [48] Klaus Jähnisch, Volker Hessel, Holger Löwe, and Manfred Baerns. Chemistry in microstructured reactors. *Angewandte Chemie (International ed. in English)*, 43(4):406–446, 2004.
- [49] J Wagner and J M Köhler. Continuous synthesis of gold nanoparticles in a microreactor. *Nano letters*, 5(4):685–691, 2005.
- [50] Jianing Yang, Yingjie Liu, Cory B Rauch, Rauch L Stevens, Randall H Liu, Robin Lenigk, and Piotr Grodzinski. High sensitivity PCR assay in plastic micro reactors. *Lab on a chip*, 2(4):179–187, 2002.
- [51] Robert C Wootton, Robin Fortt, and Andrew J de Mello. On-chip generation and reaction of unstable intermediates-monolithic nanoreactors for diazonium chemistry: azo dyes. *Lab on a chip*, 2(1):5–7, 2002.
- [52] V Hessel and H Löwe. Mikroverfahrenstechnik: Komponenten–Anlagenkonzeption–Anwenderakzeptanz–Teil 1. *Chemie Ingenieur Technik*, 2002.
- [53] Bo Zheng, L Spencer Roach, and Rustem F Ismagilov. Screening of protein crystallization conditions on a microfluidic chip using nanoliter-size droplets. *Journal of the American Chemical Society*, 125(37):11170–11171, 2003.
- [54] L Zhang, J M Koo, L Jiang, and M Asheghi. Measurements and modeling of two-phase flow in microchannels with nearly constant heat flux boundary conditions. . . . *Systems*, 2002.
- [55] Michihiko Nakano, Jun Komatsu, Shun-ichi Matsuura, Kazunori Takashima, Shinji Katsura, and Akira Mizuno. Single-molecule PCR using water-in-oil emulsion. *Journal of biotechnology*, 102(2):117–124, 2003.
- [56] Deniz Pekin, Youssr Skhiri, Jean-Christophe Baret, Delphine Le Corre, Linas Mazutis, Chaouki Ben Salem, Florian Millot, Abdeslam El Harrak, J Brian Hutchison, Jonathan W Larson, Darren R Link, Pierre Laurent-Puig, Andrew D Griffiths, and Valérie Taly. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. *Lab on a chip*, 11(13):2156–2166, 2011.

- [57] Irving R Epstein, Vladimir K Vanag, Anna C Balazs, Olga Kuksenok, Pratyush Dayal, and Amitabh Bhattacharya. Chemical Oscillators in Structured Media. *Accounts of Chemical Research*, 45(12):2160–2168, 2012.
- [58] G F Christopher and S L Anna. Microfluidic methods for generating continuous droplet streams. *Journal of Physics D: Applied Physics*, 2007.
- [59] Carsten Cramer, Peter Fischer, and Erich J Windhab. Drop formation in a co-flowing ambient fluid. *Chemical Engineering Science*, 59(15):3045–3058, 2004.
- [60] T Thorsen, R W Roberts, F H Arnold, and S R Quake. Dynamic pattern formation in a vesicle-generating microfluidic device. *Physical Review Letters*, 2001.
- [61] S L Anna, N Bontoux, and H A Stone. Formation of dispersions using “flow focusing” in microchannels. *Applied physics letters*, 2003.
- [62] Helen Song, Delai L Chen, and Rustem F Ismagilov. Reactions in droplets in microfluidic channels. *Angewandte Chemie (International ed. in English)*, 45(44):7336–7356, 2006.
- [63] Jean-Christophe Baret. Surfactants in droplet-based microfluidics. *Lab on a chip*, 12(3):422–433, 2012.
- [64] Thomas Ward, Magalie Faivre, Manouk Abkarian, and Howard A Stone. Microfluidic flow focusing: drop size and scaling in pressure versus flow-rate-driven pumping. *Electrophoresis*, 26(19):3716–3724, 2005.
- [65] Elodie Sollier, Coleman Murray, Pietro Maoddi, and Dino Di Carlo. Rapid prototyping polymers for microfluidic devices and high pressure injections. *Lab on a chip*, 11(22):3752–3765, 2011.
- [66] J C McDonald, D C Duffy, J R Anderson, D T Chiu, H Wu, O J Schueller, and G M Whitesides. Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis*, 21(1):27–40, 2000.
- [67] Y Xia and G M Whitesides. Soft lithography. *Annual review of materials science*, 1998.
- [68] Shohei Kaneda, Koichi Ono, Tatsuhiro Fukuba, Takahiko Nojima, Takatoki Yamamoto, and Teruo Fujii. Modification of the glass surface property in PDMS-glass hybrid microfluidic devices. *Analytical sciences : the international journal of the Japan Society for Analytical Chemistry*, 28(1):39–44, 2012.

- [69] I M Lifshitz and V V Slyozov. The kinetics of precipitation from supersaturated solid solutions. *Journal of Physics and Chemistry of Solids*, 1961.
- [70] M J Rosen and J T Kunjappu. Surfactants and interfacial phenomena, 2012.
- [71] Qun Zhong, Smiti Bhattacharya, Steven Kotsopoulos, Jeff Olson, Valérie Taly, Andrew D Griffiths, Darren R Link, and Jonathan W Larson. Multiplex digital PCR: breaking the one target per color barrier of quantitative PCR. *Lab on a chip*, 11(13):2167, 2011.
- [72] Jenifer Clausell-Tormos, Diana Lieber, Jean-Christophe Baret, Abdeslam El Harrak, Oliver J Miller, Lucas Frenz, Joshua Blouwolff, Katherine J Humphry, Sarah Köster, Honey Duan, Christian Holtze, David A Weitz, Andrew D Griffiths, and Christoph A Merten. Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms. *Chemistry & Biology*, 15(5):427–437, 2008.
- [73] F J Ghadessy, J L Ong, and P Holliger. Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8):4552–4557, 2001.
- [74] Yunpeng Bai, Ximin He, Dingsheng Liu, Santoshkumar N Patil, Dan Bratton, Ansgar Huebner, Florian Hollfelder, Chris Abell, and Wilhelm T S Huck. A double droplet trap system for studying mass transport across a droplet-droplet interface. *Lab on a chip*, 10(10):1281–1285, 2010.
- [75] C Holtze, A C Rowat, J J Agresti, J B Hutchison, F E Angilè, C H J Schmitz, S Köster, H Duan, K J Humphry, R A Scanga, J S Johnson, D Pisignano, and D A Weitz. Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab on a chip*, 8(10):1632–1639, 2008.
- [76] Maximilian Weitz, Jongmin Kim, Korbinian Kapsner, Erik Winfree, Elisa Franco, and Friedrich C Simmel. Diversity in the dynamical behaviour of a compartmentalized programmable biochemical oscillator. *Nature chemistry*, 6(4):295–302, 2014.
- [77] Etienne Fradet, Craig McDougall, Paul Abbyad, Rémi Dangla, David McGloin, and Charles N Baroud. Combining rails and anchors with laser forcing for selective manipulation within 2D droplet arrays. *Lab on a chip*, 11(24):4228–4234, 2011.
- [78] Helen Song and Rustem F Ismagilov. Millisecond kinetics on a microfluidic chip using nanoliters of reagents. *Journal of the American Chemical Society*, 125(47):14613–14619, 2003.

- [79] Xing-Hu Ji, Wei Cheng, Feng Guo, Wei Liu, Shi-Shang Guo, Zhi-Ke He, and Xing-Zhong Zhao. On-demand preparation of quantum dot-encoded microparticles using a droplet microfluidic system. *Lab on a chip*, 11(15):2561–2568, 2011.
- [80] Koshi Hasatani, Mathieu Leocmach, Anthony J Genot, André Estévez-Torres, Teruo Fujii, and Yannick Rondelez. High-throughput and long-term observation of compartmentalized biochemical oscillators. *Chemical communications (Cambridge, England)*, 49(73):8090–8092, 2013.
- [81] Soo Hyeon Kim, Shino Iwai, Suguru Araki, Shouichi Sakakihara, Ryota Iino, and Hiroyuki Noji. Large-scale femtoliter droplet array for digital counting of single biomolecules. *Lab on a chip*, 12(23):4986–4991, 2012.
- [82] Mathieu Morel, Denis Bartolo, Jean-Christophe Galas, Maxime Dahan, and Vincent Studer. Microfluidic stickers for cell- and tissue-based assays in microchannels. *Lab on a chip*, 9(7):1011–1013, 2009.
- [83] M J Booth and T Wilson. Strategies for the compensation of specimen-induced spherical aberration in confocal microscopy of skin. *Journal of microscopy*, 200(Pt 1):68–74, 2000.
- [84] J M Murray. Methods for Imaging Thick Specimens: Confocal Microscopy, Deconvolution, and Structured Illumination. *Cold Spring Harbor Protocols*, 2011(12):pdb.top066936–pdb.top066936, 2011.
- [85] Upinder S Bhalla, Prahlad T Ram, and Ravi Iyengar. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, 297(5583):1018–1023, 2002.
- [86] M Laurent and N Kellershohn. Multistability: a major means of differentiation and evolution in biological systems. *Trends in biochemical sciences*, 24(11):418–422, 1999.
- [87] James E Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Current Opinion in Cell Biology*, 14(2):140–148, April 2002.
- [88] Lydia Robert, Gregory Paul, Yong Chen, François Taddei, Damien Baigl, and Ariel B Lindner. Pre-dispositions and epigenetic inheritance in the *Escherichia coli* lactose operon bistable switch. *Molecular Systems Biology*, 6:357, 2010.

- [89] N Vijesh, S K Chakrabarti, and J Sreekumar. Modeling of gene regulatory networks: A review. *Journal of Biomedical . . .*, 2013.
- [90] Y A Kuznetsov. *Elements of Applied Bifurcation Theory*. Springer Science & Business Media, June 2004.
- [91] N Ramakrishnan and U S Bhalla. Memory switches in chemical reaction space. 2008.
- [92] Paul François and Vincent Hakim. Design of genetic networks with specified functions by evolution in silico. *Proceedings of the National Academy of Sciences of the United States of America*, 101(2):580–585, 2004.
- [93] S Widder, J Macía, and R Solé. Monomeric bistability and the role of autoloops in gene regulation. *PLoS ONE*, pages 1–7, 2009.
- [94] A Yamagata, R Masui, Y Kakuta, S Kuramitsu, and K Fukuyama. Over-expression, purification and characterization of RecJ protein from *Thermus thermophilus* HB8 and its core domain. *Nucleic Acids Research*, 29(22):4617–4624, 2001.
- [95] Pei-Chung Hsieh and Romualdas Vaisvila. Protein engineering: single or multiple site-directed mutagenesis. *Methods in molecular biology (Clifton, N.J.)*, 978:173–186, 2013.
- [96] R J Russell and G L Taylor. Engineering thermostability: lessons from thermophilic proteins. *Current Opinion in Biotechnology*, 6(4):370–374, August 1995.
- [97] Adrien Padirac, Teruo Fujii, and Yannick Rondelez. Quencher-free multiplexed monitoring of DNA reaction circuits. *Nucleic Acids Research*, 40(15):e118, 2012.
- [98] J D Crawford. Introduction to bifurcation theory. *Reviews of modern physics*, 1991.
- [99] Nathanaël Aubert, Clément Mosca, Teruo Fujii, Masami Hagiya, and Yannick Rondelez. Computer-assisted design for scaling up systems based on DNA reaction networks. *Journal of the Royal Society, Interface / the Royal Society*, 11(93):20131167, 2014.
- [100] N Hansen and A Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2011.

- [101] Nikolaus Hansen, Sibylle D Müller, and Petros Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.
- [102] R T Marler and J S Arora. Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, 2004.
- [103] A E Eiben and J E Smith. Multiobjective Evolutionary Algorithms. *Introduction to Evolutionary Computing*, 2015.
- [104] Anthony J Genot, Teruo Fujii, and Yannick Rondelez. Computing with competition in biochemical networks. *Physical Review Letters*, 109(20):208102, 2012.
- [105] Arthur Prindle, Jangir Selimkhanov, Howard Li, Ivan Razinkov, Lev S Tsimring, and Jeff Hasty. Rapid and tunable post-translational coupling of genetic circuits. *Nature*, 508(7496):387–391, 2014.
- [106] Hendrik W H van Roekel, Lenny H H Meijer, Saeed Masroor, Zandra C Félix Garza, André Estévez-Torres, Yannick Rondelez, Antonios Zagaris, Mark A Peletier, Peter A J Hilbers, and Tom F A de Greef. Automated design of programmable enzyme-driven DNA circuits. *ACS synthetic biology*, 4(6):735–745, 2015.
- [107] Lei Dai, Daan Vorselen, Kirill S Korolev, and Jeff Gore. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science*, 336(6085):1175–1177, 2012.
- [108] Rui Ma, Jichao Wang, Zhonghuai Hou, and Haiyan Liu. Discreteness-induced stochastic steady state in reaction diffusion systems: self-consistent analysis and stochastic simulations. *Physica D: Nonlinear Phenomena*, 205(1-4):248107, 2012.
- [109] Sui Huang. Reprogramming cell fates: reconciling rarity with robustness. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 31(5):546–560, 2009.
- [110] Jeroen Raes and Peer Bork. Molecular eco-systems biology: towards an understanding of community function. *Nature reviews. Microbiology*, 6(9):693–699, 2008.
- [111] Scott F Gilbert. The Generation of Novelty: The Province of Developmental Biology. *Biological Theory*, 1(2):209–212, 2006.

- [112] D H Lee, K Severin, and M R Ghadiri. Autocatalytic networks: the transition from molecular self-replication to molecular ecosystems. *Current opinion in chemical biology*, 1(4):491–496, 1997.
- [113] R Chignola, P Dai Pra, L M Morato, and P Siri. Proliferation and death in a binary environment: a stochastic model of cellular ecosystems. *Bulletin of mathematical biology*, 68(7):1661–1680, 2006.
- [114] Amanda C Williams and Samuel M Flaxman. Can predators assess the quality of their prey’s resource? *Animal Behaviour*, 83(4):883–890, 2012.
- [115] Gui-Quan Sun, Juan Zhang, Li-Peng Song, Zhen Jin, and Bai-Lian Li. Pattern formation of a spatial predator–prey system. *Applied Mathematics and Computation*, 218(22):11151–11162, 2012.
- [116] Adrien Padirac, Teruo Fujii, André Estévez-Torres, and Yannick Rondelez. Spatial waves in synthetic biochemical networks. *Journal of the American Chemical Society*, 2013.
- [117] Nicholas R Markham and Michael Zuker. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research*, 33(Web Server issue):W577–81, 2005.
- [118] Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415, 2003.
- [119] Hans-Ulrich Dodt, Ulrich Leischner, Anja Schierloh, Nina Jährling, Christoph Peter Mauch, Katrin Deininger, Jan Michael Deussing, Matthias Eder, Walter Zieglgänsberger, and Klaus Becker. Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain. *Nature methods*, 4(4):331–336, 2007.
- [120] O Decroly and A Goldbeter. Birhythmicity, chaos, and other patterns of temporal self-organization in a multiply regulated biochemical system. In *Proceedings of the National ...*, 1982.
- [121] Eric Tan, Barbara Erwin, Shale Dames, Tanya Ferguson, Megan Buechel, Bruce Irvine, Karl Voelkerding, and Angelika Niemz. Specific versus non-specific isothermal DNA amplification through thermophilic polymerase and nicking enzyme activities. *Biochemistry*, 47(38):9987–9999, 2008.
- [122] Kasturi Saha, Yoshitomo Okawachi, Bonggu Shim, Jacob S Levy, Reza Salem, Adrea R Johnson, Mark A Foster, Michael R E Lamont, Michal

- Lipson, and Alexander L Gaeta. Modelocking and femtosecond pulse generation in chip-based frequency combs. *Optics express*, 21(1):1335–1343, January 2013.
- [123] S Kijani, U Yrlid, M Heyden, M Levin, and J Borén. Filter-Dense Multicolor Microscopy. *dx.plos.org*.