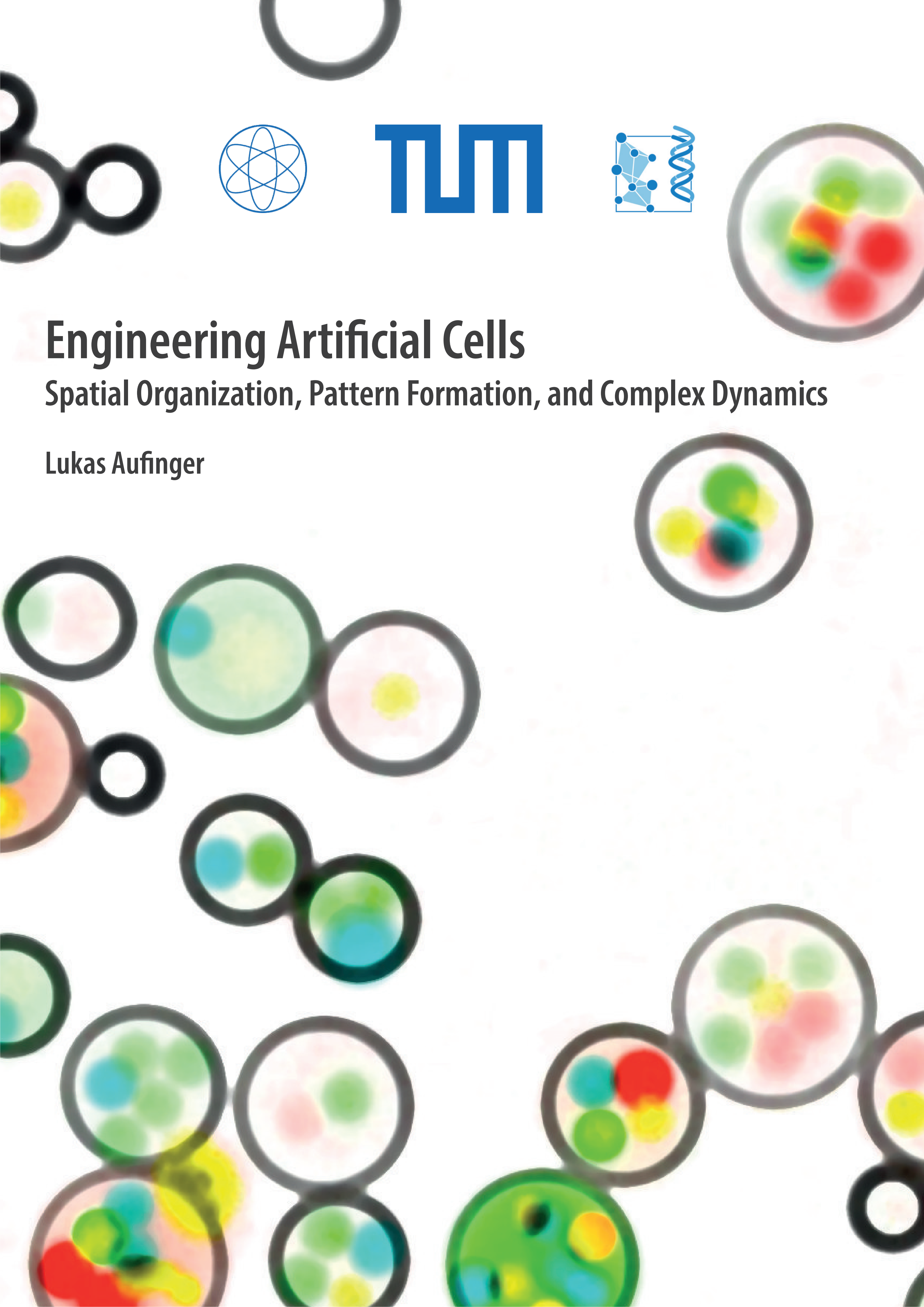
Engineering Artificial Cells

Spatial Organization, Pattern Formation, and Complex Dynamics

Lukas Aufinger



Technische Universität München

Fakultät für Physik

Engeneering Artificial Cells

Spatial Organization, Pattern Formation, and Complex Dynamics

Lukas Aufinger

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitz: Prof. Dr. Martin Zacharias

Prüfer der Dissertation:

1. Prof. Dr. Friedrich C. Simmel
2. Prof. Dr. Ulrich Gerland

Die Dissertation wurde am 27.12.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 14.04.2022 angenommen.

Contents

Acknowledgements **v**

Abstract **vi**

Zusammenfassung **vii**

List of Publications **viii**

Introduction **1**

1 Background and Theory **5**

1.1 Cell-free Gene Regulatory Networks 5

1.1.1 Reaction Environments 7

1.1.2 Gene Regulation 9

1.1.3 Nodes with Multiple Inputs and Outputs 14

1.1.4 Parameter Tuning 17

1.2 Communicating Artificial Cells 21

1.2.1 Vesicles 22

1.2.2 Droplet Interface Bilayer Networks 23

1.2.3 Water-in-oil Emulsion Droplets 25

1.2.4 Hydrogels 27

1.2.5 Microfluidic Reactors 29

1.2.6 Others 32

1.3 Spatial Differentiation and Pattern Formation 33

1.3.1 Patterning Mechanisms 33

1.3.2 Patterning Precision in the Presence of Noise 36

1.3.3 GRN Complexity and Robustness 37
1.3.4 Parameter Optimization using Evolutionary Algorithms 39

2 Material and Methods 41

2.1 Instruments 41
2.1.1 Plate Reader (BMG Labtech, Clariostar) 41
2.1.2 Inverted Epifluorescence Microscope (Nikon, Ti2-E) 41
2.1.3 Inverted Epifluorescence Microscope (Olympus, IX81) and Microfluidics Setup 41
2.1.4 Quake Valve Controller (DIY build) 43
2.1.5 Microfluidic Reservoir Cooling Unit (DIY build) 44
2.2 Microfluidics 46
2.2.1 SU-8 Photolithography 46
2.2.2 AZ 40XT Photolithography 46
2.2.3 Multi-Layer Photolithography 47
2.2.4 PDMS Soft-Lithography 48
2.2.5 Quake Valve Soft-Lithography 48
2.2.6 Droplet Generation 49
2.2.7 Ring reactor 50
2.3 Molecular Biology 53
2.3.1 Gene Template Assembly 53
2.3.2 In Vitro Transcription 54
2.3.3 Cell-free Gene Expression 54
2.3.4 DNA Purification 54
2.3.5 RNA Purification 55
2.3.6 Protein Purification 55
2.3.7 Cell Extract Preparation 56
2.3.8 Preparation of PITC-Agarose Microgel Spheres 56

3 Artificial Gel-based Organelles for Spatial Organization of Cell-free Gene Expression Reactions 59

3.1 Agarose Microgels 60
3.1.1 Functionalization 60
3.1.2 Microgel Synthesis 61

3.1.3 Mesh Size Estimation 61

3.2 Encapsulation Efficiency 62

3.2.1 Statistical Independence 63

3.2.2 Multiple particles with Identity 64

3.3 Transcription and RNA Localization 64

3.3.1 Transcription 64

3.3.2 Transcription and Capture 65

3.4 Spatially Separated Transcription and Translation 66

3.4.1 Toeholdswitch Circuit in Solution 66

3.4.2 Transcription in Gel and Translation in Solution 66

3.4.3 Translation in Gel 66

3.4.4 Transcription in Gel and Translation in Gel 67

3.5 Summary 68

4 Synthetic Cell-based Materials Extract Positional Information from Morphogen Gradients 69

4.1 Minimization of Measurement Uncertainty 71

4.1.1 Image Processing Workflow 72

4.1.2 Normalization 73

4.1.3 Illumination Uncertainties 75

4.1.4 Segmentation Uncertainties 76

4.2 Estimation of Positional Information and Positional Error from Data 78

4.2.1 Estimation of Probability Distribution Functions 78

4.2.2 Calculation of Positional Information 80

4.2.3 Calculation of Positional Error 83

4.3 Positional Error and Positional Information in Droplet Assemblies 85

4.3.1 Experimental Observations 85

4.3.2 ODE Model Reproduces Observations Qualitatively 87

4.3.3 Simulated Impact of Different Noise Sources 88

4.3.4 Tuning the Apparent Threshold Constants 90

4.4 Comparison with the Gap Gene Network of *Drosophila* 91

5 Period Doubling Bifurcations in a Forced Cell-Free Genetic Oscillator 93

5.1 Free Oscillator 94

5.1.1 ODE Model of the Oscillator Circuit 94

5.1.2 Operation of the Free Oscillator 97

5.1.3 Intrinsic Timescale of the Oscillator 97

5.2 Forced Oscillator 98

5.2.1 External Forcing of the Genetic Oscillator 98

5.2.2 Period Doubling 99

5.2.3 Bifurcation Diagram 101

5.2.4 Chaotic Dynamics in the Oscillator Model 101

5.3 Period Doubling in Natural and Synthetic Circadian Clocks 102

Conclusion 105

Bibliography 107

Acknowledgements

First of all, I would like to thank Fritz for the opportunity to freely pursue such exciting research topics. It has been a privilege to work in a lab where everything is possible and I much appreciate your creative input, technical guidance, and support of long shot ideas. I wish to thank Helene, Andrea, and Tobias for ensuring smooth lab operation and a safe work environment, as well as Susanne, and Daniela for making organizational tasks so much easier. I gratefully acknowledge financial support by the European Research Council (project AEDNA, grant agreement no. 694410), the federal state of Bavaria, and BaCaTec. The value of funding creative fundamental research without immediate implications is more apparent then ever.

Further, I wish to thank Elisabeth and Aurore for their advice on cell extract production, purification of TetR, mTurquoise2, and mVenus, and providing several gene fragments. I would also like to thank Aurore for the fun and fruitful collaboration regarding positional information, especially for generating a tremendous amount of beautiful data, as well as Ulrich Gerland for early discussions on this topic.

I would also like to thank Zoe Swank, Sebastian Maerkl, and Nadanai Laohakunakorn for openly sharing their knowledge and tricks on their microfluidic ring reactor setup, device manufacturing, and operation. Thanks to Joe for hacking the microscope and implementing the device operation and automation software for the ring reactor setup, and to Katzi for help with the design of electronics and with repairs of critical devices.

Finally, I wish to thank my now graduated students Katzi and Anna for diligently pursuing their projects, and especially my previous supervisor Matthaeus for teaching my everything and sparking my interest in genetic oscillators and cell-free systems. I want to thank Anna, Katzi, Thomas F., Louis, and John for proof-reading this thesis, as well as all my colleagues at E14, my family, and friends for the great time and support over the last years.

Abstract

A central goal of bottom-up synthetic biology is to replicate the complex phenomena that are observed in biological organisms. Here, we explore the concepts of spatial organization, pattern formation, and temporal synchronization by implementing cell-free gene regulatory networks in the simplified environment of artificial cells. This work paves the way towards synthetic organisms that can be used for technological applications or serve as minimal systems to study fundamental principles of life.

Zusammenfassung

Ein zentrales Ziel der Bottom-up Synthetischen Biologie ist die Replikation komplexer Phänomene von biologischen Organismen. Hier untersuchen wir Konzepte wie räumliche Organisation, Musterbildung und zeitliche Synchronisation mithilfe zellfreier Genregulationsnetzwerke in der vereinfachten Umgebung künstlicher Zellen. Diese Arbeit ebnet den Weg für den Einsatz synthetischer Organismen in technischen Anwendungen, oder als minimale Systeme für das Studium fundamentaler Prinzipien des Lebens.

List of Publications

L. Aufinger, J. Brenner, and F. C. Simmel. “Period Doubling Bifurcations in a Forced Cell-Free Genetic Oscillator”. *bioRxiv* (preprint). DOI: `10.1101/2021.12.15.472802`

A. Dupin*, **L. Aufinger***, I. Styazhkin, F. Rothfischer, B. Kaufmann, S. Schwarz, N. Galensowske, H. Clausen-Schaumann, and F. C. Simmel. “Synthetic cell-based materials extract positional information from morphogen gradients”. *bioRxiv* (preprint). DOI: `10.1101/2021.04.25.441320`

J. Müller, A. C. Jäkel, D. Schwarz, **L. Aufinger**, and F. C. Simmel. “Programming Diffusion and Localization of DNA Signals in 3D-Printed DNA-Functionalized Hydrogels”. *Small* 16.31 (2020), p. 2001815. DOI: `10.1002/smll.202001815`

F. Katzmeier*, **L. Aufinger***, A. Dupin*, J. Quintero*, M. Lenz, L. Bauer, S. Klumpe, D. Sherpa, B. Dürr, M. Honemann, I. Styazhkin, F. C. Simmel, and M. Heymann. “A low-cost fluorescence reader for in vitro transcription and nucleic acid detection with Cas13a”. *PLOS ONE* 14.12 (2019), e0220091. DOI: `10.1371/journal.pone.0220091`

L. Aufinger and F. C. Simmel. “Establishing communication between artificial cells”. *Chemistry - A European Journal* 25.55 (2019), pp. 12659–12670. DOI: `10.1002/chem.201901726`

L. Aufinger and F. C. Simmel. “Artificial Gel-based Organelles for Spatial Organization of Cell-free Gene Expression Reactions”. *Angewandte Chemie International Edition* 57.52 (2018), pp. 17245–17248. DOI: `10.1002/anie.201809374`

* These authors contributed equally.

Introduction

Biological organisms are self-sustaining complex systems that self-organize according to a genetic program encoded in their DNA. This molecular blueprint has been optimized through billions of years of evolution in the earth's environment to exhibit fascinating capabilities such as cognition, communication, or efficient chemical synthesis. These processes can be classified as self-organization phenomena that describe the (spontaneous) generation of order, emerging from internal interactions between the system's components. From a statistical physics point-of-view, self-organizing systems, while not requiring an external operator, have to obey the 2^{nd} law of thermodynamics, which implies that organisms are open systems that generate local order by consuming energy and producing entropy [1].

In contrast, complex man-made technologies are often a product of large organizations, careful design by engineers, and assembly of precision manufactured parts. Realizing these differences, it is natural to ask whether we can rationally engineer self-organizing systems to open a variety of exciting opportunities for technological applications. For instance, this could enable the creation of functional materials, or even 'synthetic organisms' at scales ranging from kilometers down to nanometers. In analogy to plants, synthetic organisms could for instance be grown inexpensively from seed, consuming only water, CO_2, and sunlight. The life cycle of synthetic organisms would be intrinsically circular, self-sustainable and not generate permanent waste. Most importantly, however, synthetic organisms may to some degree be able to heal themselves, eliminating the need for frequent maintenance and repairs, in contrast to many man-made machines such as Olympus microscopes.

The design of self-organization phenomena is a major challenge of synthetic biology, which seeks to identify and refine design processes to engineer systems of similar complexity as the natural examples that emerged through evolution. Even simple, single-cellular organisms like bacteria regulate their behavior with complex gene regulatory networks (GRNs) that may involve thousands of genes, nonlinearity, and feedback and are hence intrinsically difficult to understand. How genes interact in complex GRNs on an abstract level is studied by systems biology [2]. A central approach is to treat GRNs as a network graph and identify basic features such as network motifs based on graph theory [2–4]. Systems biology builds on a rich foundation of knowledge of the sequence and function of genes that was obtained by molecular biologists using a complementary reductionist approach.

The disciplines of molecular, systems, and synthetic biology approach complexity from several perspectives. First, a system with many components may be described by systems level principles that emerge from interactions between its components [5]. Second, there are relatively simple systems that create complicated emergent behavior as described by nonlinear dynamics and chaos theory [6]. Third, synthetic biologists follow a constructionist approach to build increasingly complex systems from the bottom-up [7]. Together, these approaches may lead to a comprehensive understanding of biological as well as synthetic organisms.

The first demonstrations of synthetic biology were the implementation of simple sub-systems like memory [8], communication [9], and oscillation [10] into existing living cells. These works showed that rational design and engineering could be used to manipulate and expand the functionality of the complex GRNs in biological organisms. Since its origins, the ability to design synthetic genetic circuits rapidly expanded through the adoption of traditional engineering principles like modularity, standardization of parts, parameter tuning, computer-aided design, and rapid prototyping, combined with an expanding toolkit of gene synthesis, sequencing, and manipulation techniques like PCR and CRISPR [11, 12]. This development enabled to engineer the behavior of whole systems of bacteria to generate patterns [13], oscillate in synchrony [14], or communicate between compartmentalized colonies [15].

Besides the engineering of novel behavior in existing biological cells, one major branch of synthetic biology is the creation of artificial cells. Artificial cells are cell-sized compartments that mimic some aspects of natural cells and are constructed in a bottom-up approach [16]. While one motivation is to create the basic building block of a self-replicating synthetic life form, e.g. to find the minimum requirements for life [17, 18], artificial cells can also provide a simplified environment that facilitates the engineering of cellular functions in the absence of a complex genetic background. In some cases, this approach may even generate insight, raise new questions, or confirm concepts about biological systems [19].

One common approach to generate artificial cells is to compartmentalize the transcription-translation machinery extracted from natural cells. The compartments, for instance vesicles, are then able to express proteins such as green fluorescent protein (GFP) [20]. While, as stated somewhat ironically by Luisi, Ferri, and Stano [18] in 2006: "[...] after having produced GFP, none of the systems [...] has been found capable of reproducing itself and giving rise to a chain of multiplying GFP-producing systems," there are now promising efforts to utilize the original transcription-translation machinery to regenerate some of its components [21]. Compartmentalized cell extract may hence indeed be a potential starting point for synthetic organisms.

In lieu of a truly self-sufficient artificial cell, it is still possible to explore aspects of self-organization either by utilizing the finite energy resources in a closed reaction or by continuously supplying the reactions with fresh machinery. For example, such artificial cells have been used to implement oscillations [22, 23], bistability [24], and communication [25, 26], as well as systems level phenomena like pattern formation [27], synchronized oscillations [28], and multi-cellular spatial organization [29]. From here, one goal is to gradually increase the complexity of the GRNs in the simplified environment of the artificial cells.

Synthetic biologists have suggested and developed a variety of methods and design principles to realize more complex synthetic GRNs. For example, computer-aided design combined with large parts libraries enables the automated design of large logic networks [30]. Spatial organization can be used to create a modular hierarchy to minimize undesired side reactions and cross-talk [31]. The design of pattern forming GRNs can be based on simple core-motifs and regulatory mechanisms [32], whereas established design principles for genetic oscillators are negative feedback, time delay, and non-linearity [33]. In most cases, reaction turnover and energy supply are crucial factors to keep systems in out-of-equilibrium conditions [34]. To account for variations in environmental parameters or manufacturing, core-motifs can be expanded towards more robust GRN topologies [35]. It is now crucial to further optimize the design-build-test cycle by refining design principles, developing experimental methods and molecular mechanisms, and quantifying how precisely self-organizing synthetic systems achieve their intended function.

In this thesis, we attempt to bridge the gap between externally assembled and self-organizing technology by exploring artificial cells controlled by cell-free GRNs. To this end, we focus on three key aspects: spatial separation of reaction components, spatial differentiation in the presence of noise, and temporal synchronization of a genetic oscillator. In Chapter 1 we first present the theoretical principles and the molecular building blocks that are available to construct GRNs, communicating artificial cells [36], and spatial differentiation. In Chapter 2 we describe the microfluidic techniques and instruments, as well as molecular techniques that were used to implement and characterize the experiments.

With this set of tools, we first equipped artificial cells with artificial gel-organelles that spatially organize the reaction components of a simple GRN [37] (Chapter 3). By localizing genes and mRNA in dedicated organelles, we spatially decouple a cell-free transcription-translation reaction. Conceptually, this approach can aid the realization of more complex GRNs by mitigating undesired cross-reactions and facilitating a modular design.

In Chapter 4 we then investigate the robustness of an engineered patterning process with respect to molecular noise and manufacturing tolerances [38]. Thereto, we utilize the concept of positional information to quantify how well a simple GRN embedded in assemblies of artificial cells can differentiate those cells by interpreting a morphogen gradient. This analysis allows to pinpoint the sources of variability in the final patterns and hence suggests routes to increase the fidelity of molecular patterning processes.

Finally, we study the synchronization of an oscillating GRN with an external clock [39] (Chapter 5). To this end, we utilize a semi-continuous microfluidic ring-reactor to keep the reactions out of equilibrium for up to 48 h. The reactor was programmed to periodically expose the genetic oscillator to molecular inputs. Under certain conditions, periodic forcing leads to non-trivial non-linear phenomena such as period doubling. Again, we suggest design principles regarding the use of molecular oscillators as timekeepers.

We conclude by summarizing our results and highlight our key findings and technological advances. We then identify current experimental constraints, such as energy supply and GRN design, and discuss potential routes for future developments. Finally, we discuss the perspectives and applications of self-organizing synthetic organisms and potential implications for biological organisms.

1 Background and Theory

In order to engineer a community of artificial cells (ACs) that is capable of spatial and temporal self-organization and differentiation, the ACs need to be able to communicate via signaling molecules, sense environmental cues, and process the contained information to trigger an appropriate response. This capability can be implemented using gene regulatory networks (GRNs), which play a key role in many natural pattern formation systems, such as the gap gene network in *Drosophila* [40]. GRNs were thoroughly studied by system biologists who identified the core network motifs and mechanisms required for robust pattern formation [41–43]. Although natural systems often involve additional regulatory motifs, synthetic biologists have successfully designed and implemented pattern forming GRNs in natural and artificial cells based purely on core motifs [13, 27, 35, 44–47]. As described in Section 1.1 there are several experimental platforms that provide reaction environments to implement GRNs that can sense and process multiple input signals in a quantitative and tunable manner.

The transport of signaling molecules through an ensemble of ACs is mostly mediated by diffusion and therefore largely dependent on material properties of the ACs and their surroundings [36]. Hence, there is a limited set of signaling molecules that can simultaneously promote communication among a given type of ACs and be sensed and synthesized in a given reaction environment. As reviewed in Section 1.2, the nature of the ACs also affects other aspects like energy supply and reaction turnover that crucially affect the systems dynamics.

If engineered appropriately, the communication within a consortium of identical ACs, combined with the signal processing by a GRN, can generate complex collective behavior. Focusing on pattern formation, we first present the most fundamental mechanisms that guide the design of an artificial patterning system (Section 1.3). From an engineering perspective it is furthermore interesting, to quantify the precision of a resulting pattern and to study the complexity of a patterning system against its robustness. Finally, a parameter optimization strategy for any given patterning system with tunable GRN and communication parameters is presented.

1.1 Cell-free Gene Regulatory Networks

GRNs can be visualized by a network graph as shown in Figure 1.1a. A node represents a molecular species, often a regulatory protein, an RNA, or a small molecule. The edges represent the interaction between two nodes, which involves transcriptional as well as translational regulation. GRNs can be implemented in a range of reaction environments such as in vitro transcription and cell-free gene expression systems that each support different sets of reactions and regulatory mechanisms (Figure 1.1b, Section 1.1.1).

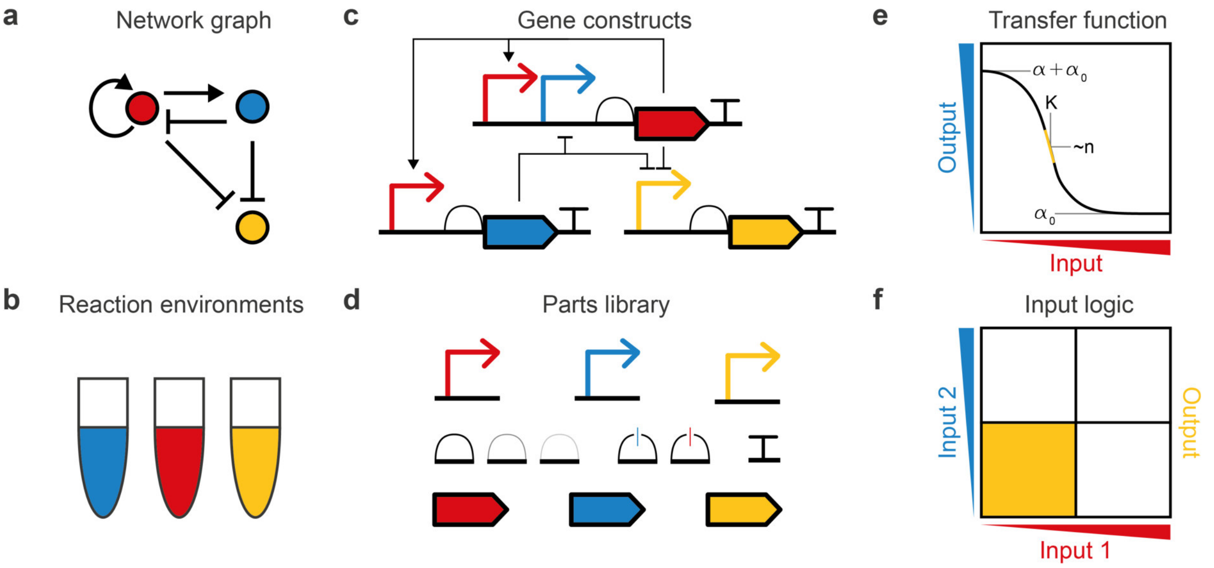


Figure 1.1: Gene Regulatory Networks. **a)** An exemplary network graph where colored circles represent nodes, arrows indicate an activating interaction, and bar-ended lines indicate inhibitory interaction. **b)** GRNs can be implemented in different reaction environments. **c)** A network graph can be translated into gene constructs that exhibit the desired regulatory interactions. **d)** A parts library offers a set of promoters (arrows), ribosome binding sites (half circles), ribo-switches (half circles with switches), terminators (T) and genes (squarrows). Each part may have different strengths or operators to promote different regulatory interactions. **e)** The input-output relation or transfer function of an individual node can often be described by Hill-type functions. **f)** Nodes with multiple inputs can integrate these signals with different logic, depending on the construct's architecture.

To convert a network graph into gene constructs (Figure 1.1c) engineers can use a variety of pre-characterized parts libraries (Figure 1.1d, Section 1.1.2). Although most regulatory mechanisms consist of multiple reactions, the corresponding input-output relation or transfer function can often be described by a Hill-type equation $f(c_{in})$ (Figure 1.1e, Section 1.1.2) with first-order degradation

$$\dot{c}_{out} = f(c_{in}) - \delta c_{out} = \alpha \frac{1}{1+(K/c_{in})^n} + \alpha_0 - \delta c_{out} \ . \tag{1.1}$$

Here, c_{in} and c_{out} are the concentrations of input and output, respectively, $\alpha_{max} = \alpha + \alpha_0$ is the maximum expression rate, α_0 is the minimum, or leak expression rate, K is the threshold constant, n is the Hill coefficient describing the steepness of the regulatory interaction, and δ is the degradation rate.

Generally, GRNs may consist of nodes with multiple inputs or outputs. In this case, multiple input signals can be processed with different logic (Figure 1.1f), which, as outlined in Section 1.1.3, depends on the architecture of the gene constructs. Finally, the parameters α, K, and n that describe the regulatory interactions can be fine-tuned within constraints to obtain the desired behavior. To this end, several methods are illustrated in Section 1.1.4.

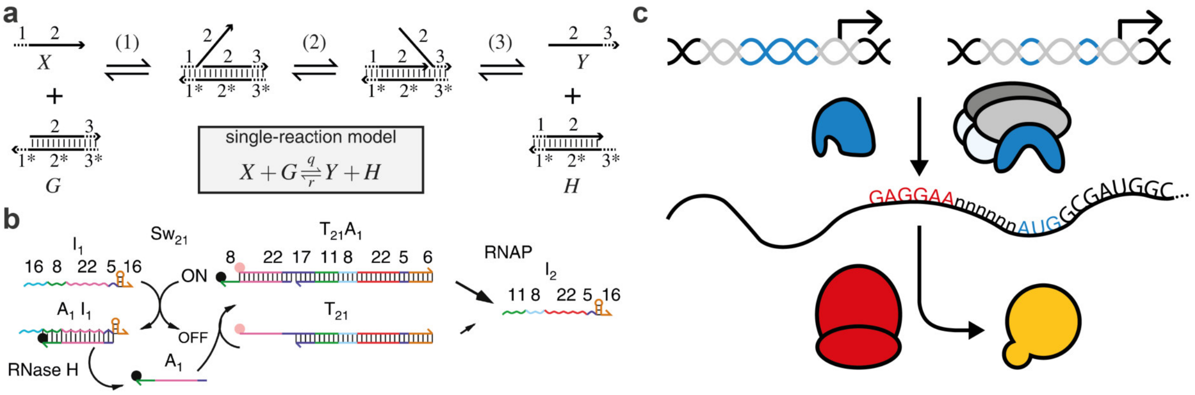


Figure 1.2: Reaction environments. a) TMSD reactions can be used to emulate arbitrary chemical reaction networks. (1) Input strand X binds to toehold domain 1^* and (2) initiates strand displacement of output Y. (3) Y dissociates. Reproduced from reference [48] under the PNAS open access option. **b)** A genelet can be used to regulate an IVT reaction. The state of the transcription template T_{21} is controlled by an activator strand A_1. Upon addition of input strand I_1, A_1 is removed from the template via TMSD and transcription is switched off. Transcription can be switched on by degradation of I_1 by RNase H to release A_1, resulting in the production of output I_2. Reproduced from reference [49] with permission of John Wiley and Sons, © 2006. **c)** Simplified illustration of a transcription-translation reaction in a CFPE system. By default, PURE uses T7 RNA polymerase (top-left), whereas TXTL uses *E. coli* RNA polymerase (top-right) to transcribe a DNA template into mRNA. The ribosome (red) binds the Shine-Dalgano sequence (red) and proceeds to translate the mRNA producing an output protein (yellow). Both reactions can be regulated by proteins or RNA as input molecules, as detailed in Figure 1.3.

1.1.1 Reaction Environments

There are several biochemical reaction environments that are commonly used to implement in vitro GRNs (Figure 1.2). We start by briefly introducing DNA-only, enzyme-free circuits based on toehold mediated strand displacement (TMSD) [50–52], then introduce in vitro transcription (IVT) systems [49, 53], and finally focus on cell-free protein expression (CFPE) systems, that enable transcription-translation networks. For CFPE there are mainly two complementary approaches, crude cell extract (often abbreviated TXTL for transcription/ translation) [54] and Protein Synthesis Using Recombinant Elements (PURE) [55, 56].

As shown in Table 1.1 the systems differ in terms of their complexity, e.g. measured by the number of proteins in the reaction mixture, and the capability to synthesize molecules. The form of DNA from which the nodes are constructed affects the time and cost to implement the GRN. While DNA oligos up to 90 bp for TMSD and some genelet circuits can be ordered from DNA synthesis companies in a few days, cloning techniques for the construction of plasmid gene templates for TXTL reactions can require several weeks. PCR-based techniques to produce linear gene templates can reduce the prototyping time down to a single day [57]. While commercial kits for IVT and CFPE reactions can come at a rather prohibitive cost, the cost per reaction drops significantly when using homemade solutions [54, 58].

TMSD is a reaction where one strand of DNA is displaced from a duplex by a third DNA strand initiated by a single-stranded overhang called toehold [50, 52] (Figure 1.2a). DNA strands can be computationally designed based on thermodynamic models [66] and the TMSD reaction rate constants can be tuned over 5 orders of magnitude by changing the length of the toehold [51]. This allows to implement arbitrary GRNs [48], such as logic circuits [67] or neural networks [68] that feature a large number of components [69]. The scaling of TMSD reaction circuits suffers from side reactions such as leaky strand displacement (SD without a toehold) which are amplified at high concentrations. As a consequence, reactions are typically performed at low concentrations which compromises reaction speed. Fast reactions with many components at high concentrations can be achieved using a leakless TMSD design [70].

Expanding on this reaction framework, IVT circuits utilize RNA polymerases, such as the single-subunit ≈ 90 kDa T7 RNA polymerase, that bind to the promoter sequence of a DNA template to transcribe RNA. Wiring of an IVT circuit requires RNA regulated transcription of RNA, which can for instance be achieved via switchable transcription templates termed 'genelets' [49, 71] (Figure 1.2b). Genelets are based on an incomplete promoter that can be completed with an activator strand to regulate the activity of the T7 RNA polymerase. Using TMSD, the output of one genelet can be coupled to the input of another.

This system was used to realize a transcriptional batch oscillator [53] in which degradation of RNA is realized by RNase H, and energy is supplied in the form of stock rNTPs. This oscillator was used to control the actuation of molecular machines [72] and later even linked to the self-assembly of µm-sized DNA nanotubes [73]. When the oscillator components were partitioned into emulsion droplets it was observed that periods and amplitudes varied more than naively expected [74].

To broaden the scope of IVT networks, there are a variety of alternative single-subunit phage polymerases that are commercially available and orthogonal to the T7 RNA polymerase, such as T3 and SP6 RNA polymerase, as well as orthogonal T7 variants found via directed evolution [75]. As an alternative regulation mechanism it is possible to inhibit T7 and SP6 RNA polymerase with RNA aptamers [76].

One drawback of IVT reactions is that T7 RNA polymerase is known to produce unspecific RNA products, e.g. by transcribing from ssDNA [77] or self-coded 3'-extension [78, 79]. This limits the scalability of IVT networks as the side reactions can occur on the single stranded 5' overhang of the genelet design as well as on the activator strand, causing inhibition of the genelet. By introducing a hairpin clamp (HPC) genelet design, this undesired side reaction can be sufficiently mitigated, so that residual inhibitory side products could be degraded by RNaseH [80]. Interestingly, Wu et al. [81] recently discovered a T7 RNA polymerase mutant with strongly reduced unspecific activity, which may be used to further improve the fidelity of future IVT circuits.

CFPE systems further expand on IVT networks by adding the capability to synthesize proteins which allows to closely mimic natural GRNs. A simplified reaction scheme is shown in Figure 1.2c. First, RNA polymerase generates mRNA by binding to a promoter and transcribing the gene template.

Table 1.1: Comparison of reaction environments. Costs per reaction were estimated as follows: [a] The cost is limited by oligo synthesis. We conservatively assume the cost of a network with 50 components with about 500 €, a yield of > 5 nmol and a final concentration of 100 nM. [b] Using homemade T7 RNA polymerase and assuming rNTPs are cost limiting (Jena Bioscience, #NU-1024L). [c] HiScribe T7 High Yield RNA Synthesis Kit (NEB, #E2040S). [d] [58]. [e] PURExpress (NEB, #E6800L). [f] [54]. [g] myTXTL Sigma 70 Master Mix Kit (BioCat, #507096-ARB). Conversion rate 0.86 € /$. Last updated in September 2021.

system	TMSD	IVT	PURE	TXTL
no. of proteins	0	1-2	36 + ribosomes [56]	≈ 800 [59]
synthesis capability	none	RNA	RNA, proteins, DNA [60]	RNA, proteins, DNA [61], ATP [62–64], metabolites [65], small molecules [25]
reactions	TMSD	TMSD, genelets	T7, toehold switches	transcription factors, sigma factors
DNA construction	oligo synthesis	oligo synthesis, PCR	PCR, plasmid	plasmid, PCR
cost/10 µL	< 0.1 € [a]	0.1 € [b] (2.3 € [c])	0.8 € [d] (9.2 € [e])	0.25 € [f] (8.8 € [g])

Then, ribosomes bind the mRNA and decode the sequence into a chain of amino acids which then folds into the mature protein. To move from IVT to CFPE the reaction mix hence has to contain the essential translation machinery.

The PURE system (Protein synthesis using recombinant elements) consists of ribosomes and 36 individually purified proteins including translation factors, tRNAs, tRNA synthetases, kinases and pyrophosphatase for energy regeneration, and T7 RNA polymerase [55, 56]. The PURE system is well defined and due to the absence of nucleases is compatible with systems involving linear or single stranded DNA.

PURE has been used for DNA self replication [60], therapeutics [82], and diagnostics [83]. Although PURE can be used to implement GRNs, it was found that it was difficult to engineer a strong promoter with tight repression using phage polymerases, and supplementing *E. coli* RNA polymerase lead to low transcription rates [23]. Due to its minimalistic nature, PURE is often considered as an ideal starting point for the creation of self-regenerating artificial cells [21, 84]. For applications for which the cost of commercial kits (Table 1.1) can be prohibitive, a simplified manufacturing process with one-pot purification of the 36 PURE proteins allows the production of homemade PURE at a more than 10-fold reduced cost [58].

In contrast to PURE, *E. coli* S30 cell extract is produced by taking crude *E. coli* cell lysate and progressively removing cell wall material and genomic DNA. As all transcription and translation machinery is extracted in a single step, production cost is relatively low [54]. The extract contains the core *E. coli* RNA polymerase, a 5 subunit, 379 kDa enzyme complex [85], ribosomes and other machinery for translation, but also several nucleases, proteases and transcription factors that can potentially interfere with a designed GRN.t Energy regeneration can be achieved via different metabolic pathways that are conserved from the host cells [63, 64, 86].

Cell extract was used to create artificial cells [20], and to construct GRNs [22, 24, 25, 28, 29, 87–92] based on a rich repertoire of transcription factors that regulate the *E. coli* RNA polymerase [93–95], and a thorough theoretical understanding [96, 97]. Remarkably, results from cell-free circuit prototyping have been shown to be portable to live *E. coli* cells, paving the road to use cell extract as a rapid prototyping platform [89, 98]. The drawback that the nucleases contained in the cell extract digest linear DNA products rapidly generated by PCR can be avoided by adding inhibitory proteins GamS [57], Ku [99], or short linear DNA containing $\mathcal{X}$-sites [100] to target the RecBCD complex [101], which is the main source of degradation.

1.1.2 Gene Regulation

Depending on the reaction environment, the nodes of a synthetic GRN are constructed of DNA, RNA, or proteins, whereas the edges represent transcriptional or translational regulation. In this section we first give an overview of common mechanisms and components for both types of regulation and then motivate a generic modeling framework.

Transcription can, in general, be divided into three phases: initiation, elongation and termination [104, p.341]. During the initiation phase, the RNA polymerase first binds to the promoter. The promoter sequence for phage RNA polymerases is typically ≈20 bp long [105], while binding of *E. coli* RNA polymerase is mediated by the σ-factor sub-unit, which recognizes two ≈6 bp recognition boxes located -10 and -35 bp upstream of the transcription start site. The promoter sequence determines the affinity of the RNA polymerase for the promoter and therefore promoter strength.

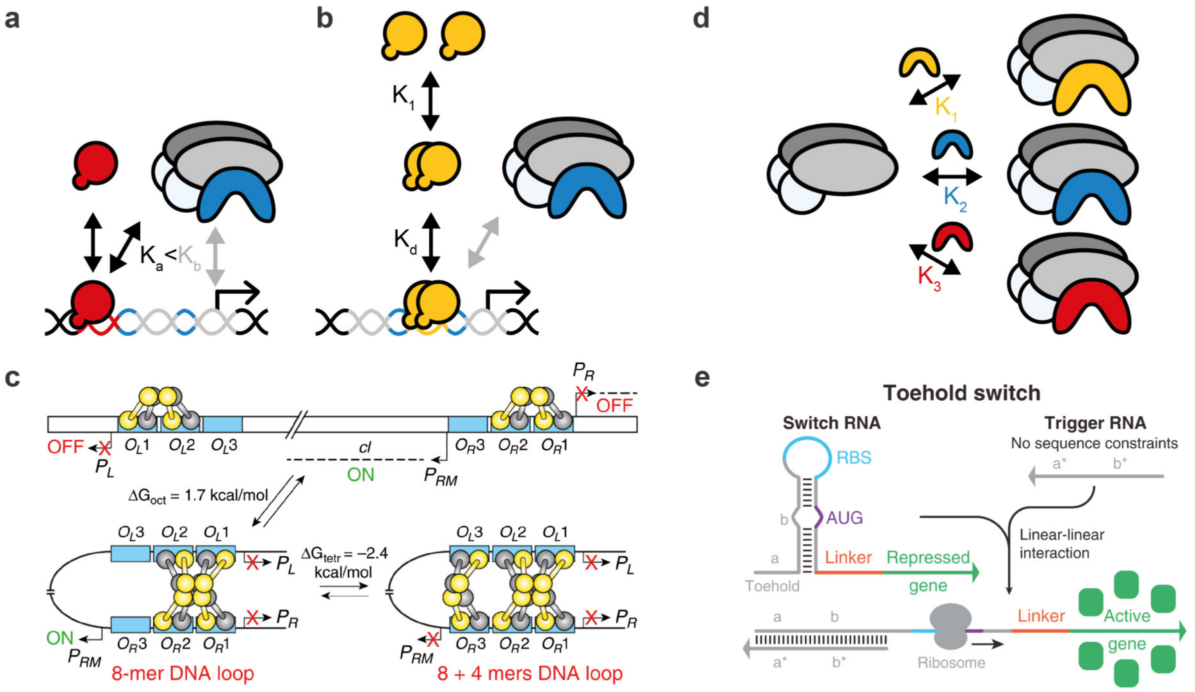


Figure 1.3: Gene Regulation. a) Cartoon illustration of gene activation. The affinity of the polymerase for its promoter increases when activator (red) is bound to an operator sequence upstream of the -10/-35 boxes (blue). **b)** Example for cooperative binding of a repressor in a 2-step reaction, where the repressor first forms a dimer with K_1 and then binds the promoter with K_d. The form of the regulatory function depends on which reaction is limiting. **c)** Example for transcriptional regulation in a natural system. The repressor cI, which is involved in the regulation of the lysogenic cycle of λ-phage, regulates its own production. The mechanism involves 3 pairs of operators that are bound in a hierarchical order with increasing repressor concentration. The formation of tetramers and octamers causes DNA loop formation. Reproduced from reference [102] under the *PNAS license*, © 2011 the authors. **d)** Sigma factors compete for core RNA polymerase. **e)** A toehold switch is an example for synthetic translational regulation. In the inactive state RBS and start codon are repressed in a stem loop. A trigger strand can bind the toehold and and activate translation via TMSD. Reproduced from reference [103] with permission of Elsevier, © 2014.

After binding, the polymerase proceeds to unwind the DNA helix to form the open transcription complex. During the elongation phase RNA is produced by incorporating rNTPs according to the DNA template, until the reaction is terminated, typically by a hairpin stem that induces unbinding of the RNA polymerase.

Transcriptional regulation by external inputs can occur during all of the three phases, for instance via transcription factors (TFs) that interfere with the binding of RNA polymerase to the promoter. To activate or repress transcription, TFs typically bind to an ≈ 20 bp long operator sequence. Well studied examples for activators or repressors that recrute or block RNA polymerase (Figure 1.3a,b) in natural GRNs are LuxR [106, 107], AraC [108], TetR [109], LacI [110, 111], or cI [112]. Some TFs like LuxR can function as both activator and repressor, depending on the location of the operator site within the promoter [113]. Some TFs can serve as molecular sensors by specifically binding to small molecules. Upon binding, TFs alter their conformation which changes the affinity to its operator to induce gene expression. Note, that regulation of phage RNA polymerases by protein-based TFs is possible [114, 115], but often less effective than regulation of *E. coli* RNA polymerase [23].

To construct a synthetic GRN, an engineer can either resort to TFs found in natural systems, or design customized synthetic TFs. Synthetic TFs can be designed de novo using programmable systems such as zinc-finger arrays [116], transcription activator-like effectors (TALEs) [117, 118], or CRISPR interference [119–121]. These systems can be targeted to custom defined operator sequences and come with varying size, specificity, affinity, and design difficulty. Some of these properties can be tuned by characterizing combinatorial libraries [92]. Alternatively, synthetic TFs can be derived by modifying existing ones. For instance, continuous evolution was used to discover a 63-amino acid small dual activator-repressor peptide [122]. By combining natural domains for ligand binding with synthetic DNA recognition domains it was also possible to design inducible TFs with custom defined recognition sites [123].

How a given TF regulates a gene depends on the design of the promoter. Natural promoters can involve a sophisticated interplay of various mechanisms. For instance, regulation of the lysogenic cycle of phage λ by the repressor cI involves multimerization, loop formation, and multiple operators with different affinities [102] (Figure 1.3c). For synthetic promoters, simpler, smaller, and more modular designs are often preferable. For instance, operator sites can be placed upstream (distal), downstream (proximal) or between (core) the -10 and -35 boxes to construct synthetic hybrid promoters [124]. Rules for the design of synthetic promoters that respond to multiple inputs are discussed in Section 1.1.3.

Another factor that influences promoter occupancy are the concentrations of RNA polymerase and DNA template themselves. For instance, single-sub-unit phage polymerases can be used analogously to transcriptional activators [125]. Because *E. coli* RNA polymerase is larger, the concentration of core polymerase is usually not altered. However, the affinity of the RNA polymerase to specific promoters can be regulated by expressing different σ factors. Sigma factors are used to respond to different stresses by regulating the expression of several stress-response genes. For example, σ^{70} is the housekeeping factor [126], σ^{32} regulates the response to heat shock [127], σ^{28} is used to express flagellae, i.e. control mobility [128], and σ^{54} triggers the response to nitrogen starvation [129]. The corresponding promoters are characterized by -10 and -35 boxes that are specific to each sigma factor. Because the different sigma factors compete for core RNA polymerase (Figure 1.3d), they tend to not only up-regulate their target genes, but also down-regulate others [130].

Gene expression can also be regulated on a translational level. The basal translation rate depends largely on the 5' untranslated region (5' UTR), particularly on the ribosome binding site (RBS), its distance to the start codon (AUG), and secondary structure. The RBS strength for a given gene can be predicted based on thermodynamic models [131]. Translational regulation can for instance be engineered using toehold switches [103] (Figure 1.3e) that respond to short single stranded RNA or DNA inputs, or synthetic riboswitches [132–134] that respond to small inducer molecules by exposing/ repressing the RBS. Such riboregulators can be designed de novo, can feature multi-input logic [135], and can be used on paper-based reactions [136] to detect RNA viruses [83] or single-nucleotide-polymorphisms (SNPs) [137].

Although transcription and translation reactions are complicated multi-step reactions that involve hundreds of molecules, the time scale for TF binding and RNA and protein production reactions are typically much smaller (seconds to minutes) than typical system time scales, as set, e.g., by degradation rates (hours) [2, pp.9]. Hence, transcription and translation rates are approximately proportional to the fraction of promoter that is bound to polymerase or ribosomes bound to the RBS, respectively. In the following, we outline a typical modeling approach for transcriptional regulation. While under conditions where translation rates are not limiting or regulated, treating transcription-translation reactions as a single-step reaction may be sufficient to accurately describe the system, extending the model to explicitly include translational regulation follows the same principles [96, 97, 138].

To model promoter binding, we consider the simple binding reaction

$$\mathrm{X} + \mathrm{R} \rightleftharpoons \mathrm{C} \,, \tag{1.2}$$

where X denotes the free promoter, R the free RNA polymerase and C is the polymerase-promoter complex. At steady-state conditions

$$\frac{dx}{dt} = -k_{+}xr + k_{\_}c = 0 \,, \tag{1.3}$$

with the association and dissociation rates k_+ and k_-, respectively. Lower-case x, r, and c denote the concentration of X, R and C, respectively. After rearranging and including conservation of mass we are left with the following system of equations

$$K_d = \frac{k_-}{k_+} = \frac{xr}{c} \,, \tag{1.4}$$

$$x_0 = x + c \,, \tag{1.5}$$

$$r_0 = r + c \,. \tag{1.6}$$

Here, x_0 and r_0 denote the total promoter and polymerase concentration, respectively, which are typically the system variables. K_d is the equilibrium dissociation constant for the binding of RNA polymerase to its promoter. Combining these equations gives

$$0 = \frac{x^2}{K_d} + x \cdot \left(1 + \frac{r_0}{K_d} - \frac{x_0}{K_d}\right) - x_0 \,, \tag{1.7}$$

which solved for x yields the exact solutions

$$x_{1,2} = -\frac{K_d}{2}\left(1 + \frac{r_0}{K_d} - \frac{x_0}{K_d}\right) \pm \frac{K_d}{2}\sqrt{\left(1 + \frac{r_0}{K_d} - \frac{x_0}{K_d}\right)^2 + 4\frac{x_0}{K_d}} \,. \tag{1.8}$$

Because $x > 0$, we keep only the positive solution. Because exact solutions for most regulatory binding reactions besides this simple case are impractical, we can often approximate by assuming that $x_0 \ll r_0$. Here, series expansion to first order, or realizing that $x_0 \ll r_0$ implies $r_0 \approx r$ and rearranging Equation (1.4) to (1.6) gives

$$\frac{x}{x_0} = \frac{1}{1 + \frac{r_0}{K_d}} \,. \tag{1.9}$$

Since we were interested in the fraction of bound promoter, the transcription rate as a function of polymerase concentration r_0 is

$$\dot{c}_{out} = \alpha(x_0 - x) = \frac{\alpha x_0 r_0}{K_d + r_0} \; , \tag{1.10}$$

where $\alpha \cdot x_0$ is the maximum transcription rate, which depends on the promoter sequence and RNA polymerase. Note, that here K_d is defined in an isolated context, however, realistic systems may exhibit unspecific binding sites due to the genomic background [139, 140]. In this case, K_d can be replaced by a phenomenological apparent binding constant K_{app}. In Section 1.1.4 we discuss a mechanism for which we can derive an approximate equation for K_{app} that depends on experimental variables.

To mathematically describe gene regulation by TFs, we can in principle apply the same equilibrium analysis as performed above for the binding of RNA polymerase. The transcription rate will be determined by the fraction of free operator concentration. However, we need to account for the specific mechanism, as multiple TFs may bind the same operator cooperatively. For instance, we can consider the binding of n monomers $\mathrm{R_M}$ that form a repressor R to an operator X in a multi-step reaction

$$\mathrm{X} + \mathrm{nR_M} \overset{K_1}{\rightleftharpoons} \mathrm{X} + \mathrm{R} \overset{K_d}{\rightleftharpoons} \mathrm{C} \; . \tag{1.11}$$

Instead of binding directly to the promoter, the TF first forms an n-mer with dissociation rate K_1, which then binds the operator with dissociation rate K_d. Alternatively, the reaction could start with the binding of one monomer which then facilitates the binding of additional monomers, which would lead to a similar result.

Following Sneppen and Zocchi [141, pp.160], the first equilibrium gives $K_1 = \frac{r_M^n}{r}$ and the second is given by Equation (1.10). Inserting r gives

$$\frac{x}{x_0} = \frac{K_d K_1}{K_d K_1 + r_M^n} = \frac{K_{app}^n}{K_{app}^n + r_M^n} \; , \tag{1.12}$$

where we defined the apparent binding constant $K_{app} = (K_d K_1)^{1/n}$. The total monomer concentration then is

$$r_0 = r_M + nr + nc = r_M + nr + n(x_0 - x) \; . \tag{1.13}$$

Inserting Equation (1.12) and $K_1 = \frac{r_M^n}{r}$ into Equation (1.13) yields $(2n)$th order polynomials for which exact solutions, if they exist, tend to be lengthy and hardly intuitive. We will therefore consider the limit cases in which one of the two reactions K_1 and K_d dominates.

If $K_d \ll K_1^{1/n}$ we can approximate the binding reaction by

$$\mathrm{X} + \mathrm{nR} \rightleftharpoons \mathrm{XR_n} \, . \tag{1.14}$$

As before, steady-state, conservation of mass, and approximating $r \approx r_0$ yields

$$\frac{x}{x_0} = \frac{1}{1 + \frac{r_0^n}{K_1}} \, . \tag{1.15}$$

This is the Hill function with Hill coefficient n, which is here equivalent to the number of TF monomers n. If $K_1^{1/n} \ll K_d$, most TF will be in the multimerized form and the reaction resembles a simple binding reaction with $n = 1$. In a reaction in which neither of the limit cases applies, we therefore expect to measure some intermediate n.

For repressors the regulatory term is proportional to the fraction of free promoter $f(c_{in}) \propto \frac{x}{x_0}$, while for activators it is proportional to the fraction of occupied promoter $f(c_{in}) \propto 1 - \frac{x}{x_0}$. So far, we have omitted that even at very high repressor or zero activator concentrations, RNA polymerase may be able to bind the promoter and transcribe a small amount of RNA, e.g. by pushing away TFs [142]. We can account for this so-called leaky expression with a constant α_0 that offsets the Hill curves. In summary, most transfer functions are phenomenologically well described by Hill functions of the form

$$f(c_{in}) = \alpha \frac{1}{1 + \frac{K^n}{c_{in}^n}} + \alpha_0 \, , \tag{1.16}$$

where $n > 1$ for activation and $n < -1$ for repression.

1.1.3 Nodes with Multiple Inputs and Outputs

GRNs commonly feature nodes with multiple inputs and outputs. Multiple outputs can be constructed relatively easily, e.g. by transcriptionally fusing output genes into one operon that is regulated by a single promoter, or by placing multiple output genes under separate promoters that are regulated by the same input (Figure 1.4a). The latter allows to fine-tune the transfer functions for each gene individually, so that e.g. the gene expression responds to different input levels. There is no strict limitation on the number of outputs other than the total number of genes that can be expressed simultaneously.

Regulation of one gene by multiple inputs, however, can be more complicated, as the chosen architecture determines how the inputs are being processed. For instance, we can either put different input promoters in front of copies of the same gene, or put multiple promoters in front of one copy of the gene (Figure 1.4b), to form a tandem promoter [143] (Figure 1.4c). Alternatively, multiple operator sites can be incorporated into a single promoter to construct combinatorial promoters [124, 144] (Figure 1.4d).

How inputs are being combined is often well described by boolean logic with respect to presence or absence of the input transcription factors. For instance, for tandem promoters (Figure 1.4c), two

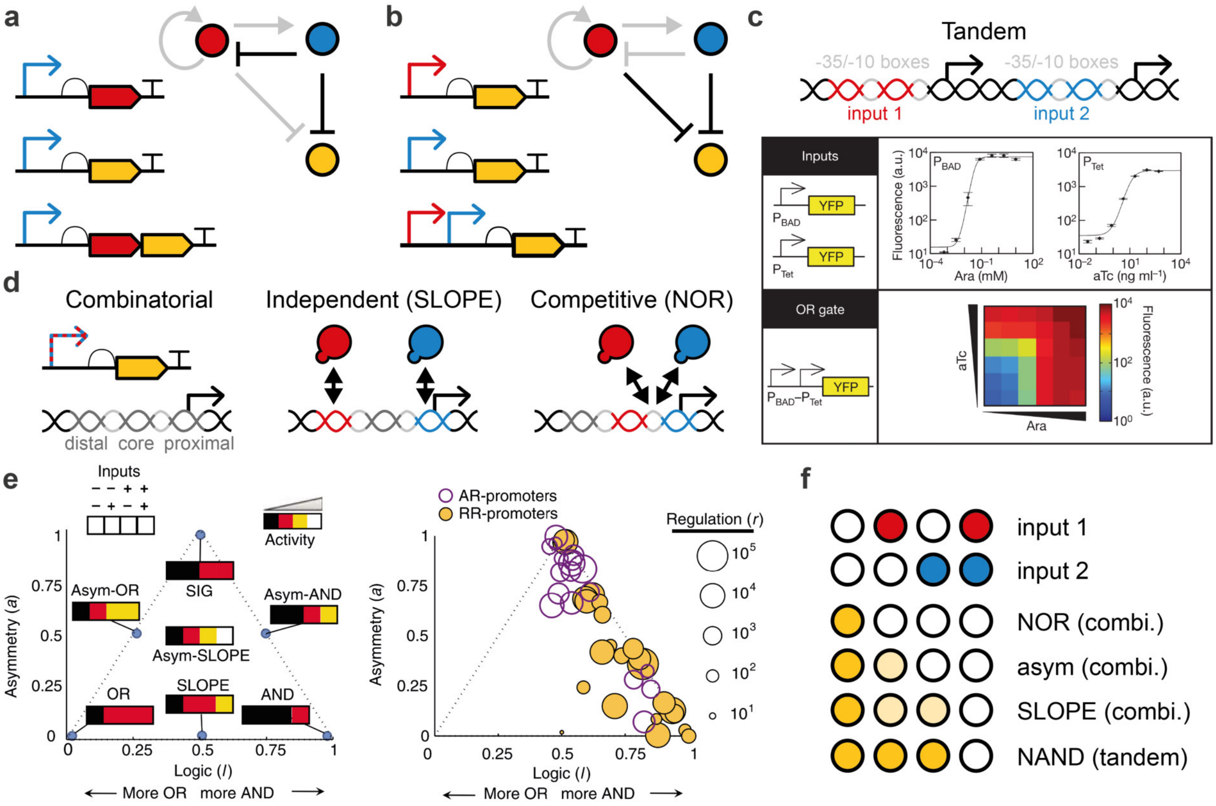


Figure 1.4: Multi-input promoter logic. While **a)** multiple outputs can be simply achieved by placing multiple genes under the control of the same promoter, **b)** there are different architectures to control one gene by multiple inputs that lead to distinct behaviors, depending on the interaction between the inputs. **c)** A tandem promoter consists of two individual promoters with the same direction that respond to a single TF as input. The transfer functions of two individual promoters combine with OR logic to yield the dual-input transfer function of the tandem promoter. Reproduced from reference [143] with permission of Springer Nature, © 2010. **d)** A combinatorial promoter is a single promoter that is regulated by multiple transcription factors. The positioning of the operators at three potential sites around the -10/-35 boxes (distal, core, proximal) affects the promoter logic. Closely placed operator sites tend to cause steric interference between TFs resulting a competitive interaction and more NOR-like logic, while distant operators often cause more SLOPE-like behavior, as defined in (e). **e)** Extended promoter logic scheme (left) and the logic behaviors of experimentally characterized combinatorial promoters (right) by Cox, Surette, and Elowitz [144]. AR=activator-repressor, RR=repressor-repressor, SIG=single-input-gate. Combinatorial promoters behave as AND-gates (NOR-gates if repressors are defined as inputs), with potential asymmetry if the operators are of different strength, but not as OR-gates (NAND-gates). Reproduced from reference [144] with permission of John Wiley and Sons, © 2007. **f)** Extended logic scheme as in (e), but defined with repressors as inputs. In (c,e) inputs are defined as the inducer molecules instead of transcription factors, which inverts the logic scheme compared to (f).

repressors tend to bind such that the presence of both is necessary to repress the gene (NAND logic). In contrast, for combinatorial promoters (Figure 1.4d,e), the presence of either repressor is often sufficient to sufficiently repress the gene (NOR logic). Note that we here have defined promoter logic with respect to the presence of active transcription factor (Figure 1.4f), whereas in the literature promoter logic is sometimes defined with respect to the presence inducer molecules. Hence, if the input is a repressor, the input state has to be negated to convert between the two conventions. Accordingly, what we define as NOR corresponds to AND, and NAND corresponds to OR.

However, real promoters will generally deviate from the idealized boolean logic. To account for this, Cox, Surette, and Elowitz [144] introduced an extended logic scheme for dual-input combinatorial promoters (Figure 1.4e). Besides the ideal NAND (OR) and NOR (AND) cases and the case in which one TF is dysfunctional and the promoter reduces to a single input gate (SIG), they consider two intermediate cases. If the two input TFs are of similar strength, but the presence of one alone does lead to intermediate expression levels, the logic is called SLOPE-like and the output is roughly equal to the SUM of the inputs. If the two inputs are of dissimilar strength, i.e. the presence of one leads to intermediate expression levels, whereas the presence of the other is sufficient to trigger full activation or repression, the logic is called 'asymmetric'. Interestingly, the authors found that combinatorial promoters predominantly show NOR-, SLOPE-, or asymmetric-NOR-like behavior [144] (Figure 1.4e). Also, many of the promoter variants tested were dysfunctional in the sense that they behaved similar to a SIG. The behavior of a combinatorial promoter is mainly determined by the relative position and strength of operator sites and the activity of the unregulated promoter. Repression is most effective at the 'core' position, followed by 'proximal' and 'distal', whereas activation is only effective at distal. If operator sites are distant, so that TFs can bind independently, promoters tend to behave SLOPE-like, whereas spatial proximity between operator sites produces more NOR-like (AND-like) logic [144]. The behavior can be fine tuned, by changing the promoter and operator strengths [145].

In some cases we are not interested in a digital response, but rather like to exploit the continuity of regulatory functions, for instance to analogically compute sums and logarithms [146]. To give a more detailed description of the input-output relation we can find a multi-dimensional version of the Hill-function, as shown for the 2-dimensional case in (Figure 1.4c). If the gene i is regulated by $j = 1, ..., N$ individual promoters, the N-dimensional regulatory function is simply the sum of the individual Hill-functions

$$f_i(c_1, ..., c_N) = \sum_j^N \frac{\alpha_{ij}(c_j/K_{ij})^{m_{ij}}}{1 + (c_j/K_{ij})^{n_{ij}}} \; . \tag{1.17}$$

Here, $m_{ij} = n_{ij} > 0$ for activators, $m_{ij} = 0$, $n_{ij} > 0$ for repressors. If the inputs act on the same promoter non-competitively, the N-dimensional regulatory function is the product of the individual Hill-functions [35]

$$f_i(c_1, ..., c_N) = \alpha_i \prod_j^N \frac{(c_j/K_{ij})^{m_{ij}}}{1 + (c_j/K_{ij})^{n_{ij}}} \; , \tag{1.18}$$

and if the inputs bind competitively we get [35]

$$f_i(c_1, ..., c_N) = \alpha_i \frac{\sum_j^N (c_j/K_{ij})^{m_{ij}}}{1 + \sum_j^N (c_j/K_{ij})^{n_{ij}}} \; . \tag{1.19}$$

Note, that regulatory functions may not be completely independent or competitive, which can be accounted for by using interaction parameters that are related to the logic parameter shown in Figure 1.4e [144] to generalize between the cases [140, 147]. Because the potential sites to place operators are limited, there is a practical limitation of ≈ 3 inputs that can be combined on one promoter. In the limit of large n_{ij}, m_{ij}, the N-dimensional Hill functions converge to the logic approximation introduced before, although the interpretation may depend on downstream thresholds. For instance Equation (1.17) with $N = 2$, $n_{ij} = \infty$, $m_{ij} = 0$ (repressors), $K_{ij} = 1$, and $\alpha_{ij} = 1$ corresponds to SLOPE-like behavior that will be interpreted as NAND, if the downstream threshold $K > 1$ and as NOR if $K < 1$.

1.1.4 Parameter Tuning

We have now established how a GRN with a desired topology can be constructed. The interactions between nodes can be described by Hill-type transfer functions that are characterized by a maximum expression rate α, dynamic range $\frac{\alpha+\alpha_0}{\alpha_0}$, a threshold K, and cooperativity n. Additionally, degradation of the output gene with a finite degradation rate δ is crucial for the function of most GRNs to avoid accumulation of reaction products. Assuming first-order degradation, the production rate of gene c_{out} is then given by

$$\dot{c}_{out} = f(c_{in}) - \delta c_{out} \ , \tag{1.20}$$

where $f(c_{in})$ is given by Equation (1.16). Solving this for $f(c_{in}) = \alpha$ gives

$$c_{out}(t) = \frac{\alpha}{\delta}\left(1 - \exp(-\delta t)\right) \ . \tag{1.21}$$

Hence, the degradation rate does not only affect the steady-state expression levels $c_{out,ss} = \frac{\alpha}{\delta}$, but importantly also defines the system time scale $\tau \propto \frac{1}{\delta}$.

To obtain the desired behavior, these parameters may need to be tuned, so that the output range of one node matches the input range of the downstream node [30] (Figure 1.5a). To this end, Brophy and Voigt [11] give an overview of common tuning mechanisms shown in Figure 1.5b. Typical parameter ranges that can be reached are summarized in Table 1.2.

The maximum expression rate α and dynamic range $\frac{\alpha+\alpha_0}{\alpha_0}$ can be changed by varying gene dosage, promoter strength [145], or RBS strength [131]. Degradation of reaction products can occur via different pathways, such as degradation by nucleases or proteases, consumption in a reaction, or dilution either by cellular growth, or in a cell-free reaction with continuous exchange of reagents

Table 1.2: Parameter ranges for CFPE reactions. The 'typical' range represents parameter values that can be expected for most commonly used genetic parts, whereas reaching the 'limit' range may require the use of optimized custom parts and dedicated tuning mechanisms. [a] In vivo.

parameter	typical	limit	ref.
K	$1 - 100$ nM	$0.1 - 10000$ nM	[89]
n	1-4	1-20	[28, 49, 89]
α	$0.1 - 1\ \mu\mathrm{M\,h}^{-1}$	$0.01 - 10\ \mu\mathrm{M\,h}^{-1}$	[63]
$\alpha_0/(\alpha+\alpha_0)$	0.01-0.2	0.0001-1	[111][a]
δ	$1/10\ \mathrm{min}^{-1} - 1/100\ \mathrm{min}^{-1}$	$1\ \mathrm{min}^{-1} - 1\ \mathrm{d}^{-1}$	[28, 89, 96]

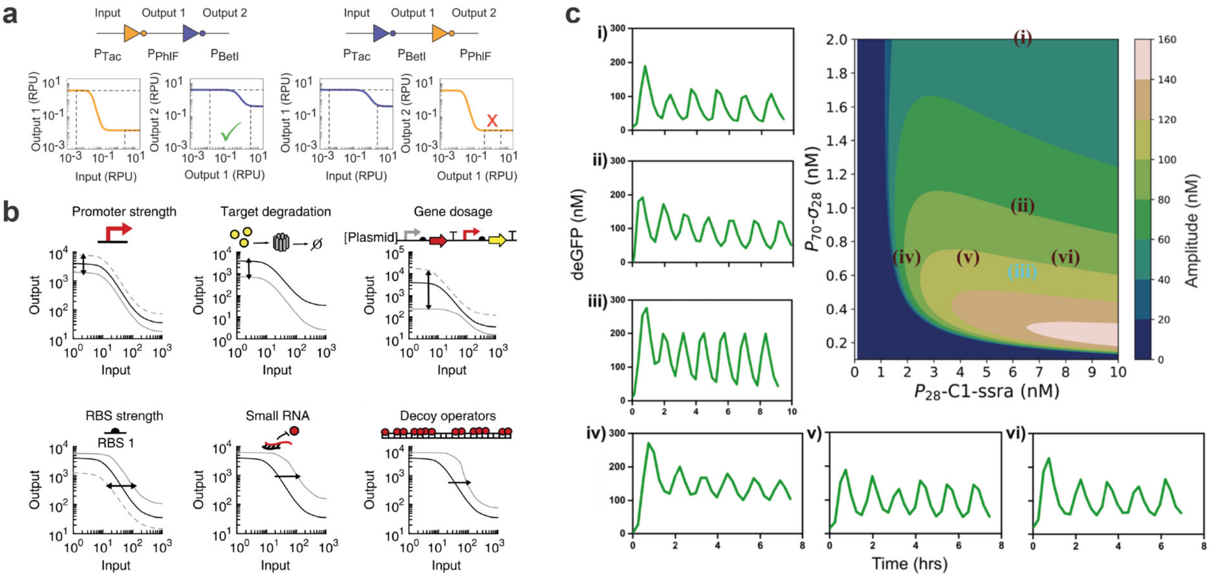


Figure 1.5: Parameter tuning. **a)** To match the output of one gene with the input of another, transfer functions have to match each others range. Reproduced from reference [30] with permission of the AAAS, © 2016. **b)** Various strategies to tune the apparent Hill parameters. Promoter strength changes α and α_0. Target degradation changes steady-state expression levels $c_{out,ss} \propto \frac{\alpha}{\delta}$. Gene dosage has similar effects to promoter strength. RBS strength changes the apparent K_d by affecting protein levels. Translational regulation by small RNA has a similar effect. Decoy operators increase the apparent K_d by reducing the effective protein concentration. Reproduced from reference [11] with permission of Springer Nature, © 2014. **c)** Example for a 2-dimensional phase space of a cell-free oscillator, varying the gene dosage of activator and inhibitor species. Reprinted from reference [90] under the license *CC-BY-NC-ND 4.0*.

(Section 1.2.5). An example of how gene dosage affects the amplitude of an oscillating GRN is shown in Figure 1.5c [90].

Thresholds K will generally have to roughly match the concentrations of the input $c_{in,ss} = \alpha_{in}/\delta_{in}$ (Figure 1.5a). Hence, $K\delta_{in}/\alpha_{in}$ will often be in the order of 1. Tuning can be achieved by mutating operator sequences, adding operators, or changing operator positions [144]. To achieve a dynamic behavior like oscillations it is often additionally required that the genes respond non-linearly and switch like (ultrasensitively), i.e. have a sufficiently large n [33]. This can for instance be achieved by choosing a promoter and TF with a high cooperativity or by buffering some amount of the circuit component with a stroichiometric binding reaction [148]. Buffering can be achieved using sequestering proteins [28, 149], decoy operators [45], or thresholds set by single stranded nucleic acids [49, 53].

When tuning parameters, we need to consider that generally parameters are interdependent, i.e. changing one parameter will affect other parameters as well. As an example, consider the following buffering mechanism, aimed to increase n, with buffer B

$$\mathrm{BR} + \mathrm{X} \underset{}{\overset{K_b}{\rightleftharpoons}} \mathrm{B} + \mathrm{R} + \mathrm{X} \underset{}{\overset{K_x}{\rightleftharpoons}} \mathrm{B} + \mathrm{RX}\ . \tag{1.22}$$

Here, the complex BR is inactive and the repressor R is hence not available for binding to the operator X. As before, we consider the condition that $x_0 \ll r_0$, which now yields $r_0 \approx r + r_b$, where r_b is the concentration of BR. If the binding to the buffer is strong, i.e. the stoichiometric binding parameter

$\frac{b_0}{K_b} > 1$ [148], R favors the binding to B. b_0 is the total concentration of the buffer, which here serves as the tuning parameter that can be experimentally varied. For $r_0 \ll b_0$, free $r = r_b \frac{K_b}{b_0} \approx r_0 \frac{K_b}{b_0}$, while for $r_0 > b_0$, $r \approx r_0 - b_0$.

This means that the transfer function of the non-buffered system is effectively shifted by b_0. We can now try to find an apparent transfer function that approximates the shifted transfer function

$$f_{shifted}(r_0) = \begin{cases} 0 & \text{if } r_0 < b_0 \\ \frac{K_x}{K_x + (r_0 - b_0)} & \text{if } r_0 \geq b_0 \end{cases} , \tag{1.23}$$

$$f_{app}(r_0) = \frac{K_{app}^{n_{app}}}{K_{app}^{n_{app}} + r_0^{n_{app}}} . \tag{1.24}$$

The apparent Hill coefficient $n_{app}(b_0)$ and threshold constant $K_{app}(b_0)$ are now functions of the tuning parameter b_0. Of course, the apparent threshold $K_{app}(b_0) \approx K_x + b_0$ is simply the original threshold K_x shifted by b_0.

One way to estimate n_{app} is to simply fit the shifted with the apparent Hill function [149]. However, to obtain a closed form approximation, we can use a local sensitivity measure of the transfer function, as defined in [148]

$$\zeta = \frac{d\ln f(I)}{d\ln I} = \frac{d\ln f(I)}{df(I)} \frac{df(I)}{dI} \frac{dI}{d\ln I} = \frac{I}{f(I)} \frac{df(I)}{dI} , \tag{1.25}$$

with input $I = r_0$ and output $f(I)$. ζ is also called logarithmic gain and represents the slope of the transfer function on a logarithmic scale. We can now for instance require that ζ at threshold is equal for the shifted and the apparent transfer function. Because by definition the factor $\left.\frac{I}{f(I)}\right|_{I=K_x+b_0} = 2(K_x + b_0)$ for both functions, we can simply equate the slopes at threshold

$$\left.\frac{K_x}{(K_x + r_0)^2}\right|_{r_0=K_x} = \left.\frac{n_{app} K_{app}^{n_{app}} r_0^{n_{app}-1}}{(K_{app}^{n_{app}} + r_0^{n_{app}})^2}\right|_{r_0=K_x+b_0=K_{app}} , \tag{1.26}$$

$$\frac{1}{4K_x} = \frac{n_{app}}{4K_{app}} , \tag{1.27}$$

In summary, we have found linear approximations for the apparent Hill parameters as a function of the tuning parameter

$$K_{app}(b_0) \approx K_x + b_0 , \tag{1.28}$$

$$n_{app}(b_0) \approx 1 + \frac{b_0}{K_x} . \tag{1.29}$$

Critically, this means that K_{app} and n_{app} can not be tuned independently with a buffering mechanism. Similar arguments can be made for other tuning mechanisms. For instance, if we increase x_0 to increase the maximum expression rate $\alpha_{app}(x_0)$, Equation (1.8) suggests that $K_{app}(x_0)$ would increase as well. Parametrization of apparent Hill parameters with tuning parameters is therefore a useful tool to estimate the experimentally accessible parameter space. The appropriate relations can be verified or determined experimentally. Sensitivity analysis and robust GRN design, as outlined in Section 1.3.4, can simplify parameter tuning and broaden the 'good' parameter range.

1.2 Communicating Artificial Cells

The contents of this section are based on the publication:

L. Aufinger and F. C. Simmel. “Establishing communication between artificial cells”. *Chemistry - A European Journal* 25.55 (2019), pp. 12659–12670. DOI: 10.1002/chem.201901726.

Broadly defined, artificial cells (ACs) are synthetic, cell-sized (1—100 μm) compartments that mimic some behavior of natural cells [16]. Here, we mainly consider ACs that are capable of gene expression via CFPE or are otherwise programmed to communicate with each other, with a particular emphasis on AC communities that are designed to exhibit reaction-diffusion phenomena. To this end, ACs have to be engineered to send out signaling molecules into the environment. The signaling molecules then spread by diffusion and can be sensed by receiver ACs, which then respond according to their genetic program.

The nature of the reaction compartments determines how and which molecules can be exchanged between them. We can use this for a rough categorization. Closed compartments typically are enclosed by a membrane which typically only permits exchange of small, non-polar molecules. Open compartments are typically membrane-free and therefore take up and release even large bio-macromolecules, keeping only reaction components that are specifically immobilized or attached to the AC through physicochemical interactions. In natural systems open compartments often occur as sub-compartments, e.g. in the form of droplets formed through liquid-liquid phase separation [163] or in the

Table 1.3: Examples for communicating ACs. The systems are roughly ordered from closed to open compartments. Active senders produce the signaling molecule *in situ*, whereas passive senders rely on a stock of signaling molecule that is slowly released. The release can be induced by expression of αHL pores. C: Number of signaling molecules employed in the longest reaction cascade. R: Reciprocal communication, i.e. the number of closed communication loops. RD: Reaction-diffusion phenomena. αHL: α-hemolysin dependent transport. m.p.: membrane permeable. o.p.: oil permeable.

Compartment	Reaction Environment	Signaling molecules	Sender type	C/R	RD	Ref.
Bacteria	Bacteria	3OC6-HSL	active	1/0	no	[9]
Vesicles	TXTL	Doxycycline, IPTG (αHL) Arabinose, Theophylline (m.p.)	induced release	2/0	no	[150]
Vesicles/ Eukaryotes	Eukaryotic TXTL/ Eukaryotic cell	Peptide (Cx43)	passive	1/0	no	[151]
Vesicles/ Bacteria	PURE/ Bacteria	Theophylline (m.p.), IPTG (αHL)	induced release	2/0	no	[152]
Vesicles/ Bacteria	Enzymatic/ Bacteria	Carbohydrate–borate complexes (m.p.)	active	1/0	no	[153]
Vesicles/ Bacteria	TXTL/ Bacteria	3OC6-HSL, C8 HSL, 3OC12 HSL, AI-2 (m.p.)	active	2/1	no	[26]
Vesicles/ Bacteria	PURE/ Bacteria	C4-HSL (m.p.)	active	1/0	no	[154]
Vesicles/ Bacteria	TXTL/ Bacteria	3OC6-HSL, Peptide (m.p.)	active	2/0	no	[155]
Vesicles/ Mouse	TXTL/ Cancer cells	Pseudomonas exotoxin (m.p.)	active	1/0	no	[156]
Vesicles/ Proteinosomes	PURE/ Enzymatic	3OC6-HSL (m.p.), Glucose (αHL)	induced release	2/0	no	[157]
Droplet interface bilayer	Genelets, TXTL, Bacteria	Arabinose, Guanine, Rhamnose (αHL), 3OC6-HSL, DFHBI, IPTG, ATc (m.p.)	induced release, passive	1/0	pulse, stochastic differentiation	[29]
Droplet interface bilayer	Enzymatic	Glucose (αHL), Hydrogen peroxide (m.p.)	active	2/0	no	[158]
Droplets	Bacteria	3OC6-HSL (o.p.)	active	2/0	cone	[15]
Droplets	TXTL/ Bacteria	3OC6-HSL (o.p.)	active	2/0	gradient	[25]
Droplets	inorganic	Br_2, BrO_2, $HBrO_2$ (o.p.)	active	1/0	stationary Turing patterns	[159]
Clay hydrogel in polymer shell	TXTL	T3 RNA polymerase, tetR repressor	active	2/0	quorum sensing	[125]
Hydrogel in droplet	PURE	mRNA	active	1/0	no	[37]
Hydrogel	PEN toolbox [160]	ssDNA	active	3/1	quorum sensing, travelling wave	[161]
Solid particles	PEN toolbox	ssDNA	active	0/0	French flag pattern	[46]
Proteinosomes	DNA TMSD	ssDNA	active	4/1	no	[162]
Chip	TXTL	σ^{28} , anti-σ^{28}	active, sustained	1/0	travelling wave	[24]
Chip	TXTL	σ^{28} , λ-repressor	active, sustained	1/0	coupled oscillations	[28]

early stages of insect oogenesis in which morphogens can freely diffuse between nuclei [40]. Because the transition from closed to open compartments is gradual, we can use measures such as molecular weight cut-off (MWCO) to roughly rank ACs. It should be noted, however, that this is no strict classification, as there can be more specific transport mechanisms for large molecules.

Another crucial aspect about the ‘openness’ of compartments is related to energy supply and waste management. In order to survive, all cells, natural or artificial, require energy produced by photosynthesis or sugar oxidation. In a broader sense we can also consider other nutrients and even high level reaction components such as enzymes or ribosomes as a form of energy currency, as to this date ACs lack the efficiency and metabolic pathways to regenerate these components from simpler sources [21, 164]. Most closed ACs are typically stocked with a finite amount of energy that is supplied with the reaction mix, resulting in a relatively short lifetime of ≈ 5 h. The lifetime of such bulk reactions has been prolonged to > 1 d in continuous microfluidic reactors [22, 23] and up to 11 d in hydrogel-based ACs supplied through microfluidics [165]. While these are open systems that also supply high grade reaction components, the reaction lifetime in vesicle-based ACs equipped with pores for supply of amino acids and ribonucleotides was extended to ≈ 4 d [20]. It is interesting to note that natural systems in a developmental context can show characteristics of both bulk and continuous reactions. While embryos in eggs use the energy stocked in the yolk, mammalian embryos rely on continuous supply by the mother. This suggests that both strategies may be promising approaches for the development of complex behavior in ACs.

In the following sections we first review the more closed compartments like vesicles, droplet interface bilayers, and droplets. Then we consider hydrogels, microfluidic reactors and other types of open compartments. In every section, we discuss molecular transport and diffusion properties specific to these compartments and highlight some examples for communication and reaction diffusion phenomena that were implemented with the specific system. An overview of the different systems is given in Table 1.3.

1.2.1 Vesicles

Vesicle-based artificial cells are aqueous compartments separated from an external aqueous solution by a lipid bilayer membrane. This membrane creates a micro-environment that contains essential bio-macromolecules like DNA, RNA, and transcription-translation machinery and thus can serve as a simplified model of a natural cell. An artificial cell can interact with the environment via the exchange of small membrane permeable molecules. The permeability P of a molecule through the hydrophobic bilayer tends to decrease with increasing size and increasing polarity, as illustrated in Figure 1.6a.

To enable the uptake and release of non-permeable molecules like sugars or ions, natural cells utilize a plethora of specific membrane proteins and pores. In the synthetic context, mainly α-hemolysin (αHL) pores from *S. aureus* have been used so far. With a minimum pore diameter of 1.4 nm [166] αHL pores enable the permeation of many small molecules like sugars and ions. For example, αHL pores were used to extend the lifetime of a TXTL reaction within artificial cells by supplying small energy molecules from the surrounding solution [20] or to enable and trigger pore-specific exchange of signaling molecules [150, 152] (Figure 1.6b,c).

The ability to integrate membrane proteins into the lipid bilayer membrane also allows for potential alternative routes towards an energetically autonomous artificial cell, e.g. by enabling light-induced ATP regeneration via H^+-gradients [167, 168].

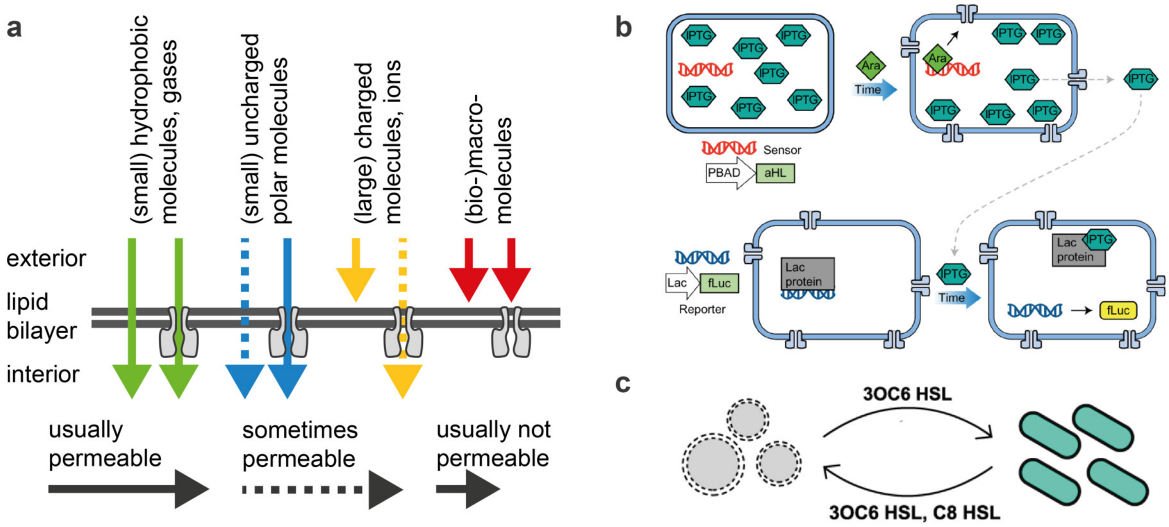


Figure 1.6: Vesicle-based ACs. **a)** Illustration of the selective permeability of lipid bilayers. While small hydrophobic molecules, such as homoserine lactone (HSL) molecules found in quorum sensing, are able to permeate unspecifically, the permeation of more polar molecules, such as glucose or isopropyl β-D-1-thiogalactopyranoside (IPTG), or ions can be pore specific. Large bio-macromolecules (> 2 kDa) are typically not permeable without the use of sophisticated import or export mechanisms found in natural membranes. **b)** Selective permeability was used to mediate communication between ACs. After sensing the presence of membrane permeable arabinose, a population of sender cells reacts by expressing αHL pores to trigger the release of IPTG. IPTG in turn is recognized by αHL containing receiver cells, which respond by producing firefly luciferase (fLuc), which is a reporter protein that generates a luminencent reporter signal. Reproduced from reference [150] with permission of Springer Nature, © 2016. **c)** Vesicle-based ACs were used to establish bi-directional communication with natural *E. coli* cells. The communication is mediated via membrane permeable HSL molecules. Reproduced from reference [26] with permission of the ACS, to whom requests related to further permission should be directed, © 2017.

1.2.2 Droplet Interface Bilayer Networks

Instead of forming vesicles, lipids can also be used to create water-in-oil emulsion droplets. Under appropriate conditions discussed below, the lipids between two droplets that are pushed close together form a lipid bilayer, commonly called a droplet interface bilayer (DIB) [171]. Many droplets can be combined in this manner to form a DIB network, which is also referred to as a synthetic tissue [172] (Figure 1.7a).

As DIBs also allow the integration of membrane proteins, it is possible to create pore-specific pathways for signaling molecules and ions [29, 169, 173] (Figure 1.7b,c) and to create artificial tissues that differentiate based on diffusible inducer molecules [29, 38] (Figure 1.7d,e) and morph based on osmosis [169] (Figure 1.7a). The DIB technique can be automated via a 3D printing process [169] to form large (mm sized), close-packed droplet lattices with a resolution of a single droplet (≈ 100 µm) [170]. The combination of 3D printing technology with genetically encoded differentiation and morphogenesis paves the road for novel synthetic hybrid materials that are manufactured in a combined top-down/ bottom-up approach [38].

To this end it is important to understand the physics underlying the formation and geometry of DIB networks, as well as molecular transport through them. A DIB between two droplets is characterized by the equilibrium contact angle θ_b (Figure 1.7f), given by

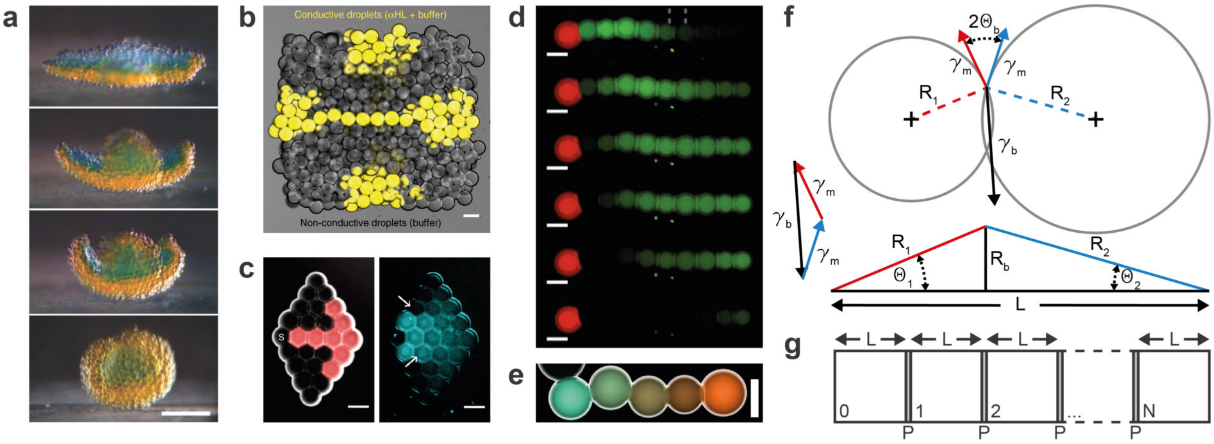


Figure 1.7: DIBs for artificial tissues. a) 3D printed synthetic tissue that morphs from a flower into a ball shape within 8 h using osmotic pressure. Scale bar, 200 µm. Reproduced from reference [169] with permission of the AAAS, © 2013. **b)** Conductive ion pathway through a DIB network using droplets containing αHL pores that were 3D printed at single droplet resolution. Scale bar, 100 µm. Reproduced from reference [170] under the license *CC BY 4.0*. **c)** Example of a diffusion pathway through an artificial tissue. Droplets labeled in red contain αHL pores, which allow the diffusion of a pore specific signaling molecule from sender droplet (S) into neighboring droplets. Arrows point at two droplets that are or are not connected to the diffusion pathway. Scale bars, 200 µm. **d)** Diffusive pulse in a linear arrangement of DIBs. The signaling molecule (DFHBI) diffuses from the sender droplet (red) into receiver droplets where it triggers an IVT circuit, resulting in a pulse wave. The total time scale is 3 h. Scale bars, 200 µm. (c,d) Reproduced from reference [29] with permission of Springer Nature, © 2018. **e)** Linear arrangement of DIBs that differentiates according to a cell-free genetic program that interprets a diffusive morphogen gradient induced from the left. Scale bar, 100 µm. **f)** Definition of parameters that describe the geometry of DIBs. In equilibrium, surface tensions γ_b and γ_m are balanced, so that θ_b can be calculated from distance measures. **g)** A linear arrangement of DIBs can be described as a series of compartments of length L, separated by a lipid bilayer with permeability P.

$$\cos(\theta_b) = \frac{\gamma_b}{2\gamma_m} \,, \tag{1.30}$$

which follows from an equilibrium of bilayer and monolayer surface tension γ_b and γ_m. Following Dixit et al. [171], the ratio of surface tensions $\frac{\gamma_b}{\gamma_m}$ can be geometrically related to the droplet radii R_1 and R_2 and their center-to-center distance L, which can be accurately measured from microscopy images. θ_b and the bilayer radius R_b can then be calculated from

$$2\theta_b = \cos^{-1}\left(\frac{L^2 - R_1^2 - R_2^2}{2R_1R_2}\right) \,, \tag{1.31}$$

and

$$R_b = \frac{R_1 R_2 \sin(\theta_b)}{L} \,. \tag{1.32}$$

In practice, the stability of the bilayer can be achieved by introducing a 'bad solvent' for lipids in the continuous phase (e.g. silicone oil in hexadecane) so that bilayer formation becomes thermodynamically favorable [174]. Careful tuning of the bad solvent fraction and lipid composition can be used to tune γ_b, which is a key parameter to control the packing of DIB networks [170].

To obtain uniform packing it is further relevant to consider the kinetics of bilayer formation, which can be biphasic [170]. After a rapid phase of initial contact (≈ 1 s), the contact angle reaches a value $\theta < \theta_b$. In the consecutive slow phase (≈ 10 min) the organic phase drains between the droplets [175, 176] and lipids arrange to form a stable bilayer approaching θ_b. Long term (≈ 1 h) morphology of DIB networks can be further influenced by water permeation due to osmotic pressure [169, 174], partitioning of water molecules into the organic phase [29], or adsorption of reagents or reaction products from the aqueous phase on the surfactant [29].

Following Dupin and Simmel [29] to describe the diffusion of a permeable solute through a DIB network, we first consider a quasi-1-dimensional array of identical droplets of length L, volume V and diffusion coefficient D that are separated by membranes with permeability P, area $A \approx \pi R_b^2$ and length $l \ll L$ (Figure 1.7g). Using Fick's law, the diffusive flux of a solute with concentration $c = \frac{n}{V}$ across a droplet j_d or bilayer j_b is given by

$$j_d = D \cdot \frac{\Delta c}{L} \,, \tag{1.33}$$

$$\text{and} \quad j_b = P \cdot \Delta c \,, \tag{1.34}$$

respectively. The total flux through the array j is then given by the series connection

$$j^{-1} = j_d^{-1} + j_b^{-1} = \left(\frac{\Delta c}{L}\right)^{-1} \cdot \left(D_d^{-1} + (P \cdot L)^{-1}\right) \,, \tag{1.35}$$

where we expanded j_b with L. We can then define an apparent diffusion coefficient [177]

$$\frac{1}{D_{app}} = \frac{1}{D} + \frac{1}{PL} \,. \tag{1.36}$$

For $P \ll \frac{D}{L}$ the diffusion is bilayer limited and $D_{app} \approx PL$, while for $P \gg \frac{D}{L}$ the diffusion through the droplet is limiting and $D_{app} \approx D$.

Finally, with $j = \frac{1}{A}\frac{dn}{dt}$ and approximating the cross section for diffusion with the bilayer area, the exchange of molecules in arbitrary droplet networks with varying volumes V_i, lengths L_i and bilayer areas A_{ij} is given by [29]

$$\frac{dn_i}{dt} = \sum_{j \in \{\text{neighbors of } i\}} \frac{A_{ij}}{V_i} \frac{DP}{PL_i + D} (n_j - n_i) \,. \tag{1.37}$$

1.2.3 Water-in-oil Emulsion Droplets

Water-in-oil emulsion droplets are aqueous compartments that are separated from a non-polar solvent by a single surfactant layer. Monodisperse emulsions with a defined droplet size can be generated by droplet microfluidic techniques [183]. Among a large range of oil-surfactant systems [176],

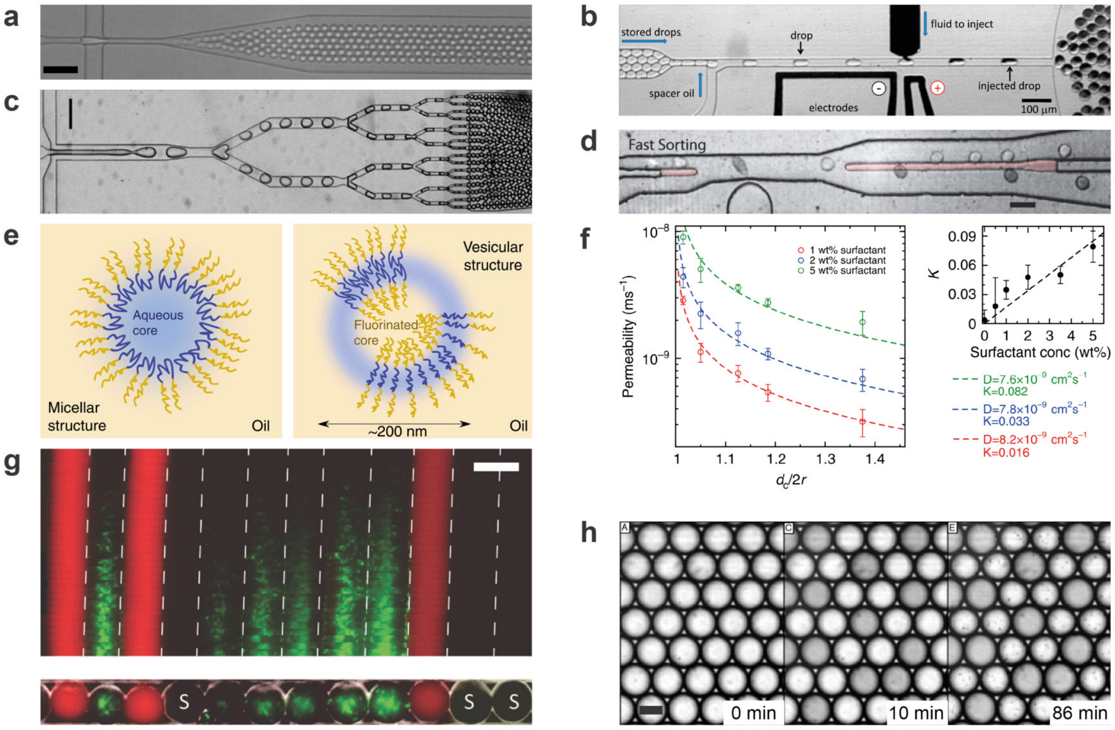


Figure 1.8: Emulsion droplets. a)-d) Examples of microfluidic droplet manipulation techniques that could be used to simulate cell cycles. **a)** Simple flow focusing device for production of monodisperse emulsion droplets. Scale bar, 100 µm. Reproduced from reference [178] with permission of The Royal Society of Chemistry, © 2008. **b)** Picoinjection allows addition of controlled volumes of liquid to existing droplets, imitating growth. Reproduced from reference [179] under the *PNAS license*, © 2010 the authors. **c)** Droplet splitting could mimic cell division. Scale bar, 200 µm. Reproduced from reference [180] with permission of The Royal Society of Chemistry, © 2011. **d)** Droplet sorting can be used for unnatural selection of certain phenotypes for directed evolution. Scale bar, 50 µm. Reproduced from reference [181] with permission of The Royal Society of Chemistry, © 2015. **e,f)** Molecular transport between fluorocarbon emulsion droplets is largely surfactant mediated. Exchange of molecules likely occurs through the aqueous phase of micelle-like surfactant assemblies. (e,f) Reproduced from reference [182] under the license *CC BY 4.0*. **g)** Linear arrangements of emulsion droplets were used to establish communication between artificial cells and *E. coli* cells, and vice versa. Scale bar, 50 µm. Reproduced from reference [25] with permission of the Oxford University Press, © 2016. **h)** Droplets were used to study Turing's thesis of morphogenesis using the inorganic Belusov-Zhabotinsky reaction. Here, the reaction first establishes a pattern. Then, the assembly morphes due to osmotic shrinking of the brighter droplet population. Scale bar, 60 µm. Reproduced from reference [159] under the *PNAS license*, © 2014 the authors.

fluoro-blockcopolymer surfactants in combination with fluorocarbon oil have emerged as the gold standard in the droplet microfluidic field, due to their chemical inertness, biocompatibility, omniphobicity, and mechanical stability provided by the large size of the surfactant ($15-30$ nm) [178]. These features make droplets very stable, but also render them intrinsically more isolated, complicating interaction with the environment and nutrient supply. However, droplets have been used as templates for vesicle based AC generation [184–186], but also as an AC system themselves [25].

In this context, the rich set of available microfluidic droplet manipulation techniques (Figure 1.8a-d) allows mimicking of cellular processes such as growth [179], division [180], or even unnatural selection [181], bypassing some of the core challenges in the artificial cell field. While such artificial cells promise less autonomy than vesicle based ones, it is possible that similar external processes played a key role in the evolution of early protocells [17].

Molecular transport in fluorocarbon emulsions is largely mediated by micelle-like surfactant assemblies, which contain an aqueous phase into which the solute molecules can partition [182, 187] (Figure 1.8e,f). Despite differences in mechanism, diffusion through closely packed droplet assemblies can be mathematically described by the same equations as diffusion through DIB networks (Equation (1.36) and (1.37)), with the membrane being replaced by an oil barrier with permeability P, which depends on geometry [159]. Exchange of small hydrophobic molecules through the oil phase was for example used to mediate communication between communities of natural cells [15], between ACs and natural cells [25] (Figure 1.8g), or to study oscillatory coupling and pattern formation in droplet lattices filled with inorganic Belusov-Zhabotinsky reagents [159, 188–191] (Figure 1.8h).

1.2.4 Hydrogels

Hydrogels are crosslinked polymer networks in an aqueous solvent. Because they exhibit elastic properties, but also allow the diffusion of solutes, without a membrane that separates them from an external aqueous solution, they are often described as solid-liquid hybrid materials. From a materials perspective reviewed, e.g., by Thiele et al. [192], hydrogels mostly fall into two main categories: hydrogels of natural origin often consist of polysaccharides, whereas synthetic ones are often based on crosslinked vinyl-, acryl- or epoxide-networks. Polysaccharides are often stiffer as the distance between flexible glycoside bonds is typically 4 carbon bonds, whereas the artificial polymers are flexible around every carbon bond. Further gel material properties are charge, hydrophilicity, swelling properties, polymer volume ratio, the crosslinking mechanism, and biocompatibility. Monodisperse microgels can be synthesized using microfluidic techniques [193–195].

In the context of ACs, hydrogels have been utilized as open compartments harboring DNA resembling artificial nuclei [37, 125, 196]. The absence of membranes allows for communication with biomacromolecules such as mRNA [37] (Figure 1.9a,b) or proteins [125] (Figure 1.9c,d), which is a promising approach to enable the design of interactions in a programmable manner.

How well a solute can enter a hydrogel is generally described by the equilibrium partitioning coefficient K defined as

$$K = \frac{[S]_{gel}}{[S]_{sol}} = \exp\left(-\frac{\Delta G_{gel} - \Delta G_{sol}}{k_B T}\right) , \tag{1.38}$$

where $[S]_{gel}$, $[S]_{sol}$, ΔG_{gel}, and ΔG_{sol} are the solute's equilibrium concentrations and Gibbs free energies in the gel and free solution, respectively. ΔG_{gel} can be influenced by steric, electrostatic or chemical interactions, and hence the total partitioning coefficient can be expressed as a product of these contributions $K = K_{steric} \cdot K_{electrostatic} \cdot K_{chemical}$ [197]. For a description of electrostatic contributions refer to [197].

The steric contribution is mainly entropic and can be described by the gels mesh size ξ, defined as the size of the solute for which $K_{steric} = 1/2$. The gel grid reduces the number of states a solute can obtain inside the gel compared to the free solution. The solutes entropy inside the gel will therefore be lower and ΔG_{gel} higher. How strongly the mesh size will affect the partitioning of a given solute in turn also depends on the stiffness of the gel fibers and the size, geometry, and stiffness of the solute

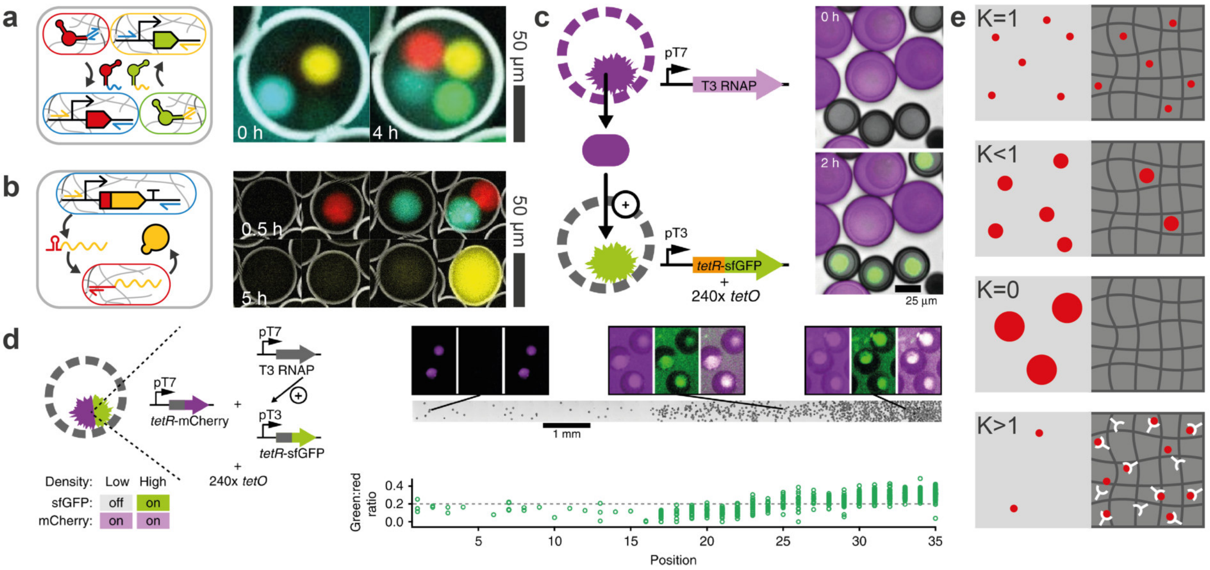


Figure 1.9: Hydrogel microparticles. **a)** Example for selective partitioning by immobilizing ssDNA ligands for a fluorescent RNA reporter transcribed from hydrogels. **b)** The method from (a) was used to spatially separate a transcription from a translation reaction. **c)** Selective partitioning with a fluorescently tagged repressor protein binding to immobilized operator sites. **d)** The method from (c) was used to generate quorum sensing like behavior in a density gradient of ACs. (c,d) Reproduced from reference [125] under the license *CC BY 4.0*. **e)** Partitioning of molecules in a hydrogel-solution two phase system is determined by hydrogel properties such as mesh size and intermolecular forces between the solute and hydrogel matrix. At small polymer volume fractions < 5% and in the absence of other interactions, molecules much smaller than the mesh size do not 'see' the hydrogel, while molecules much larger than the mesh size can not enter the hydrogel. Molecules with sizes in the order of the mesh size can enter the gel, but $K < 1$. K can be specifically increased through the presence of ligands immobilized in the hydrogel matrix.

(Figure 1.9e). Intuitively, a small globular solute with $r_g \ll \xi$ will not 'see' much of the gel and if the polymer volume fraction ϕ of the gel is small ($\phi \ll 1$), $K \approx 1$. In contrast, large globular solutes $r_g \gg \xi$ simply do not fit into the gel and $K \approx 0$. While flexible linear molecules like dextranes, which are often used to estimate mesh sizes, or RNA can be well approximated by a globular solute, more stiff linear polymers such as dsDNA may behave differently and 'reptate' [198].

Chemical interactions of the solute with the hydrogel matrix are usually simple association/ dissociation reactions between the solute S and an immobilized ligand L with dissociation constant $K_d = \frac{[S]_{gel,free}[L]}{[SL]}$. In absence of other interactions the concentration of unbound solute in the gel $[S]_{gel} - [SL] = [S]_{gel,free} = [S]_{sol}$ yielding [197]

$$K_{chemical} = \frac{[S]_{gel}}{[S]_{sol}} = 1 + \frac{[L]}{K_d} \,. \tag{1.39}$$

If $[L] \gg [S]$ we can approximate $[L] = [L]_0 - [SL] \approx [L]_0$.

Hence, a way to engineer specific partitioning of certain molecules is to introduce binding sites to the gel network by chemically modifying the polymer with appropriate ligands. For instance, we have shown that we can immobilize ssDNA oligomers to an agarose hydrogel to specifically localize (m)RNA molecules [37]. In this context, it is advantageous to use a neutrally charged gel (at physiological pH) with a large mesh size, such as agarose, to minimize unspecific contributions to ΔG_{gel}.

Diffusion in gels can be complicated and depends on the same properties as the partitioning coefficient, so we usually have to measure the apparent diffusion coefficient D_{gel} experimentally. In the case of a globular solute and exclusively steric contributions to ΔG_{gel}, however, the ratio of D_{gel} to the free diffusion coefficient D_0 approximately equals K [199]

$$D_{gel} \approx D_0 \cdot K \ . \tag{1.40}$$

In a gel with immobile ligands with K_d, only the unbound fraction of solute can diffuse which gives [200, 201]

$$D_{gel} \approx D_0 \frac{1}{1+[L]/K_d} \ . \tag{1.41}$$

1.2.5 Microfluidic Reactors

Microfluidics can not only be used to generate and manipulate ACs, but microfluidic reactors themselves can be powerful engines to simulate cellular behaviors. The striking feature of microfluidic reactors is their intrinsic capability to continuously supply reactions with nutrients, in analogy to e.g. microfluidic blood capillaries in higher organisms. The ability to keep reactions within ACs out of equilibrium for extended periods of time is a central requirement for the study of more complex dynamic behaviors under steady-state conditions.

Advanced microfluidic devices come with a high degree of active control and automation [23], enable the control of communication channels via diffusion [22], and enable screening of genetic components such as promoter libraries at a throughput of up to 768 TXTL reactions on a single device [92]. While most microfluidic devices are manufactured from PDMS, the incorporation of hydrogels [203] or even construction of gel-only chips [204, 205] may enable the supply of spatially extended and differentiated synthetic tissues with nutrients. In the following, we first introduce diffusion-based and automated, valve-based microfluidic reactors and then discuss the perspectives of microlfuidic gel reactors in the context of pattern formation.

Diffusion-based continuous microfluidic reactors consist of reaction chambers containing patches of genes immobilized to their surface. Such 'DNA-brushes' can be generated using lithographic patterning [206, 207] or DNA microarray spotting techniques [92], which allows to equip multiple reaction chambers with different genes. The reactions are kept in a steady state by passive diffusive exchange of fresh CFPE mix and reaction products through a narrow capillary that connects the reaction chambers with a supply channel [22] (Figure 1.10a). In a similar manner, individual chambers can be directly connected to allow for communication between neighboring cells [24].

This technique was used to implement oscillatory networks [22], a propagating wave [24] (Figure 1.10b), study the synchronization between coupled oscillators [28] (Figure 1.10c), and stochastic switching of a bistable GRN [208]. In this context it is important that key parameters such as expression and dilution rates, or diffusive coupling between cells, can be tuned by varying the geometry of the device.

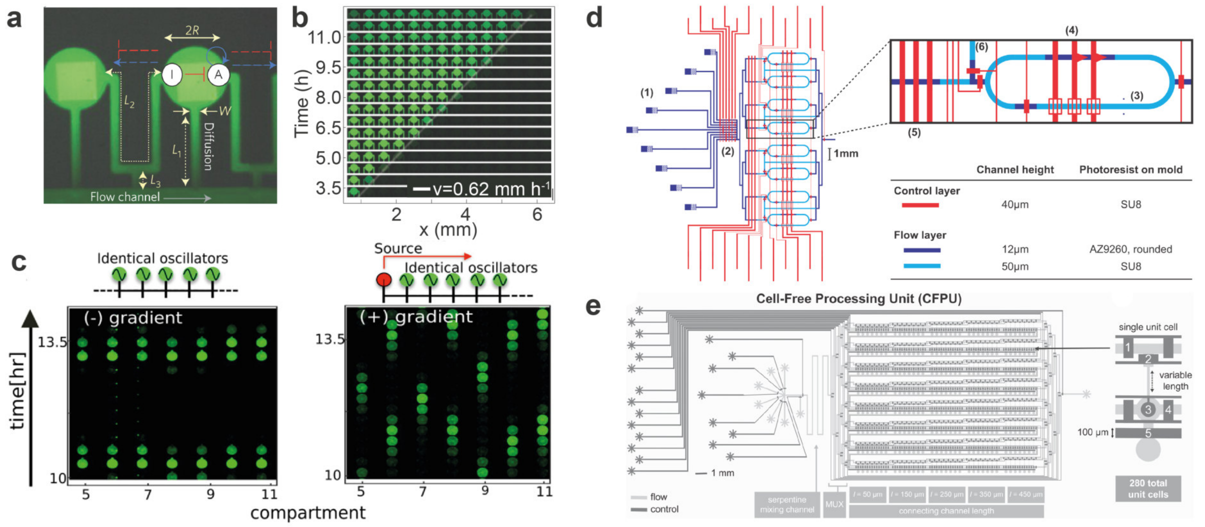


Figure 1.10: Microfluidic reactors as artificial cells. **a)** On-chip DNA compartments (circles) are continuously fed with fresh TXTL reagents through a capillary (L_1, W). The DNA is immobilized in the form of a DNA brush (green square). Artificial cells communicate through capillaries L_2, whereby the ratio L_2/L_3 can be varied to tune coupling strength. **b)** Propagating gene expression wave in an array of coupled on-chip ACs. (a,b) Adapted from reference [24] with permission of Springer Nature, © 2015. **c)** The same type of ACs was used to study synchronization and pattern formation in arrays of coupled ACs. Reproduced from reference [28] under the *PNAS license*, © 2017 the authors. **d)** Quake valve based ring reactor. The chip consists of flow channels (blue) and control lines (red), which when pressurized close valves to block the flow. Reagents are supplied through inlets (1) and selected by a multiplexer (2). At the heart of the chip are 8 rings (3) that serve as the reaction compartments. The reactions within the rings are stirred and fed using peristaltic pumps (4,5). To change between reagents, reagents are flushed through bypass channels (6). In a typical experiment, the ring is periodically fed with a defined amount of fresh reagent, the reaction is mixed and images are acquired, as defined by an automated program. Reproduced from reference [23] under the PNAS open access option. **e)** The 'cell-free processing unit' allows screening of TXTL reactions at a high throughput of 280 parallel reactions. The 35-fold increase in capacity compared to the 8-ring reactor was achieved by switching to passive dilution via diffusion through small capillaries (as in (a)), pulse-width modulated reagent mixing, and a DNA microarray. Reproduced from reference [202] under the license *CC BY 4.0*.

If the diffusive dilution through the capillary is slow compared to the protein expression, the system can be approximated as a constant source (the reaction chamber) and sink (the supply channel). Then, a linear gradient will establish over the length L of the capillary in equilibrium (Figure 1.10a). With a channel height h, a reaction chamber volume $V_R = A_{chamber} \cdot h$, and capillary volume $V_C = L \cdot w \cdot h$, the protein lifetime in the chamber is given by [22]

$$\tau_d = \frac{V_R}{V_C}\frac{L^2}{D} = \frac{A_{chamber}}{Dw}L \ . \tag{1.42}$$

With a similar time scale τ_c defined for diffusive coupling between neighboring chambers, communication can be described by [22]

$$\frac{dc_i}{dt} = \frac{c_{i-1} - 2c_i + c_{i+1}}{\tau_c} - \frac{c_i}{\tau_d} \ , \tag{1.43}$$

which resembles a discrete diffusion equation with degradation. Note, that one geometrical constraint is that convective flow through the reaction chambers has to be limited. In practice this can be achieved by choosing the height of capillaries and reaction chambers h, much smaller than the height of the supply channel, as the flow resistance for rectangular channels scales as $R \propto h^{-3}$. Since h canceled in Equation (1.42) this does not affect the diffusive timescales.

Another powerful platform for prototyping and characterization of dynamic GRNs are valve-based reactors, as developed by the Maerkl lab [23, 202] (Figure 1.10d,e). The PDMS devices consist of two layers, a flow layer to flush reagents and a control layer used to direct the fluid flow on-chip by actuating so-called ‘Quake valves’ and peristaltic pumps [209]. These devices can run fully automated custom programs enabled by software controlled valve actuation and image acquisition. For instance, the so-called 8-ring device, described in detail in [210, 211] (Figure 1.10d), consists of 8 individually addressable rings in which the reaction takes place. The rings can be periodically fed with up to 9 different fluids and then mixed by pumping the solution in circles around the ring. The fraction of the ring volume that is replaced in each cycle is called the refresh ratio $0 < R < 1$. With the time interval between dilution cycles t_{int}, the decay constant is

$$\tau_d = -\frac{t_{int}}{\log(1-R)} \ . \tag{1.44}$$

This dilution allows to keep reactions in a steady state for extended periods of time (> 24 h), which enables implementation of dynamic systems such as cell-free oscillators [23, 89, 90]. Notably, the periodic turnover by dilution simulates the dilution by cell growth, which allowed the transfer of oscillators prototyped in an 8-ring reactor to living cells [89]. Instead of feeding all the reaction components, some proteins of the transcription-translation machinery can also be regenerated by co-expressing them in the reactors themselves, which can be a starting point towards self-regenerating ACs [21]. Other valve-based devices such as the MITOMI platform [212] use gene libraries on a DNA microarray to enable the characterization of cell-free gene GRNs at a high throughput of up to 768 batch reactions [92], or 280 continuous reactions when combined with diffusive feeding through a capillary [202] (Figure 1.10e). These examples highlight the potential of microfluidic reactors for the prototyping of ACs and cell-free GRNs.

Another type of continuous flow reactors with potential applications in pattern formation are (micro)fluidic gel reactors. As described in Section 1.2.4, hydrogels are an attractive material for the study of pattern formation because it allows for the diffusion and supply of reagents, but prevents convection that would interfere with reaction-diffusion phenomena. This fact was exploited in early reaction-diffusion experiments to create spiral waves [213–215] and led to the first experimental evidence of Turing patterns [216–218]. For the realization of Turing patterns, the gel played another crucial role. To visualize the pattern, an immobile starch indicator or gel consisting of poly(vinyl alcohol) was used. As the color change is a result of a complex formation between the indicator and the activator in the reaction, the presence of indicator results in a decrease of the apparent diffusion coefficient of the activator which is a necessary condition for Turing pattern formation.

The application of hydrogels in a microfluidics context has been reviewed by Goy, Chaile, and Madrid [203]. The main application areas are cell culture [219–221], tissue engineering [222], and biosensing. Even gel-only chips with potential interest in tissue engineering have been realized [204, 205]. Also, hydrogels have been frequently used to generate gradients within a microfluidic device [219, 223], for instance for cell mobility and differentiation studies [224], or DNA based TMSD circuits [225].

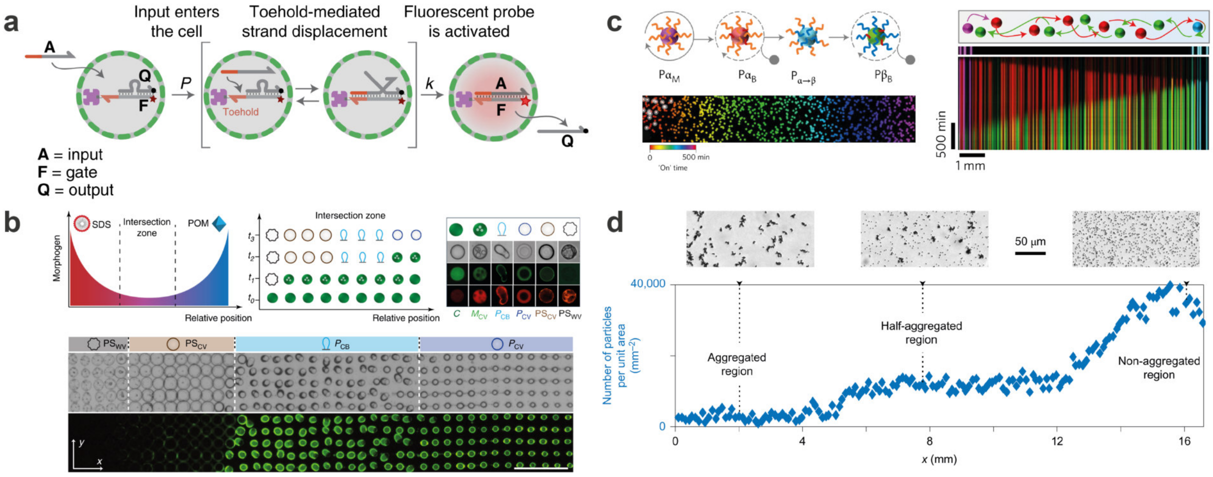


Figure 1.11: Other types of ACs. **a)** Porous polymersomes communicating via ssDNA signals. Reproduced from reference [162] with permission of Springer Nature, © 2019. **b)** Coacervate based ACs react to two opposing gradients of an additive by changing their morphology in a spatially differentiated manner. Scale bar 500 µm. Reproduced from reference [226] under the license *CC BY 4.0*. **c)** DNA-coated colloidal particles sending a DNA triggered wave through a chamber and back. Reproduced from reference [161] with permission of Springer Nature, © 2017. **d)** DNA-coated colloidal particles form a French flag pattern with differentiated levels of aggregation as programmed by a DNA circuit based on the PEN toolbox [160]. Reproduced from reference [46] with permission of Springer Nature, © 2017.

1.2.6 Others

Additional systems used to form ACs are coacervates [226–229], polymersomes [162, 230, 231], colloidosomes [232], proteinosomes [157, 233], or solid particles [46, 161].

Polymersomes, for instance, are shells surrounded by a porous polymer membrane. The MWCO can be tuned such that communication via short ssDNA is possible whereas the functional identity is maintained by DNA immobilized to larger, non-permeable streptavidin complexes [162, 231] (Figure 1.11a).

Coacervates are membrane-free droplets that are formed by liquid-liquid phase separation (LLPS) and contain high concentrations of polymers. Coacervates not only play an important role as protocells in origin of life hypotheses [234], but LLPS generally is an important concept in modern biology [163]. The morphology of some coacervate based ACs can change in response to additives, which was used to mimic spatial differentiation and morphogenesis based on physicochemical interactions [226] (Figure 1.11b).

ACs can also exist of solid colloidal particles that are coated with DNA. For instance, such particles were used to realize propagating waves [161] (Figure 1.11c) and to synthesize a French flag pattern [46] (Figure 1.11d).

1.3 Spatial Differentiation and Pattern Formation

To provide a framework to engineer synthetic patterns, we can look into patterning concepts in natural systems. Biological development is the process by which single stem cells evolve into complex multicellular organisms. Development typically involves morphogenesis (the evolution of mechanical force fields), spatial differentiation (chemical pattern formation), and molecular differentiation (the generation of specialized cell types from pluripotent stem cells). While these three processes are often interdependent [235], we here focus on the spatial differentiation by GRNs in the context of noise.

Generally, we can describe a pattern-forming synthetic gene regulatory network with a partial differential equation of the form

$$\frac{\partial c_i(\boldsymbol{x},t)}{\partial t} = D_i \Delta c_i(\boldsymbol{x},t) + f_i(\{c_j(\boldsymbol{x},t)\},\{k_{ij}\}) \, , \tag{1.45}$$

where $c_i(\boldsymbol{x},t)$ $(i = 1,...,N)$ are the molecular species that form a GRN with N nodes. The regulatory interactions are governed by the functions $f_i(\{c_j(\boldsymbol{x},t)\},\{k_{ij}\})$, which are described by regulatory parameters $\{k_{ij}\} = K_{ij}, n_{ij}, \alpha_{ij}, \alpha_{0,i}$, as outlined in Section 1.1. The diffusion of the molecular species through the system is described by the diffusion coefficients D_i, as covered in Section 1.2.

We first present common mechanisms and concepts that are used to describe pattern formation in biological systems (Section 1.3.1). We next present how the precision of a resulting pattern in the presence of noise can be quantified using measures such as positional information (Section 1.3.2). In Section 1.3.3 we present design approaches to determine the topology of a pattern forming GRN. Finally, we propose a strategy to find parameter ranges that lead to robust pattern formation using evolutionary algorithms (Section 1.3.4).

1.3.1 Patterning Mechanisms

Chemical patterning in biology typically involves pattern inducing molecules, called morphogens, that are regulated and interpreted by a GRN through various mechanisms. The two major concepts are positional information (PI) and reaction diffusion (RD) [238]. PI was first proposed in 1969 by Lewis Wolpert as a universal concept for pattern formation [239]. The PI is provided by a morphogen gradient and interpreted by the cells via a GRN that measures the local morphogen concentrations using appropriate thresholds. The central argument propagated by Wolpert is the so-called 'French flag problem' motivated by the scale invariance observed in patterning systems (Figure 1.12a). RD mechanisms, as first proposed by Alan Turing in 1952 [237], allow the generation of a periodic pattern from an initially homogeneous state (Figure 1.12b). RD ideas were later refined by Gierer and Meinhardt, who studied a simple system consisting of a short range, autocatalytic activator and a long range inhibitor [240].

While PI can explain the differentiation into regions with different and arbitrary phenotypes, it fails to explain how the initial gradient or 'prepattern' is generated in the first place. In contrast, RD explains the generation of a pattern from an initial uniform state, but it is not evident how the periodic pattern generates regions with unique phenotypes. Hence, PI and RD are often described as opposing ideas, or as Wolpert phrased it, "[RD] is the antithesis of the positional information approach" [236]. This,

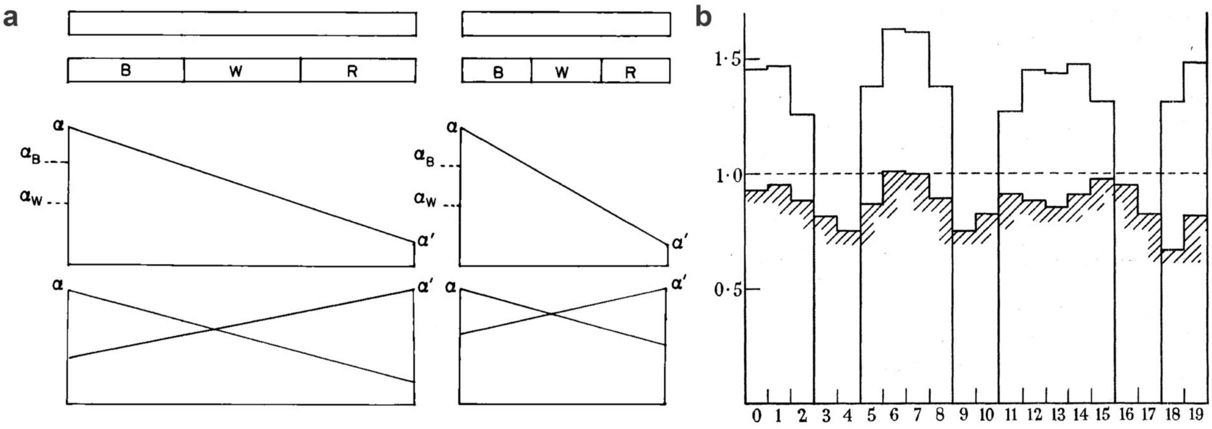


Figure 1.12: Historical sketches of positional information and reaction-diffusion mechanisms. **a)** Wolpert's positional information is illustrated by the French flag problem. Top: A long (left) and short (right) line of cells differentiates into three distinct regions (B: blue, W: white, R: red). Crucially, the differentiation is independent of the system size. Middle: This could be achieved by first establishing a linear morphogen gradient between the poles α and α' acting as source and sink, respectively. A GRN would then decide into which phenotype to differentiate by measuring the morphogen concentration compared to threshold values α_B and α_W. Bottom: Alternatively, PI could be provided by the ratio of two opposing gradients. Reproduced from reference [236] with permission of Elsevier, © 1971. **b)** Turing's RD mechanism using an activator-inhibitor system. The x-axis denotes a ring of 20 discrete cells with periodic boundary conditions, separated by a permeable membrane. The y-axis shows the evolution of the concentration profile of a morphogen (dotted line: initial uniform state, hatched line: incipient pattern, solid line: final pattern). Reproduced from reference [237] with permission of The Royal Society.

however, does not mean that both ideas are mutually exclusive, as there are various modes in which RD and PI work together [238]. For example, RD plays a crucial role in improving the robustness of dorso-ventral patterning in *Drosophila* [41], which is often regarded as a classical example of PI [40]. Alternatively, an RD mechanism could be used to generate the prepattern which is then interpreted in terms of positional information. In fact, Turing wrote [237]:

> "Most of an organism, most of the time, is developing from one pattern into another, rather than from homogeneity into a pattern."

Also Wolpert acknowledged this possiblity [241]:

> "[...] A number of models, mainly based on Turing's (1952) original ideas of a reaction-diffusion mechanism, have shown how [...] prepatterns could be generated [...]. All these mechanisms generate wave-like patterns that could generate repeated structures. A crucial feature of all these is that the peaks and troughs are all the same. [...] But, by combining them with positional information, which will make the waves non-equivalent, a very large variety of patterns can be generated [...]."

Biological examples of RD mechanisms include repetitive patterns like hair follicles [242, 243], feather buds [244, 245], and pigmentation patterns on flowers [246], but also the formation of rugae [247] and digits [248, 249]. One of the archetypical examples of a PI mechanism is the well studied gap gene network in *Drosophila*, as reviewed by Jaeger [40] (Figure 1.13).

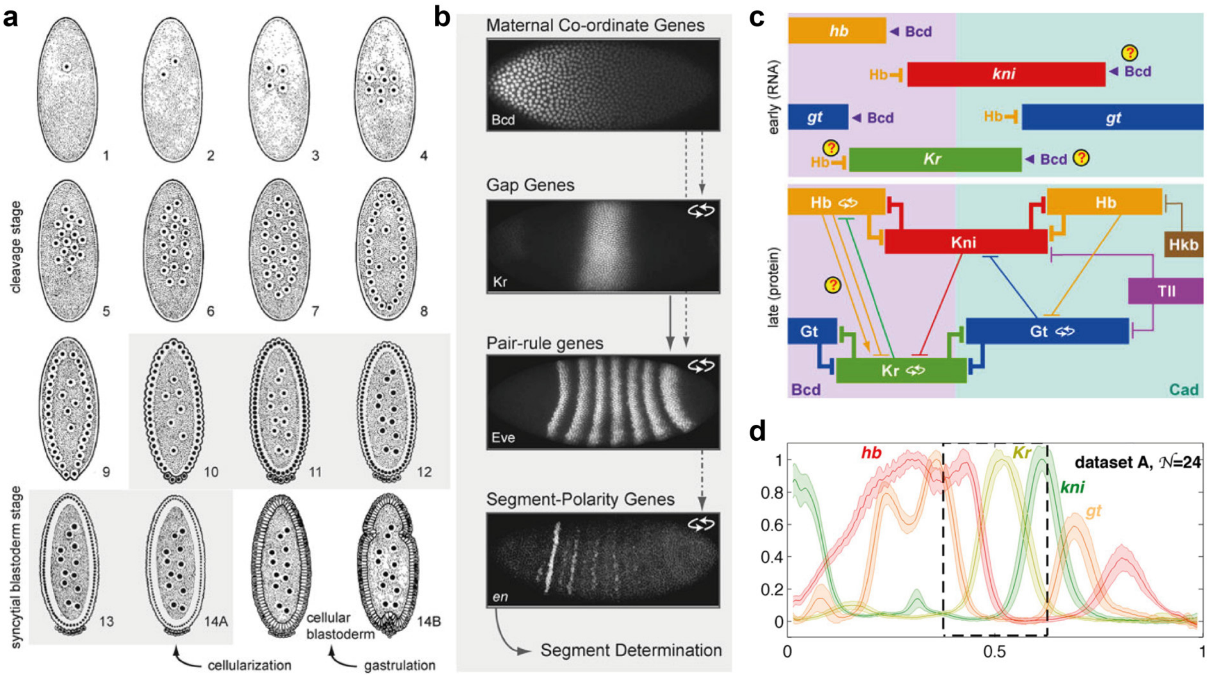


Figure 1.13: The gap gene network in *Drosophila*. a) The early phase of *Drosophila* development consists of an initial cleavage stage, followed by the blastoderm stage and gastrulation. The cleavage stage is characterized by nuclei divisions, as indicated by the number. Until the end of the blastoderm stage (gray), nuclei reside in a common cytoplasm, without separation by an additional membrane. **b)** GRN hierarchy that leads to the transformation of the maternal co-ordinate genes to segment-polarity genes. Arrows indicate regulatory interactions between GRNs, whereas circular arrows indicate cross-regulatory interactions within a GRN. The full length of the embryo is ≈ 500 µm. **c)** Regulatory interactions in the gap gene network. The early phase primarily features regulation by the maternal gradients, whereas the late phase is characterized by cross-regulatory interactions. The cross-regulatory interactions feature autoactivation, strong mutual repression between complementary gap genes and weak asymmetric repression between overlapping gap genes, resulting in positive feedback and incoherent feed-forward loops. (a-c) Reproduced from reference [40] under the license *CC BY-NC 2.0*. (a) Reproduced from reference [250] with permission of The Company of Biologists, © 1983. (b) Contains images from the FlyEx database [251, 252], used with permission of the Oxford University Press. **d)** Normalized experimental gap gene profiles along the AP axis and their standard deviation represented by the shaded area. Reproduced from reference [253] with permission of the Oxford University Press, © 2014.

The early phase of *Drosophila* development preceding gastrulation and the differentiation into different cell types can be divided into two phases, illustrated in Figure 1.13a. Starting from a single nucleus, nuclei divide until, after approximately 10 divisions, some form a single cell layer envelope, the blastoderm. Until gastrulation, nuclei inside the blastoderm reside in a shared cytoplasm without separation by additional membranes. It is in this phase that the essential body plan is established by a sequence of pattern forming GRNs (Figure 1.13b). The pre-pattern is formed by the maternal co-ordinate genes like *bicoid* (*Bcd*), which form morphogen gradients along the anterior-posterior (AP) axis of the embryo as a result of the localized distribution of maternal mRNA. These morphogen gradients then regulate the expression of the gap genes, which in turn regulate the pair-rule genes. These finally regulate the expression of the segment-polarity genes, which determine the differentiation into distinct body segments during gastrulation.

The gap gene network (Figure 1.13c) is not only regulated by the maternal morphogen gradients, but also features cross-regulation between the gap genes themselves. The GRN mainly consists of repressive interactions, featuring strong mutual repression between complementary gap genes and weak repression between overlapping gap genes. We can identify several network motifs, e.g. autoregulation, positive feedback loops and incoherent feed forward loops, that are common in developmental GRNs [2, p.99]. Positive feedback sharpens the boundaries between complementary genes and leads to bi-stability, hence differentiation will be permanent [43]. The weak repression results in several incoherent feed-forward loops, which have been found to be essential components of stripe forming mechanisms [44].

As shown in Figure 1.13d, the gap gene expression profiles are remarkably reproducible across multiple embryos. The high precision and robustness was one of the major critiques against a purely PI-based mechanism, as this would require very precise maternal gradients as well as very precise measurements by the GRN [254]. It was therefore argued that cross-regulation by diffusible morphogens may be a crucial component to achieve a high positional precision [255].

While RD and PI are very powerful ideas, there are also other mechanisms and theoretical concepts. For instance, genetic oscillators are used to clock a propagating morphogen wave front that directs the generation of somites in vertebrates [256] or to form circular stripe patterns with genetically engineered bacteria [45]. Some patterning mechanisms, such as the one underlying the pigmentation of lizard skin scales, which is primarily based on interactions between neighboring cells, can even be described by computational concepts like cellular automata (CA) [257]. Such a mechanism is an attractive option for the generation of synthetic patterns, as CA allow for universal computation [258] and therefore enable programmable pattern formation.

1.3.2 Patterning Precision in the Presence of Noise

The positional precision that can be achieved by a patterning mechanism in the presence of noise can be quantified with statistical measures such as positional error (PE) and positional information (PI) [253], as described in detail in Section 4.2.1 to Section 4.2.3). Briefly, the PE $\sigma_x(x)$ at position x (Figure 1.14a) can be calculated from the Fisher Information $\mathcal{I}(x)$, which in combination with the Cramér-Rao bound sets a lower bound for the PE

$$\frac{1}{\sigma_x^2(x)} \leq \mathcal{I}(x) = \int dg\; p(g|x)\left(\partial_x \ln p(g|x)\right)^2 \; . \tag{1.46}$$

The conditional probability distribution function (pdf) $p(g|x)$ can be estimated from experimental data such as shown in Figure 1.13d. For Gaussian pdfs with mean $\mu_g(x)$ and variance $\sigma_g^2(x)$, the Fisher Information is [253]

$$\mathcal{I}(x) = \frac{(\partial_x \mu_g(x))^2}{\sigma_g^2(x)} + 2\frac{(\partial_x \sigma_g(x))^2}{\sigma_g^2(x)} \; , \tag{1.47}$$

which means that information is stored in the slope of the gene expression gradient relative to the variation in gene expression levels, but also in the gradient of the variations. PE measures the local uncertainty of a position estimate based on a measurement of the gene expression levels and has been found to be as low as 1% across the AP axis of *Drosophila*, corresponding to a distance of about one nucleus [259].

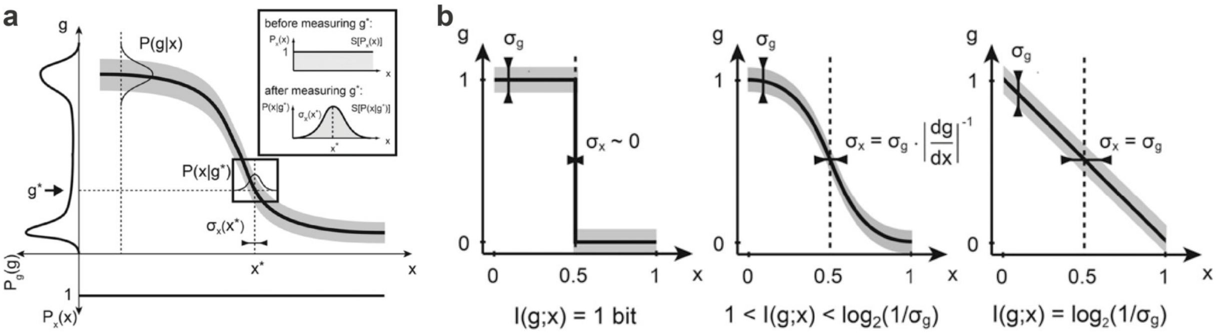


Figure 1.14: Positional error and positional information. **a)** The positional error in a gene expression gradient at position x reflects the accuracy of a position estimate based on a measurement of the gene expression gradient and its noise gradient. **b)** The positional information quantifies the number of regions that can be distinguished based on a noisy gene expression gradient. (a,b) Reproduced from reference [253] with permission of the Oxford University Press, © 2014.

PI can be formally defined as the Shannon mutual information that the gene expression levels g contain about position x [253]

$$I(x;g) = \int dx\ p_x(x) \int dg\ p(g|x) \log_2 \frac{p(g|x)}{p_g(g)}\ . \tag{1.48}$$

As illustrated in Figure 1.14b, PI is measured in bits and quantifies the global capability to generate distinct regions based on a gene expression profile. While single gap genes in *Drosophila* provide a PI of about 2 bit, a joint measurement of the four gap genes can provide more than 4 bit of information, allowing to in principle distinguish more than $2^4 = 16$ regions [253]. Regarding the debate of PI versus RD (Section 1.3.1) it is interesting to note that Alan Turing had independently developed a concept similar to Shannon's mutual information to measure correlations between two variables, but used base-10-logarithm and hence called the unit of measurement '(deci)ban' [260].

1.3.3 GRN Complexity and Robustness

Developmental GRNs can be characterized by various metrics including network complexity and robustness. Network complexity can be simply defined by the number of genes (nodes) and regulatory interactions (edges) [44], but may involve other structural measures such as network modularity compared to random networks, hierarchy levels, or network motifs. Robustness typically refers to a GRNs ability to generate a target pattern in the presence of fluctuations in parameters or network structure [35]. Note that network properties are often interdependent. For example it has been suggested that the modularity and robustness observed in many natural developmental GRNs emerge as a consequence of their evolution [261]. For the design of synthetic patterns, it is typically desirable to find network topologies with low complexity to simplify construction, characterization, and troubleshooting. Simultaneously, the GRN should generate the target pattern robustly in the presence of uncertainties in experimental parameters. As increasing complexity with additional regulatory interactions can improve robustness, this often results in a trade-off.

One approach to find a pattern forming GRN is to screen all possible networks with a given number of nodes for the target behavior. The network topologies can then be classified with a network atlas

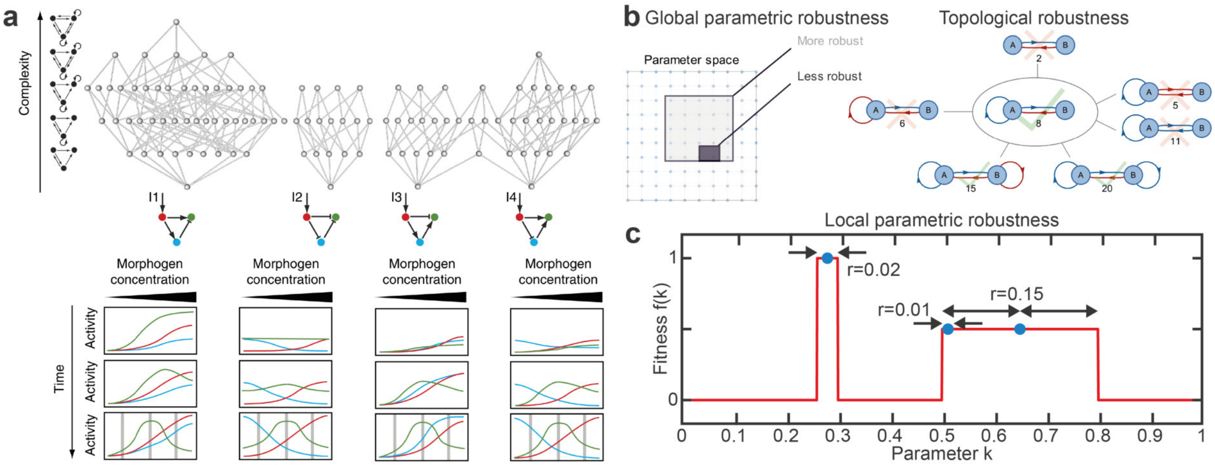


Figure 1.15: GRN complexity and robustness. **a)** Network atlas classifying all stripe forming 3-node networks according to their simplest topology, which are the four incoherent feed-forward loops. Reproduced from reference [44] under the license *CC BY-NC-SA 4.0*. **b)** Robustness measures. left: Global parametric robustness measured in terms of the viable fraction of parameter space. right: Topological robustness defined as the fraction of neighboring networks that show the target behavior. Reproduced from reference [35] with permission of Elsevier, © 2019. **c)** Local parametric robustness r, defined as the minimum distance in parameter space that leads to an undesired change in behavior (here measured by a fitness score). Here, $k = 0.28$ is the globally optimal solution, but $k = 0.65$ is the most robust. Although $k = 0.51$ is on the same plateau as the most robust solution, it is not robust.

(Figure 1.15a), by first connecting neighboring networks, i.e. networks that differ in only one edge, and then rank them by complexity. The networks then cluster around the most simple archetypical network that generates the target behavior. Using this approach [44] revealed that all stripe forming 3-node networks are based on the four basic incoherent feed-forward loops.

We can now ask which of the identified topologies are the most robust. In this context Scholes et al. [35] analyzed all possible 2-node and 3-node networks (with 2 diffusing nodes) for their capability to form Turing patterns. The authors concluded that while $> 61\%$ of all topologies exhibited formation of Turing patterns, there were only few topologies that showed a relatively high robustness. How to define robustness is highly dependent on the problem statement. Abstractly, robustness is a system's capability to maintain qualitative and quantitative features when parameters change, or components are added or removed. To obtain quantitative measures of robustness, Scholes et al. [35] distinguish '*global* parametric robustness', defined as the fraction of the sampled phase space that generates the target pattern, and 'topological robustness', defined as the fraction of neighboring networks that can generate the target pattern (Figure 1.15b).

While global robustness allows quantitative comparison of different network topologies to gauge which topologies would appear more likely in natural systems, it is for instance not clear how the global robustness of an N-dimensional system compares to an M-dimensional one. Another issue is that parameters have different units and that the system may be more sensitive to changes in one parameter, but more robust to changes in another parameter. From an engineering perspective, we are therefore interested in the '*local* parametric robustness' against expected changes and uncertainties in parameters σ_{k_i} that are known from characterization experiments. For a given network structure and parameter set, the local robustness can be determined with a one-at-a-time (OAT) sensitivity analysis, which additionally provides information about the shape of the parameter space. As shown in Figure 1.15c, the local robustness of a parameter k_i can be defined as the distance Δk_i to the near-

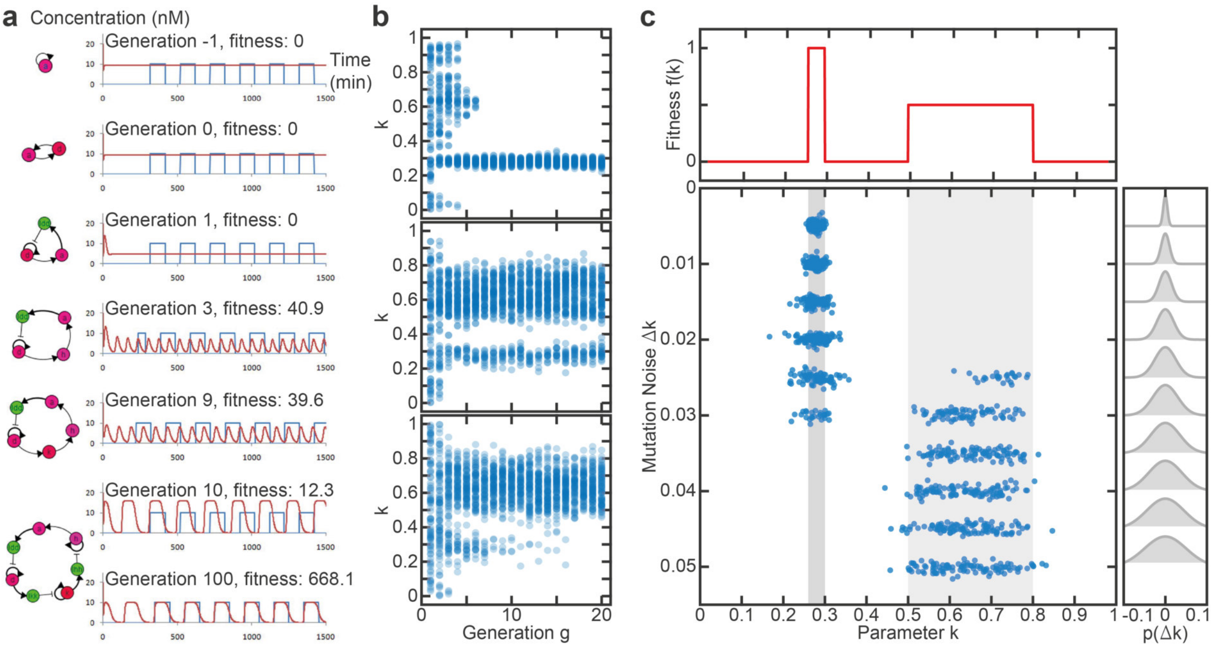


Figure 1.16: Evolving pattern forming GRNs. a) Evolution of a GRN that generates a temporal square wave function. The fitness is defined as a measure of the difference of the GRN output (red) to the target behavior (blue). Over 10 generations, the GRN grows in complexity and approaches the target behavior. Adapted from reference [263] with permission of the IEEE, © 2015. **b)** Evolution of a population of parameters in the simple fitness landscape shown in (c), over 20 generations and with varying mutation rates. The mutation noise was drawn from normal distributions with standard deviations of 0.01, 0.03, and 0.05 from top to bottom. **c)** Final parameter populations (jittered) after 20 generations for varying mutation rates. This example illustrates how the less fit, but more robust solution is preferred at higher mutation rates. Distributions on the right illustrate the width of the mutation noise compared to the fitness landscape.

est phase boundary in parameter space [262]. Similarly, we can define the relative local robustness as $\Delta k_i / k_i$. To maximize the probability to obtain the target pattern when designing a GRN, we may therefore require $\sigma_{k_i} < a \cdot \Delta k_i \; \forall i$, where $a > 0$ is a stringency parameter.

1.3.4 Parameter Optimization using Evolutionary Algorithms

To design GRNs with robust behavior, we would like to optimize the local robustness for multiple parameters simultaneously, and potentially extend the GRN search to networks with ≥ 4 nodes. As a comprehensive analysis of such large networks is typically prohibited due to computational cost, an interesting approach is to optimize GRNs using evolutionary algorithms (EAs). EAs are metaheuristics, i.e. optimization algorithms that are useful when the objective function is unknown, or not available, so gradient-based methods are not applicable. EAs follow the principle of natural evolution by maintaining a population of ‘genes’, which can be represented by topological features as well as systems parameters. The initial population then undergoes repeated rounds (generations) of mutation, selection (based on a computed fitness score) and reproduction. The two features that make EAs an attractive option for GRN design are the possibility to integrate topological and parametric evolution and the tendency to intrinsically yield robust solutions [262, 264]. For instance, Dinh et al. [263] evolved GRN topologies that produce rectangular oscillations using the PEN toolbox [160] (Figure 1.16a).

The potential of EAs to yield robust solutions is illustrated in Figure 1.16b,c. Consider the simple 1-dimensional fitness landscape as before, with a global optimum, and a less optimal, but more robust local optimum. We can now use a tournament selection (TS) algorithm. First, the fitness of a pair of parameters k_i, k_j is compared and the one with higher fitness is carried over to the next generation. Every survivor is copied and then all parameters are mutated by a quantity drawn from a normal distribution with standard deviation Δk, the mutation noise. Repeating the evolution over 20 generations for varying mutation noise, we can observe that the algorithm yields the narrow global optimum for low mutation noise, but as Δk increases, solutions settle in the less optimal, more robust optimum. Intuitively, at high mutation noise solutions in the narrow optimum tend to produce less fit offspring than in the broad optimum. Hence, the probability to propagate their genes to future generations is higher in the broad optimum then in the narrow one with higher fitness score. The fitness of the progeny is therefore an intrinsic contribution to the apparent fitness function in addition to the raw fitness score. We can now measure robustness as the resilience of a parameter set to mutations according to a given distribution, e.g. the number of healthy progeny. When experimental parameter tolerances and constraints are known (Section 1.1.4), the mutation noise can be chosen such that the EA yields experimentally viable solutions.

2 Material and Methods

We here describe key instruments (Section 2.1), microfluidic protocols (Section 2.2), and molecular biology techniques (Section 2.3).

2.1 Instruments

The instruments that were mainly used to generate the data presented in Section 3 to Section 5 are a fluorescence plate reader for bulk characterization and screening of GRN components (Section 2.1.1), and two epifluorescence microscope setups (Section 2.1.2, Section 2.1.3). The latter setup was equipped with custom built devices (Section 2.1.4, Section 2.1.5) for microfluidic experiments.

2.1.1 Plate Reader (BMG Labtech, Clariostar)

Bulk fluorescence experiments were performed using a fluorescence plate reader (Clariostar, BMG Labtech) set to appropriate wavelengths, measured through bottom optics with averaging over 20 flashes. Cell extract and in vitro transcription reactions were carried out in 384-well plates (Brand, #781687) using 15 µL reaction volumes, prepared either manually or using a droplet dispenser (Dispendix, I.DOT) with I.DOT pure plate90.

2.1.2 Inverted Epifluorescence Microscope (Nikon, Ti2-E)

The inverted epifluorescence microscope Nikon Ti2-E was equipped with an Andor Neo5.5 camera (pixel size 6.5 µm), a temperature controlled incubator (Okolab, #OKO-H201), motorized stage, and z-drive with perfect focus system (Nikon). A SOLA light engine SE II (Lumencor) was used for fluorescence excitation. The available objectives were a 4x (CFI P-Apo, NA 0.20, Nikon), 10x (CFI P-Apo, NA 0.45, Nikon), 40x (phase contrast, air, CFI P-Apo DM, NA 0.95, Nikon), and 60x (phase contrast, oil immersion, CFI P-Apo DM, NA 1.40, Nikon). We optimized the filter sets for parallel acquisition of up to 4 fluorescence channels with $< 1\%$ nominal bleed through [265], as listed in Table 2.1.

2.1.3 Inverted Epifluorescence Microscope (Olympus, IX81) and Microfluidics Setup

Microfluidic experiments were performed on a custom setup built around an Olympus IX81 epifluorescence microscope equipped with a motorized stage (Maerzhauser) and z-drive, camera Andor

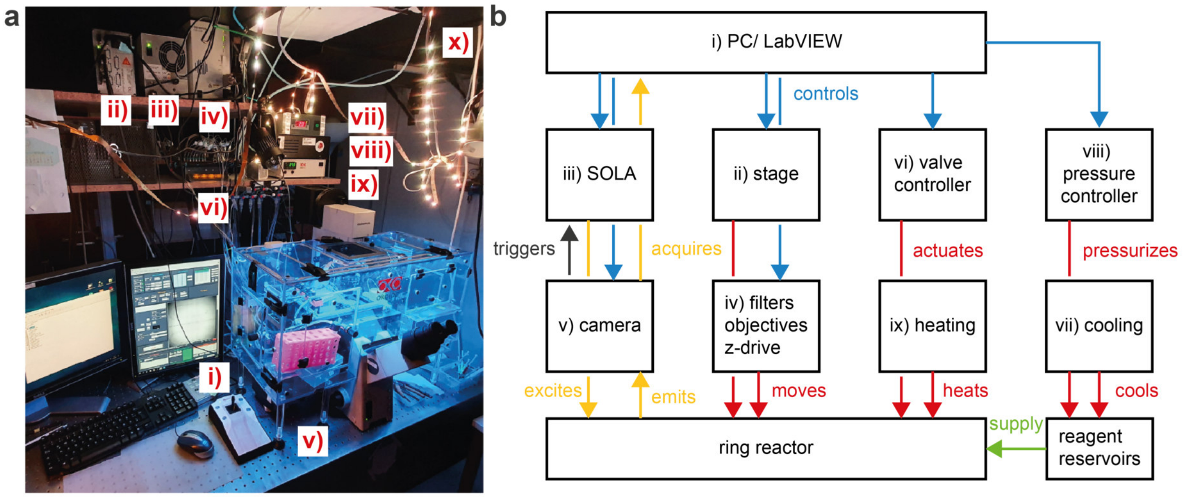


Figure 2.1: Overview of the microfluidics setup. a) Photograph and **b)** flow chart of the setup. The i) custom LabVIEW software controls the ii) xy-stage controller, iii) fluorescence illumination, iv) Olympus IX81 main unit including filter turret, brightfield shutter and z-drive, v) camera, and the vi) microfluidic valve controller. The vii) cooling unit, viii) Elveflow OB1 pressure controller, and ix) Okolab temperature incubator are independently controlled as these units do not need to be actuated during experiments. x) Atmospheric lighting ensures good visibility for the experimenter when connecting new microfluidic chips.

iXon3 DU888 (pixel size 13 μm), temperature controlled incubator (Okolab), a SOLA light engine SE II 365 (Lumencor) for fluorescence excitation, and 4x (UPlanFL N, NA 0.13, Olympus), 10x (UPlanFI, NA0.30, Olympus), 20x (UPlanFI, NA 0.50, Olympus), 40x (UPlanFI, NA 0.76, Olympus), and 100x (UPlanSApo, NA 1.40, Olympus) objectives. The motorized filter turret was equipped with the filter sets listed in Table 2.1, optimized for parallel acquisition of up to 5 fluorescence channels with $< 1\%$ nominal bleed through [265]. Brightfield images were acquired through the GFP emission filter.

For microfluidic experiments the setup, shown in Figure 2.1, was additionally equipped with a microfluidic pressure regulator (Elveflow, OB1), a custom built Quake valve controller (Section 2.1.4), and a custom built microfluidic reservoir cooling unit (Section 2.1.5). All acquisition and fluid handling components were controlled with a custom LabVIEW software that was conceived and implemented by Johann Brenner. Experiments can be performed by defining custom automated routines with integrated microscope actuation and image acquisition.

Table 2.1: Filter sets for Nikon Ti2-E (left) and Olympus IX81 (right). Ex.: excitation, DM: dichroic mirror, Em.: emission. Notation: central wavelength/ bandwidth.

Fluorophore	Ex.	DM	Em.
BF			
CFP	438/24	458	483/32
YFP	497/16	515	525/30
RFP	559/34	585	625/15
Cy5	628/40	660	692/40
GFP	472/30	495	520/35

Fluorophore	Ex.	DM	Em.
DAPI	366/10	405	442/42
CFP	438/25	458	483/32
YFP	500/20	515	535/30
RFP	559/34	588	609/34
Cy5	628/40	660	692/40
GFP	472/30	495	520/34

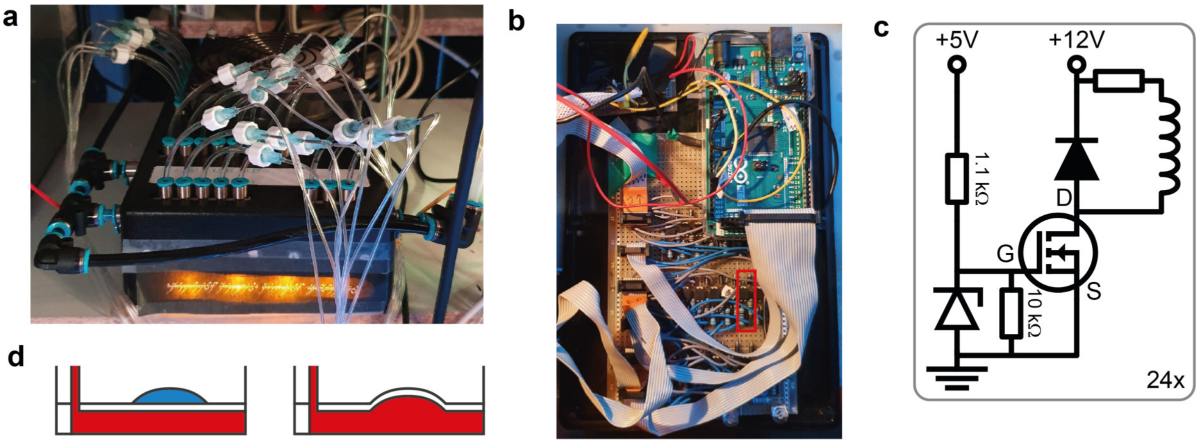


Figure 2.2: Quake valve controller. a) Outside view of the controller box. Valve manifolds are supplied by pressure input lines (black tubing) and connect to the microfluidic chip via transparent tubing. The orange status light indicates that the device is switched on and ready to rule all ring reactors. **b)** Overview of the main control circuit board. The Arduino Mega (top right) controls 24 transistor circuits that switch the solenoid valves. One unit circuit is highlighted with a red box. Ribbon cables connect between the Arduino, the transistor circuits and the solenoids. **c)** Circuit diagram of one circuit unit. The transistor is gated by the Arduino via a voltage divider stabilized with a Z-diode. A flyback diode across the inductive load avoids voltage spikes when switching. **d)** Schematic illustrating the operation of a Quake valve. When no pressure is applied to the control line (red), the valve is open and allows flow (blue). When pressurized, the PDMS membrane separating the control from the flow line is pushed into the flow channel and obstructs the flow.

2.1.4 Quake Valve Controller (DIY build)

The microfluidic Quake valve controller (Figure 2.2a) was custom built based on a total of 24 miniature solenoid valves (3/2-way, normally open), grouped in two 8x, one 5x and one 3x manifolds (Festo, #MH1-A-12VDC-N-TC-8V-PR-K05-QB-AD-BD-CX-DX, #MH1-A-12VDC-N-TC-5V-PR-K05-QB-AD-BD-CX-DX, #MH1-A-12VDC-N-TC-3V-PR-K05-QB-AD-BD-CX-DX, with cables, #NEBV-H1G2-KN-0.5-N-LE2). Input pressures were regulated between 1 and 3 bar using two pressure regulators (Festo, #MS4-LR-1/4-D5-AS). The input pressure of the 3x manifold was set with an independent pressure regulator so that the actuation frequency of a 3-valve peristaltic pump could be maximized.

The outputs of the solenoid valves were connected to the microfluidic chip using a series of adapters starting with 5 cm of 3 mm PUN tubing (Festo, #PUN-H-3X0,5-BL), connected to a 3/32-inch barb to male luer lock adapter (Masterflex, #45505-32, purchased from Fisher Scientific), followed by 0.6 mm × 80 mm (23 G) syringe needles (Sterican, #4291431) cut to a length of $\approx$ 10 mm. These were then connected to 1 m of Tygon ND 100-80 tubing (Saint-Gobain, #AAD04103, 0.5 mm ID, 1.52 mm OD, purchased from Fisher Scientific) and 0.7 mm (22 G) metal connectors cut to $\approx$ 25 mm from syringe needles (Sterican, #2050812). The metal connectors were bent into an L-shape aided by an M6 screw and inserted directly into the holes punched into the PDMS chip. The long and short side of the connectors were alternated between neighboring ports to optimize space usage.

The 24 solenoid valves were controlled by an Arduino Mega2560 Rev3 (#A000067) using the 5 V digital output pins to actuate the 12 V solenoid valves via a control circuit (Figure 2.2b,c), soldered onto a matrix board in 24 replicates. The circuit consists of an N-channel MOSFET (#IRLD110PBF) gated by the Arduino via a voltage divider (1.1 kΩ, 10 kΩ) and stabilized by a Z-diode 3.9 V (#BZX79-C3V9,113). The load circuit was powered with a 12 V, 10 A power supply and a 1 A flyback diode (#SB1100) was used to eliminate voltage spikes across the inductive load. Additionally, we used a

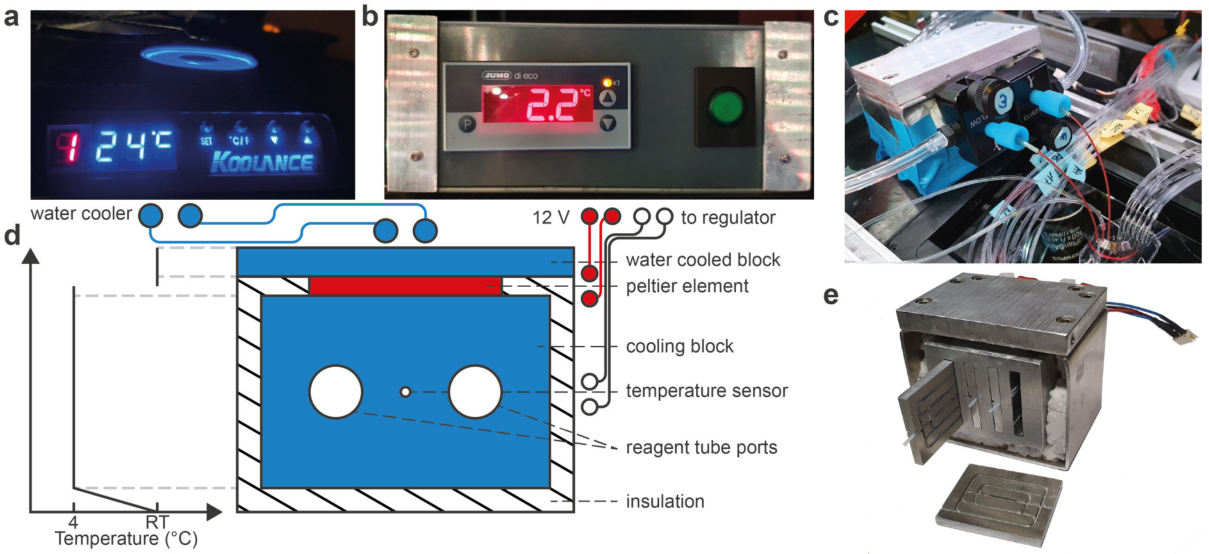


Figure 2.3: Reagent cooling unit. The cooling unit was used to cool temperature sensitive reagents in close proximity to the chip to minimize uncooled dead volume. **a)** CPU water cooler and **b)** temperature regulator are connected to **c)** the cooling unit, which consists of a peltier element that cools an aluminum block with two ports to hold 1.5 mL or 2 mL Eppendorf tubes. **d)** Schematic of the cooling unit consisting of a peltier element that is 2-point regulated by the temperature regulator using a pt100 temperature sensor. The heat generated at the hot side of the peltier is dissipated through an aluminum block that is maintained at room temperature by a CPU water cooler. To ensure a constant temperature at the sample reservoirs the block is insulated with styrofoam. The ideal temperature profile is sketched on the left. **e)** An alternative version of the cooling unit allows cooling of up to 4 times 50 µL or 8 times 25 µL of reagents stored in tubing.

decoupling capacitor 1000 µF, 16 V (#EEUFR1C102L) across the 12 V power supply and ribbon cables with appropriate connectors. All electronic components were purchased from Farnell, Reichelt, or Conrad.

The valves (Figure 2.2d) were actuated by controlling the Arduino with the custom LabVIEW software implemented by Johann Brenner. The software allows for manual switching of individual valves and simultaneous switching of multiple valves according to user-defined functions, including periodic actuation of peristaltic pumps. Valve functions were either used in manual mode or in automation routines.

2.1.5 Microfluidic Reservoir Cooling Unit (DIY build)

To allow cooling of sensitive reagents in close proximity to the chip with minimized uncooled dead volume, we built a custom designed cooling unit (Figure 2.3). The unit is built around an aluminum block, with two drilled ports that hold 1.5 mL or 2 mL Eppendorf tubes. Modeling clay was inserted into the holes to increase the contact area between the tubes and the block. An alternative, CNC-machined version of the cooling block allows cooling of up to 4 times 50 µL or 8 times 25 µL of reagents.

The cooling block is cooled to ≈ 4 °C using a peltier element (TRU COMPONENTS, #TES1-127025) that is operated at 12 V by a 2-point regulator (Jumo di eco, #701540/-821-02-061), switching at ±0.1 °C deviation from the target temperature as measured by a pt100 temperature sensor (Conrad, #181315-1/3-DIN-M-222) inserted into the center of the cooling block. The heat generated at the hot side of the peltier element was dissipated through an aluminum block with drilled channels that allowed perfusion with water that was kept at room temperature using a CPU cooler (Koolance, Exos-2 V2).

The cooling block was insulated thoroughly with styrofoam from all sides for three reasons. First, this minimizes the hysteresis caused by the 2-point regulator because the slower heating allows for a relatively faster response. Second, the temperature difference between the room temperature and the target temperature drops across the insulation, rather than across the block itself, so that the temperature within the block is more homogeneous. Third, the insulation covers any potentially cool surfaces from the surrounding air, which avoids excessive condensation of air humidity on the cooling block.

2.2 Microfluidics

In this section we describe the design, manufacturing and operation of microfluidic devices. The fabrication of master molds using photolithography is outlined in Section 2.2.1, Section 2.2.2, and Section 2.2.3. Fabrication of PDMS microfluidic chips by soft-lithography is described in Section 2.2.4 and Section 2.2.5. The operation of chips for droplet generation and the ring reactor are presented in Section 2.2.6 and Section 2.2.7, respectively.

2.2.1 SU-8 Photolithography

Chips were designed in AutoCAD. All photolithography steps were performed in a cleanroom using 2-inch or 3-inch silicon wafers (Siegert Wafer, #2P0/1-10/500±25/SSP/TTV<10) and SU-8 2050, or SU-8 3050 negative thick-film photoresist (micro resist technology), which were appropriately diluted according to manufacturer specifications to achieve resist thicknesses of ≤ 50 µm.

Wafers were first rinsed with isopropanol and acetone, dried with nitrogen and heated at 120 °C for 5 min, optionally followed by priming with Ti Prime (MicroChemicals), applied by spin-coating for 60 s at 300 $\mathrm{rpm\,s^{-1}}$ to 6000 rpm and baking for 2 min at 120 °C. Then, ≈ 2 mL of the resist were spin-coated for 15 s at 100 $\mathrm{rpm\,s^{-1}}$ to 500 rpm, 45 s at 300 $\mathrm{rpm\,s^{-1}}$ to $1500-6000$ rpm. The speed for the initial spread was increased to 1000 rpm for the thicker, more viscous resist formulations (SU-8 X025 and higher) to obtain a more homogeneous spread and the final spin speed was chosen as suggested by the manufacturer for the desired resist thickness or based on spin curves Figure 2.4a,b.

After optional relaxation overnight at room temperature to improve the homogeneity of the resist thickness, the wafers were then soft-baked (SB) by ramping from 65 °C to 95 °C following the manufacturer's instructions. The resist was exposed with $200-300$ mJ/cm^2 either on a maskaligner using Photomasks (Zitzmann GmbH, 64.000 dpi) or using a maskless aligner (MLA, Heidelberg Instruments), followed by a post exposure bake (PEB) from 65 °C to 95 °C and development with PGMEA (Merck, #484431-1L), according to the manufacturer's instructions. The wafer was cleaned by rinsing with isopropanol and dried with a gentle nitrogen blow. Leftover white residue was removed by rinsing the structures with fresh PGMEA, followed by isopropanol and a nitrogen blow.

The masters were finalized with a hardbake for 1 h at $120-150$ °C, and optionally silanized to minimize adhesion of PDMS to the master by incubation with ≈ 5 µL of trichloro(1H,1H,2H,2H-perfluorooctyl)silane (Merck, #448931-10G) in a weak vacuum for 2 h to overnight, followed by rinsing with ddH$_2$O. The resist thickness was measured with a Dektak 150 surface profiler (Veeco instruments). Discarded wafers were recycled using hot piranha (5-7:1 H_2SO_4:H_2O_2, 80 °C) for ≈ 30 min with gentle agitation.

2.2.2 AZ 40XT Photolithography

Microfluidic designs and wafers were prepared identically to SU-8. As AZ 40XT (MicroChemicals) is a positive photoresist, the designs were inverted by surrounding the structures with an appropriate box and using XOR logic.

First, ≈ 2 mL of the resist were spin-coated for 15 s at 100 $\mathrm{rpm\,s^{-1}}$ to 500 rpm, followed by 45 s at 300 $\mathrm{rpm\,s^{-1}}$ to 4000 rpm. After SB ramping from 70 °C for 20 s to 120 °C for 3.5 min and

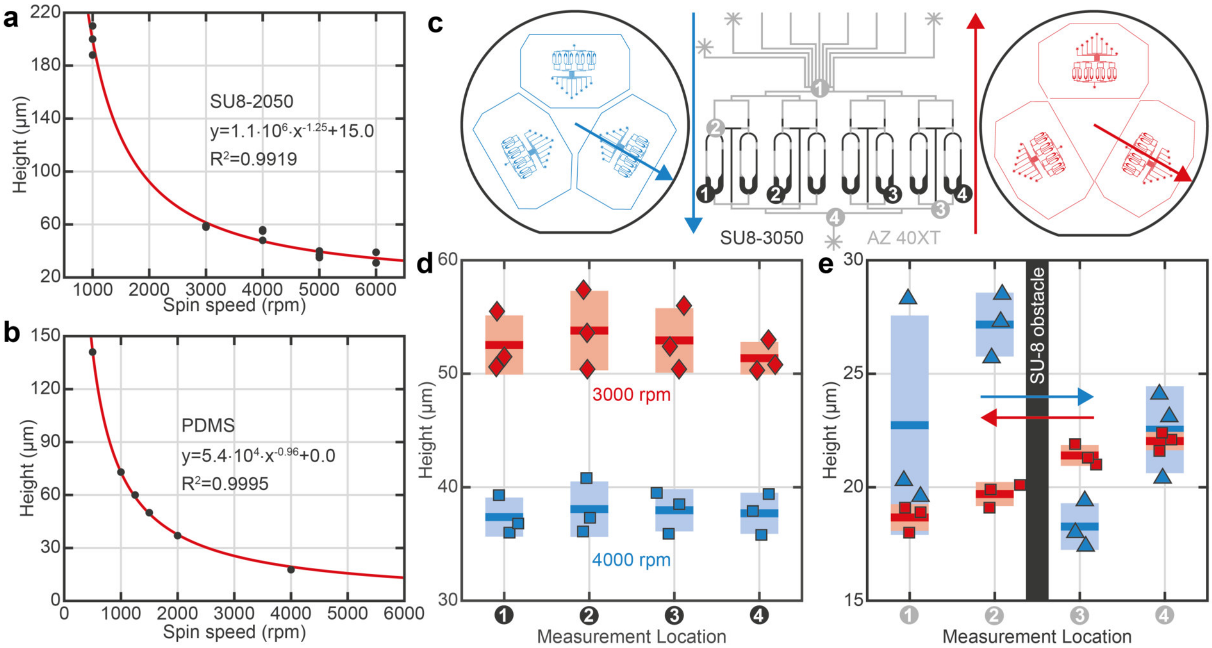


Figure 2.4: Spin curves and homogeneity of resist thickness. a) The spin curve for SU-8 2050 follows a power law. **b)** The spin curve for PDMS can be used to estimate the thickness of the membrane constituting Quake valves. **c)** Comparison of two chip orientations (left and right panel) on wafer layouts for the 8-ring flow layer that use a hybrid SU-8/ AZ 40XT design. The middle panel shows a zoom in on the structures where numbers indicate the location of height measurements in (d,e). Dark structures (SU-8) represent obstacles for the second spin with AZ 40XT. Arrows indicate the radial direction along which the resist is spun off. **d)** Height measurements for SU-8 3050 at the locations and for orientations as indicated in (c). CVs are about 5% across all locations for both chips. **e)** Height measurements for AZ 40XT for orientations and locations indicated in (c). It can be observed that the AZ 40XT structures are higher in front of the SU-8 obstacles, while being lower behind them. The inhomogeneity is less pronounced when relaxing the resist for 1 h prior to SB and switching the orientation (red). Arrows indicate the direction of spin induced resist flow.

back to 70 °C for 20 s, the resist was exposed with $200 - 300$ mJ/cm^2 using the MLA. After PEB at 70 °C for 20 s to 105 °C for 40 s and 70 °C for 20 s, the resist was developed in 2.38% aqueous tetramethylammonium hydroxide (TMAH, Merck, #331635-250ML) for ≈ 2 min and rinsed with dH_2O. Note that AZ 40XT is sensitive to isopropanol rinsing.

The masters were finished with a hardbake at 120 °C for 30 min, which allows the resist structures to reflow and adapt a round cross-section, which is necessary for proper closing of Quake valves, and silanization as for SU-8 structures.

2.2.3 Multi-Layer Photolithography

Chip designs with multiple channel heights or mixed round and rectangular channel cross-sections were manufactured in a multi-step process using alignment markers. Multiple SU-8 layers were formed without intermediate development and optional local development of the alignment markers. In contrast, hybrid SU-8 and AZ 40XT devices were manufactured with development after the SU-8 layer, as AZ 40XT is removed by the SU-8 developer PGMEA. If the SU-8 structures are of similar or larger height than the AZ 40XT structures, this can lead to buildup of AZ 40XT in front of and corresponding deficiency behind SU-8 structures when spin-coating (Figure 2.4c-e). To minimize adverse effects of uneven channel heights, like uneven closing of Quake valves, the radial orientations of the designs were optimized and the resist relaxed at RT for up to 1 hour prior to the soft bake.

2.2.4 PDMS Soft-Lithography

Generally, 12 g (25 g) PDMS and curing agent (Sylgard 184, Dow Corning) were mixed thoroughly in a ratio of 10:1, poured on a clean 2-inch (3-inch) silicon-master wrapped in aluminum foil, degassed for 15 min and baked for 60 min at 80 °C. The cured PDMS devices were carefully disassembled from the master and trimmed with a scalpel. Inlet holes were punched with either a 0.5 mm biopsy punch (WPI, #504528) when using Tygon ND 100-80 tubing (Saint-Gobain, #AAD04103, 0.5 mm ID, 1.52 mm OD, purchased from Fisher Scientific) with 0.6 mm OD metal connectors, or 1.25 mm biopsy punch (WPI, #504530) when using PTFE tubing (BOLA, #S1810-10, OD 1.6 mm, ID 0.8 mm). Instead of 0.5 mm biopsy punches, which are often of insufficient manufacturing quality for our purpose, catheter punches with a cutting edge of 0.74 mm (Syneo #CR0320245N21R4) can be used in combination with Tygon tubing and 0.7 mm OD metal connectors.

Devices were cleaned with Scotch Magic Tape and kept covered until bonding. As substrate $1'' \times 3''$, or $2'' \times 3''$ object slides were either coated with ≈ 5 g of PDMS to render the surface hydrophobic and minimize sticking of aqueous solutions, or cleaned with a kitchen sponge and soap, followed by rinsing with dH_2O, sonication in 2% Hellmanex III (Hellma, #9-307-011-4-507), sonication in ddH_2O and blow-drying with nitrogen. Finally, substrate and device were bonded by exposure to O_2 plasma (Diener Femto, 40 s, 20 sccm, 100 W), finished by baking at 80 °C for 1 h. To obtain all-hydrophobic channel walls, either all PDMS chips were recovered by baking at 200 °C for 3 h [266] or chips were flushed with 1% trichloro(1H,1H,2H,2H-perfluorooctyl)silane (Merck, #448931-10G) in FC-40 oil (Merck, #F9755-100ML) immediately prior to the experiment.

2.2.5 Quake Valve Soft-Lithography

Ring reactors and other chips that feature on-chip fluid control using Quake valves and peristaltic pumps were manufactured from two layers of PDMS as follows. For 3-inch masters (3 chips), 30 g of 5:1 PDMS (25 g PDMS + 5 g crosslinker) and 10.5 g of 20:1 PDMS (10 g + 0.5 g) were weighed and mixed thoroughly using a spatula. For 2-inch masters (1 chip), the amounts were 12 g of 5:1 and 10.5 g of 20:1 PDMS. Next, the 5:1 PDMS was completely poured over the flow-layer master wrapped in aluminum foil to form a mold, resulting in a thickness of ≈ 4 mm. The control-layer master was covered with ≈ 2 g of 20:1 PDMS, spread over the SU-8 structure by tilting the wafer, and kept in a closed petri dish at all times when outside a cleanroom.

After degassing both layers for 20 min in a dessicator, the control layer was spun off using a spin coater for 45 s at $1500-1800$ rpm, to reach a thickness of ≈ 50 µm (Figure 2.4b). The spin speed was optimized for each SU-8 master individually using the linear approximation that the thickness decreases by $\approx 4/100\ \mu\mathrm{m\,rpm}^{-1}$ in the relevant range. Because the presence of the SU-8 structures that form the control channels leads to a buildup of PDMS during spin coating, the PDMS was allowed to reflow for $45-60$ min to ensure an even surface and uniform thickness of the PDMS membrane.

Next, the PDMS was partially cured by baking for precisely $19-20$ min for the flow layer and $24-26$ min for the control layer in a pre-heated oven at 80 °C. To ensure reproducible bake times, it is crucial to place the wafers in a plastic petri dish to avoid direct heat transfer with the oven walls.

After appropriate curing the PDMS is not sticky when touched. Then, the flow layer was peeled off the master and immediately covered with Scotch Magic Tape.

After cutting out the chips with a scalpel and cleaning the structure with more Scotch Magic Tape, the flow layers were rapidly aligned on the control layer by eye. Note, that punching one of the inlet holes before the alignment can aid removal of air bubbles. Fine alignment was then performed using a stereo-microscope (Olympus, SZ61) by repeated lifting and realigning of the flow layer according to alignment markers and valve positions. Note, that the structures in the 5:1 PDMS shrink slightly compared to the original SU-8 structures on the master mold. To fit the size of the control layer structures, flow layer structures were enlarged in the original CAD drawing by $\approx 1.8\%$.

Remaining air bubbles between the two layers were removed by gentle pushing on the PDMS slab. Next, the wafer was again wrapped in aluminum foil and the remaining 20:1 PDMS was poured around the chips to facilitate later removal of the PDMS membrane from the wafer. After adding $\approx$ 10 g weights on top of the chips to avoid delamination from the wafer during bonding, chips were thermally bonded at 80 °C for 1 h.

Then, chips were again removed from the master, separated, trimmed and covered with Scotch Magic Tape. Holes for inlet, outlet and control line ports were then punched using catheter punches with a cutting edge of 0.74 mm (Syneo, #CR0320245N21R4) mounted on a drill re-purposed as a manual press. Punching precision was ensured using a budget digital WiFi microscope (Cainda, Amazon, #B081L4Y98G) coupled to a mobile phone or tablet and appropriate lighting. PDMS plugs were removed from the holes using tweezers. The resulting holes can be used with $0.6 - 0.8$ mm connectors.

Next, a $75 \times 50 \times 1$ mm^3 glass slide was first cleaned with a kitchen sponge soaked in Hellmanex III (Hellma, #9-307-011-4-507), rinsed with dH_2O, sonicated for 2 min in Hellmanex III, rinsed with ddH_2O, sonicated for 2 min in ddH_2O, and blow-dried with pressurized N_2. Both sides of the PDMS chips were thoroughly cleaned and covered with Scotch Magic Tape until they were bonded to the glass slides using O_2 plasma (Diener Femto, 40 s, 20 sccm, 100 W). Finally, bonding was finished by curing the chips at 80 °C for 1 h.

2.2.6 Droplet Generation

For simple droplet generation and encapsulation of gel beads, we used flow-focusing devices [268] with 25×40 µm^2 (width × height) and 40×60 µm^2 cross-section at the junction, respectively (Figure 2.5).

The sample reservoirs were connected to the chip with $\approx$ 20 cm of stiff and inert PTFE tubing (BOLA, #S1810-10, OD 1.6 mm, ID 0.8 mm), directly plugged into the PDMS. Prior to the experiment the chips were primed with 1% trichloro(1H,1H,2H,2H-perfluorooctyl)silane (Merck, #448931-10G) in FC-40 oil (Merck, #F9755-100ML) through the outlet, then flushed with the continuous phase and finally with the aqueous phase by slowly increasing the pressures until a stable droplet production was achieved. Droplet devices were operated with an Elveflow OB1 pressure controller with typical pressures of $100 - 300$ mbar. The continuous phase consisted of FC-40 oil (Merck, #F9755-100ML) with 2% (w/w) PFPE/PEG-surfactant [178] (Emulseo, #1903-09). The aqueous phase typically contained $30 - 100$ µL of transcription or transcription/translation mixture or 180 µL of modified PITC-agarose (Section 2.3.8). The droplet generation was monitored with one of the microscopes described in Section 2.1.2 and Section 2.1.3. The droplet generation frequency typically was in the order of $1 - 2$ kHz.

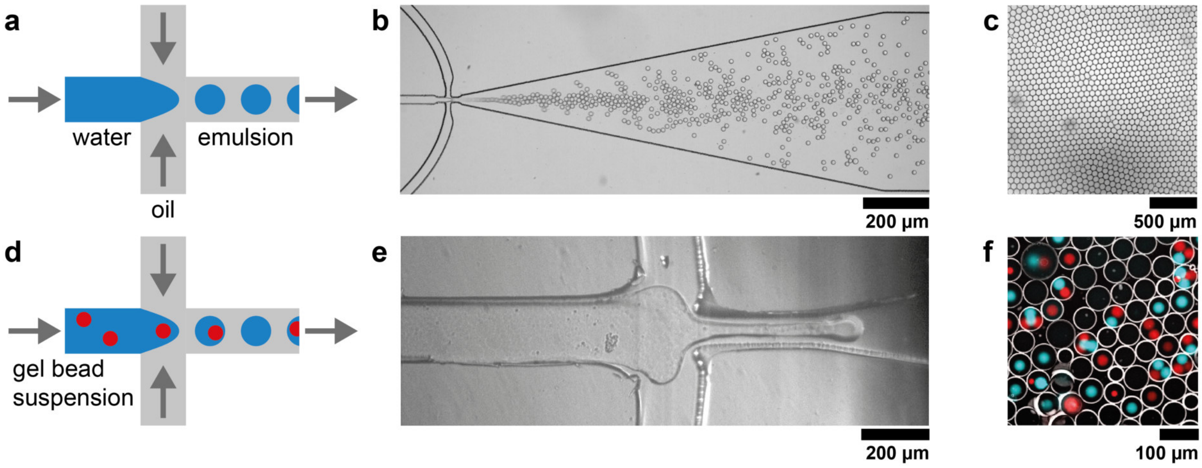


Figure 2.5: Droplet generation with microfluidic flow-focusing devices. **a)** At the flow-focusing junction (FFJ), the aqueous phase meets with the oil phase. The orifice in the downstream channel leads to droplet breakup via ‘dripping’ or ‘jetting’ (e) [267]. **b)** Microscopy image of a FFJ during operation at ≈ 1000 droplets/s. **c)** Hexagonal lattice of monodisperse droplets. **d)** Microgel spheres were encapsulated into droplets with a FFJ. **e)** Microscopy image of microgel encapsulation in the jetting regime. Hydrogel particles produce low contrast due to the similar refractive index as the surrounding water. **f)** Fluorescence false-color image of a mixture of fluorescently labeled microgel spheres encapsulated in droplets (white).

Droplets were collected in an Eppendorf tube connected to the outlet via ≈ 5 cm of PTFE tubing and imaged in custom glass chambers as described in reference [37]. Briefly, a microscopy slide and a cover slide were hydrophobically treated with water repellent (Rain-X, #5026349019096), or vapor deposition of trichloro(1H,1H,2H,2H-perfluorooctyl)silane (Merck, #448931-10G) and glued together with dental cement (picodent twinsil 22, #1300-1000) that was mixed with 75 µm glass beads (Merck, #59200-U), or 32 µm bond line spacer beads (Cospheric, #S-SLGMS-2.5 29-32um - 1g) to yield a defined height. Droplet videos were generally analyzed by segmenting the BF images in FIJI and then tracking the fluorescence intensity with a custom MATLAB plugin[1], developed by Dr. Korbinian Kapsner [269].

2.2.7 Ring reactor

Ring reactors (Figure 2.6a) were operated as described in references [23, 210]. First, control lines were connected to the corresponding ports (Figure 2.6b), ensuring that every line is filled with ≈ 10 cm of ddH_2O to avoid pushing air into the flow channels. The control lines were then pressurized with 1 bar until the air has been pushed out of every control channel. Then, the pressure was increased to $1-3$ bar until all valves were visually closed.

Next, the chip was calibrated to determine the fraction of the reactor volume R_0 that is replaced with each feed step, to account for variations in the dimensions of rings and feed pump (Figure 2.6c). First, solutions of 25 µM fluorescein in PBS and pure PBS were connected to two of the flow lines and pressurized to 200—300 mbar. The reagents were flushed into the chip and any air that was trapped in the flow layer was removed by closing the outlet valve (B5) so that the air was pushed out through the PDMS. The calibration routine starts with loading the reactors with PBS and acquiring a blank image for each reactor. Then, the reactors are loaded with fluorescein, the bypass channels are

---

[1] `https://github.com/kkapsner/Matlab.git`

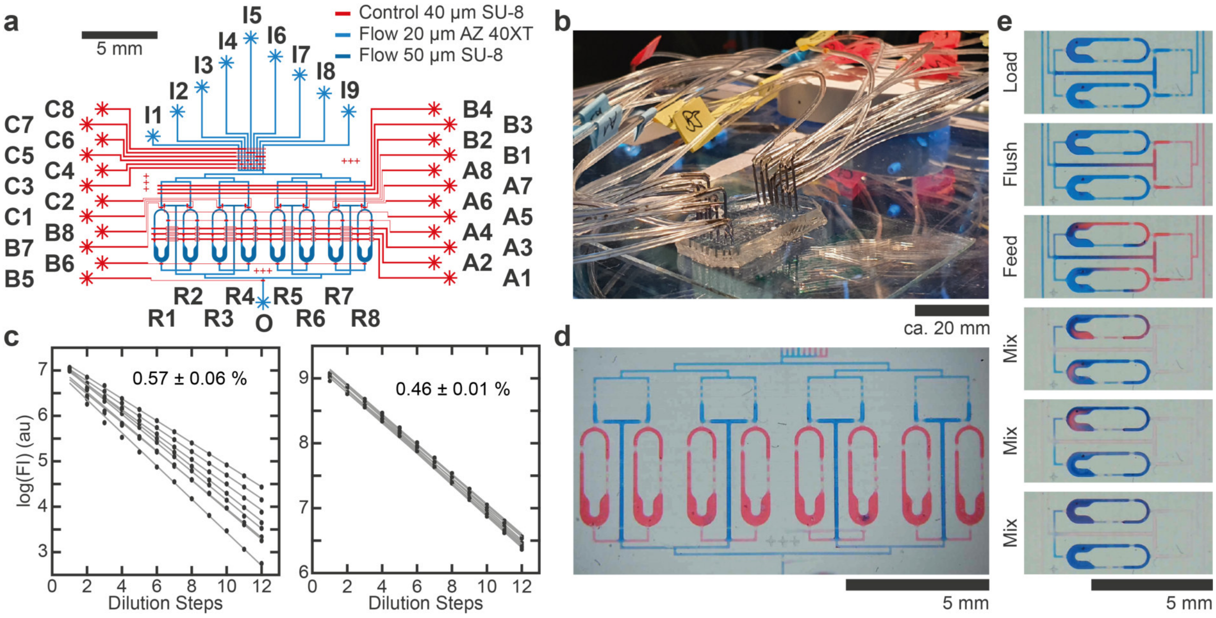


Figure 2.6: 8-ring reactor. **a)** Chip layout with control lines (red) and flow lines (blue). Labels indicate inlets (I), outlet (O), rings (R) and control line ports (A,B,C). The inlet is selected via a multiplexer (C3-8) and the flow is directed by appropriately switching the valves. A feed pump (B2-4) allows to replace defined fractions of the reactor volume with fresh reagents, followed by mixing the reactor content with a mixing pump (A1-3). **b)** Photograph of a chip connected to control lines. **c)** Example calibrations for a chip with heterogeneous (left) and more homogeneous (right) channel dimensions. The reactors are initially filled with fluorescein in PBS, which was then serially diluted with pure PBS with a fixed number of feed steps, followed by thorough mixing. As the fraction of reactor volume that is replaced with each feed cycle R_0 depends on the reactor volume, as well as on the feed pump volume, the calibration corrects for chip-to-chip as well as reactor-to-reactor variations. **d)** Photograph of a chip loaded with red (rings) and flushed with blue (bypass channels) food color. Control lines are transparent. **e)** Time series of a typical dilution and mixing cycle.

flushed with PBS (Figure 2.6d) and an image is acquired for each reactor, followed by feeding the reactors with a fixed amount of feed steps $F = 45$ and thorough mixing (500 cycles at 3 Hz). This process (Figure 2.6e) is iterated 11 times so that a dilution series is obtained. The fraction of the reactor volume R that is replaced after each dilution step n can then be determined for each reactor by measuring the fluorescence intensity I within an appropriate region of interest for each reactor. As $I_n = c \cdot (1-R)^n$, R can for instance be determined from the slope of a linear fit to $\log(I_n)$ and $R_0 = R/F$, as shown in Figure 2.6c.

Finally, the reagents for the experiments were connected to the inlets. Typically, reagents were stored in 1.5 mL Safe-Lock Tubes (Eppendorf, #0030123301) connected to the flow pressure lines controlled by the pressure controller. Uncooled reagents were connected to the chip via ≈ 20 cm of Tygon ND 100-80 tubing (Saint-Gobain, #AAD04103, 0.5 mm ID, 1.52 mm OD, purchased from Fisher Scientific) and a 1 cm long PEEK capillary (Trajan Scientific, #1301005001-10F, OD 0.8 mm, ID 0.125 mm, purchased from Fisher Scientific). Reagents that were cooled with the 2-port Eppendorf tube cooler were connected via 12.5 cm of PEEK capillary to minimize uncooled volume.

For CFPE the reservoirs typically contained 45 µL of extract with additional proteins and 45 µL of buffer with DNA templates and other additives. As extract and buffer were mixed on-chip in a 50:50 ratio, the concentrations of additives were adjusted accordingly. The extract reservoirs were typically cooled, whereas the buffer was supplemented with 2 mM TCEP (Roth, #HN95.1) to allow for storage without cooling [21, 202]. An additional flow line contained nfH_2O to flush the bypass channels between cycles to minimize accumulation of aggregates during the experiment. A typical experiment consisted of periodic feeding, mixing, and acquisition every 15 min over 24 h.

Microscope images were analyzed with custom MATLAB scripts. First, an ROI and background ROI were manually selected for each ring and the background subtracted average intensity $I - B$ was normalized against the corresponding measurement from the R_0 calibration $I_0 - B_0$. Using similarly generated 1-point reference measurements $I_{ref} - B_{ref}$ (1 µM of reporter protein in cell extract) and $I_{0,ref} - B_{0,ref}$ (25 µM fluorescein in PBS), we obtain a reporter concentration c, which is comparable across experiments

$$c = \frac{I - B}{I_0 - B_0} \cdot \frac{I_{0,ref} - B_{0,ref}}{I_{ref} - B_{ref}} \cdot 1\ \mu\text{M}\ . \tag{2.1}$$

Maxima of oscillatory trajectories were detected using *findpeaks* and auto-correlation was computed using *xcorr*. The rotation number m was determined by computing the auto-correlation of the maxima.

Simulations were performed in MATLAB using *ode23s* to solve the ODE model during the interval of one dilution cycle $t_{int} = 15$ min. After each cycle, the initial conditions of the consecutive cycle c_{i+1} were set with the final concentrations of the previous cycle c_i as $c_{i+1} = c_i \cdot (1 - R) + c_{in}$, where R is the refresh ratio and the input amplitude $c_{in} = A_{in}$, if i mod $k = 0$, and $c_{in} = 0$, otherwise. To approximate the experimental procedure, we sample the solutions at the final time point of each interval.

2.3 Molecular Biology

In this section we present the molecular biology techniques that were used to implement GRNs. Gene template assembly is described in Section 2.3.1, followed by in vitro transcription (Section 2.3.2) and cell-free gene expression (Section 2.3.3) reactions. We next present the protocols for DNA (Section 2.3.4), RNA (Section 2.3.5), and protein (Section 2.3.6) purification. In Section 2.3.7 we outline the procedure for cell extract preparation, followed by the protocol for PITC-agarose preparation in Section 2.3.8.

2.3.1 Gene Template Assembly

Genetic constructs were assembled using a standardized Golden Gate Assembly (GGA) scheme adapted from [57]. Each construct is divided into 4 functional parts: the promoter, the 5'UTR, the gene of interest, and the terminator. The parts were either pre-cloned into pSB1C3 using Esp3I (NEB, #R0734S) or PCR amplified using Phusion 2x Mastermix (NEB, #M0531S) and then assembled into pSB1A3 using BsaI-HF-v2 (NEB, #R3733S). Note, Esp3I is an isoschizomer of BsmBI with a superior activity in T4 ligase buffer and is therefore preferred in GGA reactions.

Overhang PCRs to introduce the cut sites required to assemble the initial parts were performed with primers purchased from IDT using a touch-up protocol starting at the annealing temperature [2] and ramping up to the elongation temperature of 72 °C in 1 °C increments. Other parameters were chosen according to the manufacturer's instructions.

The GGAs were performed in 10 µL reactions containing 0.5 µL of the restriction enzyme, 0.5 µL T4 DNA ligase (NEB, #M0202S), 10 fmol of the backbone and 20 fmol of each insert in 1x T4 DNA ligase reaction buffer (NEB, #B0202S). Using a thermocycler, the reactions were first heated to 37 °C for 15 min (initial cutting), then cycled 30 times between 37 °C for 1.5 min (cutting) and 16 °C for 3 min (ligation), followed by a final cutting step at 37 °C for 5 min and heat inactivation of the enzymes at 80 °C for 20 min.

Assemblies were then transformed into chemically competent DH5α, DH5αZi (when using LacI repressible promoters), DH5αZ1 (when using TetR repressible promoters), or NEB stable (NEB, #C3040H, when everything else had failed). 5 µL of the assembly mix were added to 50 µL of cell suspension and chilled on ice for 15 min. After heat shock at 42 °C for 45 s, cells were chilled on ice for 2 min and then recovered in 1 mL SOC-medium for 1 h. Next, the cultures were concentrated by centrifugation at 2,000 g for 3 min and removing 900 µL of the supernatant. The bacteria were resuspended in the remaining medium and plated using sterilized glass beads.

On the following day, 2-6 colonies were screened for correct assemblies via colony PCR with 2x OneTaq Mastermix (NEB, #M0482S) and appropriate primers in reaction volumes of 10 µL. After addition of 2 µL 6x purple loading dye (NEB, #B7024S) supplemented with 10x GelRed (Biotium, #41003), the lengths of the amplicons were verified by $1-2\%$ agarose gel electrophoresis at $150/8\ \mathrm{V\,cm^{-1}}$ for $\approx$ 20 min. Selected colonies were cultured overnight in 5 mL LB-medium and the plasmids were extracted using a Miniprep Kit (Qiagen, #27104) and veryfied by sequencing (Microsynth Seqlab, Economy Run). Concentrations were estimated using a UV/vis Nanophotometer (Implen).

---

[2] `tmcalculator.neb.com`

2.3.2 In Vitro Transcription

In vitro transcription reactions contained 1x transcription buffer (NEB, #B9012S, 40 mM Tris–HCl, 6 mM $MgCl_2$, 1 mM DTT, 2 mM spermidine, pH 7.9), supplemented with 4 mM of each rNTP, counterbalanced with additional 12 mM $MgCl_2$, typically $10 - 200$ nM of transcription template, 1 unit μL^{-1} RNase inhibitor (NEB, #M0314S), $50 - 200$ nM T7 RNA polymerase that was purified by Dr. Sandra Sagredo, and assay-specific optional additives. The reactions were filled to final reaction volumes of 15 µL (for 384-well plate), $50 - 200$ µL (for encapsulation in microfluidic droplets), or 100 µL (for RNA purification) with nuclease-free H_2O (Roth, #T143.2). Unless otherwise noted, reactions were performed at 37 °C for $6 - 20$ h.

2.3.3 Cell-free Gene Expression

Bulk CFPE reactions were either carried out in homemade *E. coli* cell extract (Section 2.3.7) or PURExpress (NEB, #E6800) in reaction volumes of 15 µL. First, the extract (E) and buffer (B) (PURE solutions A and B) that were stored at –80 °C were thawed on ice. Meanwhile, appropriate amounts of DNA templates and small molecule additives were mixed in nfH_2O. Then, B was mixed with E and the mixture was added to the DNA mixtures. Protein and RNA additives were optionally diluted in PBS and added to the final reaction mix. The final composition of a cell extract reaction was: 33% (v/v) cell extract, 42% (v/v) buffer and 25% (v/v) DNA plus additives. The final composition of a PURE reaction was: 40% (v/v) solution A, 30% (v/v) solution B, and 30% (v/v) DNA plus additives. Expression was carried out at 30 °C for $4 - 10$ h.

2.3.4 DNA Purification

Linear templates for IVT reactions (genelets) and CFPE reactions were synthesized by different methods, depending on length, required amount and whether multiple variants were screened for function. Templates below 90 bp were usually annealed in 1x transciption buffer from two full-length complementary oligos synthesized by IDT by ramping from 95 °C to 4 °C with 2 $°C\,min^{-1}$. Longer linear templates were PCR amplified from either IDT Ultramers (up to ≈ 170 bp), IDT g-Blocks, GGA reactions, purified plasmids, or colonies using appropriate primers and Phusion 2x Mastermix (NEB, #M0531S). Optionally, plasmids were digested with DpnI (NEB, #R0176S) for 1 h at 37 °C after the PCR. To screen multiple variants of genelets up to ≈ 150 bp, it can be efficient to use two IDT oligos with complementary 3'ends and fill in remaining bases with a PCR reaction. The first oligo here typically contains the T7 promoter and serves as the master strand, whereas the second oligo is variable and contains the sequence of the transcript. PCR reactions were purified using a Monarch PCR & DNA Cleanup Kit (NEB, #T1030L) or Monarch DNA Gel Extraction Kit (NEB, #T1020S) following 1-2% agarose gel electrophoresis, eluted with 30 µL of nfH_2O and quantified using a UV/vis Nanophotometer (Implen).

Plasmids were prepared by first streaking the corresponding bacteria from a glycerol stock overnight. Then, one colony was cultured overnight in 5 mL LB-medium and the plasmids were extracted using a Miniprep Kit (Qiagen, #27104) or the overnight culture was used to inoculate a 150 mL culture for a Midiprep (Qiagen, #12243). Optionally, the plasmids were further purified with phenol-chloroform extraction (PCE) using Phase Lock Gel (Quantabio, #2302820), phenol/chloroform/isoamyl-alcohol (P/C/I, Roth, #A156.1) and chloroform (Roth, #7554.1), followed by ethanol precipitation with 0.1 volume of 3 M sodium acetate pH 5.2, and 3 volumes of ice cold (–20 °C) 100% ethanol. After 1 h at

−80 °C the pellet was washed with 1 mL of cold 70% ethanol, dried briefly in a vacuum centrifuge and resuspended in nfH_2O. Concentrations were estimated using a UV/vis Nanophotometer (Implen) and the samples were stored at −20 °C.

2.3.5 RNA Purification

RNA was prepared with an IVT reaction from appropriate DNA templates, followed by DNA digestion with DNase I (NEB, #M0303S) at 37 °C for 30 min. The product was purified by PCE as for DNA, except using P/C/I for RNA extraction (Roth, #X985.1), or gel purification with a small-RNA PAGE Recovery Kit (Zymo Research, #R1070) following an 8 M Urea denaturing-PAGE for typically 90 min at 100/8 $\mathrm{V\,cm^{-1}}$. RNA was quantified with denaturing-PAGE stained with SYBR Green II (Thermo Fisher, #S7564) and RiboRuler Low Range RNA Ladder (Therom Fisher, #SM1831) for reference. Purified RNA was stored at −80 °C.

2.3.6 Protein Purification

For protein purification we used standard Ni-NTA His-tag affinity chromatography. First, the gene of interest was cloned into a linearized expression vector with a T7 promoter and a C-terminal His-tag (6xHis-pSB1A3-pT7) using GGA and transformed into BL21star(DE3). For expression, a 5 mL overnight culture was used to inoculate a 500 mL culture that was grown at 37 °C to OD 0.4-0.6, before inducing the expression by addition of IPTG to a concentration of 1 mM. The expression was conducted at 16 °C overnight. Cells or proteins were kept on ice or 4 °C during all subsequent steps. First, cells were harvested by washing once with 50 mL buffer A (50 mM Tris−HCl, 500 mM NaCl, pH 8.0, 25 mM imidazole), pelleted in a 50 mL Falcon tube, weighed, and optionally stored at −80 °C.

Next, the pellet was resuspended in 20 mL buffer A, supplemented with 1 $\mathrm{mg\,mL^{-1}}$ lysozyme (Roth, #8259.1), 1 mM PMSF (Roth, #6367.1), 1 mM benzamidine (Roth, #CN38.1), and 1 μL Turbo DNase (Thermo Fisher, #AM2238) and incubated for 1 h on ice. The cells were then lysed by sonication (Bandelin, Sonoplus) 50% Amplitude, 9 cycles, 30 s/15 s pulse. The cell debris was removed by centrifugation for 15 min at 7,000 g in the 50 mL Falcon tube and then 10 min at 20,000 g in 2 mL Eppendorf tubes.

The supernatant was loaded into HisTrap HP columns (GE, #17-5247-01) primed with buffer A using a syringe pump at 1 $\mathrm{mL\,min^{-1}}$. The protein was eluted starting with 1 column volume (CV) of buffer A containing 25 mM imidazole, then linearly increasing the imidazole concentration to 250 mM over 10 CVs, and 1 CV at 250 mM imidazole using an Äkta pure (GE). 5 μL of the collected fractions were added to an equivalent amount of 2x Lämmli buffer (Merck, #S3401) and analysed via SDS-PAGE (12% separation gel, 4% collection gel, 37.5:1 acrylamide:bis-acrylamide (Roth, #T802.1), $30 - 45$ min, 120 V, stained with Coomassie-stain (Roth, #4829.2), and pre-stained protein ladder (Roth, #8271.2)).

Sufficiently pure fractions were pooled and the buffer was exchanged with 15 mL, 10 kDa Amicon filter columns (Merck, #UFC901024), by washing 3x with 10 mL of storage buffer (20 mM Tris−HCl, pH 7.9, 150 mM NaCl, 1 mM DTT). Concentrations were measured by UV/Vis spectroscopy (Implen) using extinction coefficients from the literature. After addition of 15% Glycerol, aliquots of 5−20 μL were flash frozen in liquid nitrogen and stored at −80 °C.

2.3.7 Cell Extract Preparation

E. coli cell extract was prepared according to the protocol by Falgenhauer et al. [270] based on reference [54]. $1-6\times0.5$ L cultures of BL21(DE3)-Rosetta2 in 2xYTP medium were inoculted with 5 mL overnight cultures and grown to OD 1.8-2 at 37 °C, 250 rpm. The cells were harvested by centrifugation for 15 min at 4778 g and washed 2 times with 200 mL of buffer S30A (14 mM Mg-Glutamate, 60 mM K-Glutamate, 50 mM Tris, titrated to pH 7.7 with acetic acid, 2 mM DTT added immediately before use) per 0.5 L of culture. Cells were resuspended in 40 mL S30A per centrifuge bottle, then pelleted in Falcon tubes for 10 min at $2,000$ g, weighed and optionally stored overnight at −80 °C.

If stored at −80 °C, pellets were thawed on ice and resuspended with 0.8× the pellet mass of S30A by vortexing. 8 $\mathrm{mg\,mL^{-1}}$ of lysozyme (Roth, #8259.1) were dissolved in S30A, 0.2× the pellet mass was added to the cell suspension and the mixture was incubated for 15 min on ice. Next, the slurry was divided into 1.5 mL samples in 2 mL Eppendorf tubes and sonicated with 12 cycles of 10 s at 10% amplitude (Bandelin, Sonoplus), alternating between samples and stirring with the sonotrode between pulses. Cell debris was removed by centrifugation at $20,000$ g for 40 min, after which the supernatant was carefully transferred to fresh 1.5 mL Eppendorf tubes and incubated at 37 °C for 80 min to digest genomic DNA (run-off reaction).

Next, samples were again centrifuged at $12,000$ g for 10 min and the supernatant was carefully transferred into 10 kDa dialysis tubing (Thermo Fisher,#88243) soaked in buffer S30B (14 mM Mg-Glutamate, 60 mM K-Glutamate, titrated with 2 M Tris to pH 8.2, 1 mM DTT added immediately before use). After dialysis against 2 L of S30B for 3 h at 4 °C, the extract was recovered with a syringe and transferred to 1.5 mL Eppendorf tubes, followed by centrifugation at $12,000$ g for 10 min. Finally, the extract was distributed into 30 mL aliquots, flash frozen in liquid nitrogen and stored at −80 °C.

The accompanying buffer solution constitutes 42% of the final reaction volume. After addition of extract, DNA, and additives, the final reaction contained 50 mM Hepes pH 8, 1.5 mM ATP (Roth, #HN35.3) and GTP (Roth, #K056.4), 0.9 mM CTP (Roth, #K057.4) and UTP (Roth, #K055.3), 0.2 $\mathrm{mg\,mL^{-1}}$ tRNA (Merck, #10109541001), 26 mM coenzyme A (Merck, #C3144-10MG), 0.33 mM $\mathrm{NAD^+}$ (Merck, #481911), 0.75 mM cAMP (Merck, #A9501-1G), 68 µM folinic acid (Merck, #47612-250MG), 1 mM spermidine (Merck, #S2626-1G), and 30 mM 3-PGA (Merck, #P8877-1G) as an energy source. The final concentrations of screened components were [270]: 4 mM Mg-glutamate (Merck, #49605-250G), 60 mM K-glutamate (Merck, #49601-100G), 1.5 mM of each amino acid except leucine (Biozym, #BR1401801), 1.25 mM leucine, 2.5% (w/v) PEG-8000 (Merck, #89510-250G-F), and $0-1$ mM DTT (Merck, #10197777001). The buffer was also flash frozen in liquid nitrogen and stored at −80 °C.

2.3.8 Preparation of PITC-Agarose Microgel Spheres

PITC-agarose was prepared as described in reference [37]. The reaction conditions were adapted from FITC protein labelling protocols, while the workup was inspired from the preparation of FITC labelled dextrans [271]. Typically, 1 mmol (306 mg) of super low melting (SLM) agarose (Carl Roth, #HP45.1) was dissolved in 10 mL of carbonate buffer at pH 10 by heating to 90 °C in a Falcon tube. The solution was cooled to room temperature and 0.25 mmol (22.9 µL) of propargyl-isothiocyanate (PITC, ChemPur Feinchemikalien und Forschungsbedarf GmbH, #FL-9569-1) dissolved in 5 mL of DMSO were added. After vortexing, the reaction was carried out for 12 hours at RT on a rotator.

The product was precipitated 4 times with 2 volumes (20 mL) of cold (−20 °C) iso-propanol, washed with 1 volume of methanol, dried for 2 h in a desiccator and re-dissolved in 1 volume of ddH$_2$O. The first precipitation was carried out with 30 mL iso-propanol and the final precipitate was dried overnight. The light brown flakes were stable for storage at 4 °C for at least 9 months. Typical recovery was $\approx 70-90\%$. The degree of substitution was assessed by reacting 50 µM of the fluorogenic dye 3-azido-7-hydroxycoumarine (Carl Roth, #7811.1) with 5x dilutions of PITC-agarose in unmodified SLM agarose, in 1x PBS containing 1 mM TCEP (Roth, #HN95.1), 0.1 mM THPTA (Roth, #7822.2), 1 mM CuSO$_4$ (Roth, #P024.1). For quantification a reference curve was obtained using propargyl-NHS (Merck, #764221) as an alkyne standard, because of its higher solubility in water compared to PITC.

Copper-catalyzed click reactions were performed in aqueous solutions under standard conditions [272]. Typically, ≈ 10 mg of PITC-agarose were dissolved in ddH$_2$O to a concentration of 2% (w/v), assisted by heating at 80 °C for 15 min and repeated vortexing and stirring, using round-bottom shaped 2 mL tubes. 100 mg (100 µL) of the viscous solution were pipetted to a fresh tube using a scale to improve precision. The final coupling reaction contained 1% PITC-agarose, $0.05-10$ µM azide-modified DNA, 1 mM TCEP, 0.1 mM THPTA, and 1 mM CuSO$_4$ in a total volume of 200 µL of 1x PBS. The solution was mixed thoroughly and the reaction was allowed to proceed for at least 2 h at 45 °C on a shaker at 350 rpm. Optionally, a small amount of pre-labeled PITC-agarose (prepared in the same way as DNA-modified PITC-agarose) was added after DNA coupling, for barcoding. Depending on the reporter molecule, we used 3-azido-7-hydroxycoumarine (Carl Roth, #7811.1), 5-TAMRA-PEG3-azide (baseclick GmbH, #BCFA-037), or Cyanine-5-azide (baceclick GmbH, #BCFA-082). Uncoupled DNA and the reaction buffer were removed through excessive washing after emulsion polymerization. We note that for comparison of transcription rates at different gel concentrations, the click reaction was performed in a master mix containing 1% PITC-agarose, which was then mixed with 4% PITC-agarose and 1x PBS to ensure an equal DNA concentration for all samples.

After the click reaction the agarose solution was briefly heated at 80 °C to ensure that the agarose is melted completely. The solution was centrifuged for 5 min at 16,000 g to remove any remaining debris that would interfere with droplet production. Then, 180 µL of the supernatant were emulsified using a microfluidic flow focusing device (orifice width: 25 µL, height: 40 µL), as described in Section 2.2.6. The ≈ 30 µL large functionalized microgel spheres were collected and gelled overnight at 4 °C. The emulsion was broken after removal of excess oil by addition of 1 volume (≈ 200 µL) of perfluoro-octanol (PFO, Alfa Aesar, #B20156.09) and 1 volume of 1x PBS, followed by gentle shaking. The mixture was spun down and the aqueous supernatant containing the microgel was transferred to a fresh tube. Finally, the microgel was washed 4 times with 1 mL PBS and 4 times with 1 mL nfH$_2$O by centrifugation for 1 min at 2,000 g. The microgel was stored at 4 °C for up to one month.

3 Artificial Gel-based Organelles for Spatial Organization of Cell-free Gene Expression Reactions

The contents of this chapter are based on the publication:

L. Aufinger and F. C. Simmel. "Artificial Gel-based Organelles for Spatial Organization of Cell-free Gene Expression Reactions". *Angewandte Chemie International Edition* 57.52 (2018), pp. 17245–17248. DOI: 10.1002/anie.201809374.

As a step towards programmably communicating ACs we developed and characterized open hydrogel compartments and utilized them to spatially insulate gene expression reactions. This technology can be the basis for engineering higher level organization within AC communities. For instance, separating the nodes of a GRN into ACs with specialized functions could allow for the realization of larger GRNs by reuse of components, or initially identical ACs could differentiate based on their local neighborhood and global signals.

To this end, we exploit the properties of hydrogels as a solid-liquid hybrid material (Section 1.2.4). Hydrogels provide an aqueous environment for IVT and CFPE reactions that allows for the immobilization of individual reaction components, while simultaneously allowing for communication between ACs via biomacromolecules such as DNA, RNA, or proteins. Biomacromolecules are attractive signaling molecules, due to their direct interfacing with GRNs, as well as their inherent sequence programmability, enabling orthogonal information channels with larger capacity [36]. This enables spatial GRN architectures in which individual ACs receive, process, and send signals to exchange information to execute a molecular program.

Here, we developed a protocol for covalent immobilization of linear DNA molecules coding for IVT and CFPE reactions (PURE) inside of agarose microgel spheres (Section 3.1). Stochastically encapsulating microgel organelles with different functions into emulsion droplets allows to study varying stoichiometries simultaneously (Section 3.2). In Section 3.3 we first show that RNA signals transcribed from transcription organelles can be specifically targeted to capture organelles via hybridization to a sequence-encoded address tag. Next, we demonstrate that mRNA regulated by toehold switch riboregulators [103] can be captured and activated by their cognate DNA triggers immobilized in translation organelles (Section 3.4). When mRNA is produced from transcription organelles, this mechanism enables the spatial separation of transcription and translation processes.

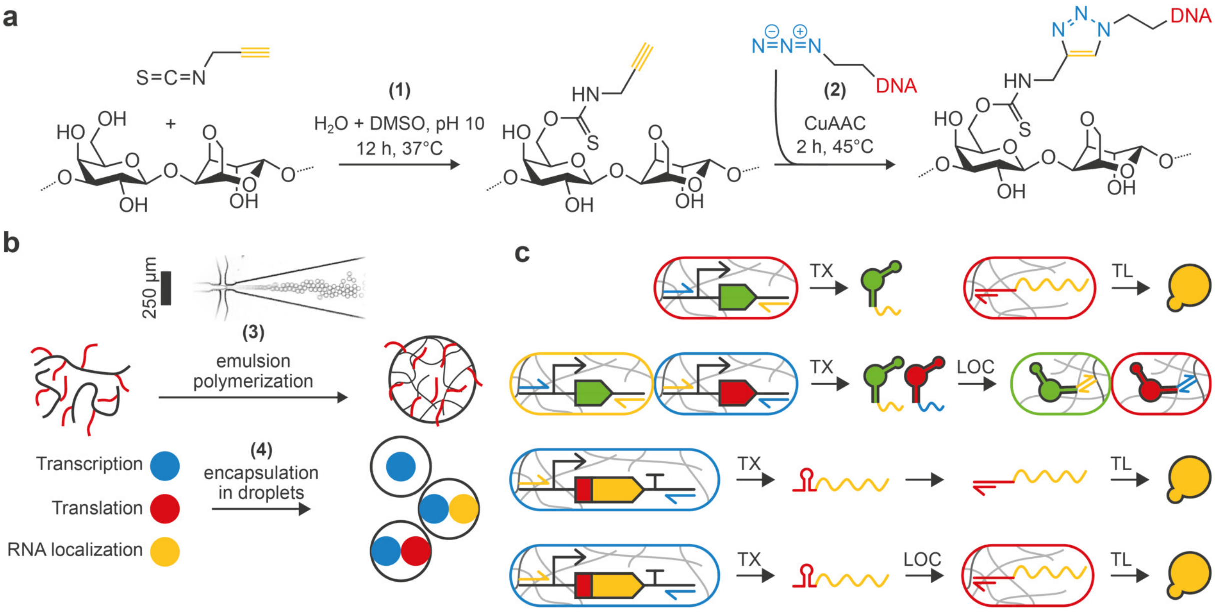


Figure 3.1: Workflow for production of DNA-functionalized microgel spheres. a) Agarose is first modified with PITC at mild conditions (1) and subsequently coupled to azide-modified DNA (2). **b)** The product is used to generate monodisperse microspheres via emulsion polymerization using a microfluidic flow-focusing device (3). Microgels with different functions are added to the reaction mixture and co-encapsulated into larger emulsion droplets (4). **c)** Combinations of microgels analyzed in this study. From top left to bottom right: Transcription, translation, transcription and orthogonal RNA localization, transcription in gel with translation in solution, and transcription in gel with RNA localization and translation in gel. A toehold riboswitch is used to ensure that translation is localized in the gel.

3.1 Agarose Microgels

For CFPE reactions to occur inside the hydrogel, it is crucial that the mesh size is large enough to permit the diffusion of large biomolecules such as ribosomes (25 − 30 nm [273]). Additionally, the hydrogel should not chemically interfere with the CFPE reaction, and exhibit minimum interactions, e.g. electrostatic interactions, with biomolecules as this could significantly affect partitioning unspecifically. For these reasons, we chose agarose as the hydrogel matrix, as it is biocompatible, charge neutral at physiological pH, and has a high polymer fiber stiffness resulting in large pore sizes in the order of 100 nm [199]. In the following, we first describe the DNA-functionalization (Section 3.1.1) and microgel synthesis (Section 3.1.2), and then characterize the mesh size (Section 3.1.3).

3.1.1 Functionalization

One reason to choose agarose is its biochemical intertness, however, this property makes chemical functionalization more challenging. Coupling reactions to sugars typically target hydroxyl groups which react as weak nucleophiles at pH above the typical pKa 10−14 [274, pp.35], while hydrolysis of the agarose chains occurs above pH 11−12. Further, suitable coupling reactions are typically most efficient when reagents are present at mM concentrations. Because DNA is typically only available at nM to µM concentrations, and also contains hydroxyl groups as well as even more nucleophilic amines, we chose to not attempt to directly couple DNA to agarose in a one-step reaction.

We therefore first chemically modify the agarose with a functional group and then couple the DNA in a separate step. A good candidate for such a functional group are azides or alkynes, so that the DNA can be coupled via copper catalyzed azide-alkyne cycloaddition 'click chemistry' which is defined by its high specificity and efficiency [275]. Possible coupling reactions with hydroxyls that occur at sufficiently mild conditions and were previously employed for agarose or other polysaccharides such as dextran are for instance epoxies [276], Mitsunobu substitutions [277], and isothiocyanates [271]. While the former two strategies generate amines, which can then be linked to another functional group via EDC-NHS chemistry in an additional step, we used the latter approach, as it allows for introduction of alkyne groups in a single step using the commercially available propargyl-isothiocyanate (PITC).

The experimental workflow shown in Figure 3.1 was inspired from Thiele et al. [196]. First, we produced a large batch (1 g, sufficient for $\approx$ 100 microgels) of the chemically functionalized hydrogel (Figure 3.1a, Section 2.3.8). To improve the solubility of PITC, we added varying fractions of DMSO to the solvent (Figure 3.2a). The degree of substitution, i.e. the amount of terminal alkynes per agarose subunit, was estimated to be $\approx 0.1\%$, which equals 25 µM of alkynes when dissolved at 10 $\mathrm{mg\,mL^{-1}}$. Second, we covalently coupled oligonucleotides or PCR amplified DNA to small amounts ($\approx$ 100 µL at 10 $\mathrm{mg\,mL^{-1}}$) of the PITC-agarose (Figure 3.1a, Section 2.3.8).

3.1.2 Microgel Synthesis

The resulting DNA hydrogel is then mixed with dye labeled PITC-agarose, encapsulated in emulsion droplets of diameter $d \approx 30$ µm employing microfluidic flow focusing (Section 2.2.6, Section 2.3.8) and gelled overnight at 4 °C (Figure 3.1b). Following this procedure we manufactured different types of microgel spheres with DNA encoded functions such as sequence-orthogonal capturing of nucleic acids, IVT of functional RNA, and CFPE. To probe these functions, the recovered microgel spheres are added to either a T7 RNA polymerase based IVT mix or the PURExpress system [55] and loaded into larger emulsion droplets ($d \approx 50-60$ µm) for observation via time-lapse epifluorescence microscopy (Figure 3.1b). Using this procedure we tested several combinations of organelles, as shown in Figure 3.1c.

3.1.3 Mesh Size Estimation

We next estimated the mesh size by studying the partitioning of fluorescein and FITC-dextrans with hydrodynamic diameters from $1-54$ nm (Figure 3.2b,c). To this end, we incubated microgels of different gel densities with the fluorescent probes and imaged them with a confocal microscope. The partitioning coefficient can then be computed from the fluorescence intensity in the gel f_{gel} and solution f_{sol} as

$$K = \frac{c_{gel}}{c_{sol}} = \frac{f_{gel}}{f_{sol}} \; . \tag{3.1}$$

As shown in Figure 3.2b, K decreases for increasing gel density, as expected [199]. For 32 nm dextrans and 1% gel density we get $K \approx 0.8$ suggesting that ribosomes should indeed be able to diffuse into the gel. If we use the definition of mesh size as the particle size at which $K = 0.5$ (Section 1.2.4), we get $\xi > 50$ nm for the full range of gel densities.

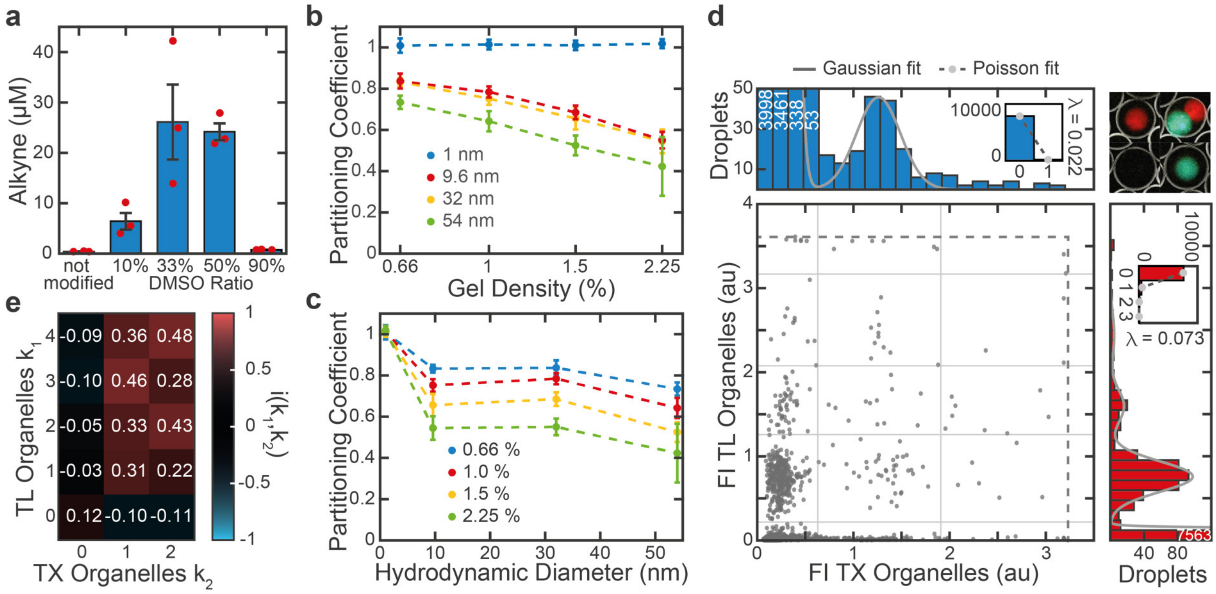


Figure 3.2: Characterization of functionalized microgel and encapsulation. a) Agarose functionalization at 1% (w/w) was measured for varying DMSO ratios in the reaction. Error bars represent standard deviations of 3 technical replicates. **b), c)** Partitioning experiments to estimate the mesh size of functionalized PITC-agarose. **d)** Scatter plot illustrating the encapsulation of two distinct organelles into droplets. FI: Fluorescence Intensity. **e)** Normalized pointwise mutual information for different combinations of encapsulated organelles indicating that co-encapsulations are favored by the encapsulation procedure, compared to an ideal Poisson process.

Interestingly, Figure 3.2b,c shows that the dextran molecules saturate at $K < 0.8$, even for the lowest gel density, while for fluorescein $K = 1$ within the measurement uncertainty. The latter is expected, as fluorescein is simply too small to ‘see’ the pores, it only ‘sees’ the fibers which occupy a fraction of < 0.03 of the volume. However, the fact that the dextrans do not saturate at a K close to 1, suggests a different type of interaction. One possible explanation is unspecific repulsive interaction between the dextrans and agarose fibers, but as neither of them are charged this seems unlikely. Another possible explanation is that the pore size distribution is biphasic, i.e. there is a significant population of small pores $\xi < 10$ nm that can only be penetrated by fluorescein, and a population of large pores $\xi \gtrsim 100$ nm. This would agree with the suprafiber structure of agarose [278]. Agarose chains bundle via hydrophilic interactions to form suprafibres that are very stiff and therefore form the population of large pores. The suprafibers in turn can occupy a larger volume than the individual fibres alone and hence contain small pores to fill the void volume. Additional support for the plausibility of this thesis is that we used low melting agarose, which is produced from native agarose by hydroxyethylation, which results in less hydrophilic interactions, a lower melting temperature and less dense suprafibers [278].

3.2 Encapsulation Efficiency

For the encapsulation of the gel organelles into emulsion droplets, we simply mixed different types of gel organelles with reaction solution. We then used a flow-focusing chip with a larger junction (40 µm instead of 25 µm) to load them into larger droplets at random. If the number of encapsulated particles per droplet k is an *independent* random variable, the encapsulation corresponds to a Poisson process. Then, in accordance with previous studies [279], the probability of encapsulating k gel organelles into a droplet follows a Poisson distribution (e.g. insets in Figure 3.2d)

$$p_{Poisson}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \,, \tag{3.2}$$

where λ equals the average number of particles per droplet. In the following, we first statistically analyze the encapsulation process (Section 3.2.1) and then estimate the probabilities to obtain a particular composition with different encapsulation mechanisms (Section 3.2.2).

3.2.1 Statistical Independence

We can use our data to make an argument about the statistical independence of the encapsulated gel organelles to identify possible deviations from the idealized process. In Figure 3.2d we have encapsulated a mixture of k_1 red and k_2 blue gel organelles. For the observer the organelles are therefore easily distinguishable, but behave identically for the encapsulation mechanism. Hence, if we can show that k_1 and k_2 are independent random variables, also k is an independent random variable and the same applies if the organelles have multiple identities.

If k_1 and k_2 are independent random variables, the joint pdf factorizes $p(k_1,k_2) = p(k_1)\cdot p(k_2)$. We here consider two measures for statistical independence. First, the mutual information of the two variables

$$I(k_1;k_2) = \sum_{k_1,k_2} p(k_1,k_2)\log_2 \frac{p(k_1,k_2)}{p(k_1)\cdot p(k_2)} \tag{3.3}$$

for our sample data is 0.023 bit, while 0 bit correspond to complete statistical independence. Second, the normalized point-wise mutual information

$$i(k_1;k_2) = \frac{-1}{\log_2 p(k_1,k_2)} \cdot \log_2 \frac{p(k_1,k_2)}{p(k_1)\cdot p(k_2)} \,. \tag{3.4}$$

is $i \approx 0$ for single and $i \approx 0.3$ for multiple encapsulations (Figure 3.2e), where $i(k_1;k_2) = 0$ signifies statistical independence and $i(k_1;k_2) = \pm 1$ indicates positive, or negative dependence.

In summary, this suggests that assuming statistical independence is a good approximation of the encapsulation process. This assumption should hold as long as the fraction of the volume in a droplet that is occupied by gel organelles is small, so that gel organelles do not sterically interfere. If the gel organelles have a constant diameter $d \approx 25$ µm and their concentration $C = \lambda V_{droplet}^{-1}$ is kept constant as well, this means that to optimize for co-encapsulation of multiple organelles, the droplet volume should be adjusted accordingly. For instance, to encapsulate two distinct organelles, we use droplets with a diameter of $D \approx 50$ µm, while for 4 organelles, we use $D \approx 100$ µm. The fact that co-encapsulations are actually over-represented compared to the ideal independent case may be related to our observation that particles frequently tend to queue in the microfluidic channel prior to encapsulation, which aids co-encapsulation.

3.2.2 Multiple particles with Identity

We now determine the multivariate probability function for the encapsulation of multiple particles with identity. Assuming independence, the joint probability for n independent stochastic variables k_i is

$$p(k_1, ..., k_n; \lambda_1, ..., \lambda_n) = \prod_{i=1}^{n} p_{Poisson}(k_i; \lambda_i) \; . \tag{3.5}$$

For maximization of the probability of a particular composition $\{k_i\}$ the optimum choice is $\lambda_i = k_i$. In this case, the probability of co-encapsulating exactly one of each of four distinct particles as in Figure 3.4a is $p_{Poisson}(1,1,1,1;1,1,1,1) = p_{Poisson}(1,1)^4 = e^{-4} \approx 0.018 = 1.8\%$.

For applications it may be desirable, however, to produce droplets, which predominantly contain organelles with a specific stoichiometry. To optimize the composition, we can first realize that with some calculus

$$p(k_1, ..., k_n; \lambda_1, ..., \lambda_n) = p_{Poisson}(k; \lambda) \cdot p_{Multinomial}(k_1, ..., k_n; k, p_1, ..., p_n) \; . \tag{3.6}$$

where $p_{Multinomial}(k_1, ..., k_n; k, p_1, ..., p_n) = k!(k_1! \cdot ... \cdot k_n!)^{-1} p_1^{k_1} \cdot ... \cdot p_n^{k_n}$ is the multinomial probability distribution with $p_i = \lambda_i \lambda^{-1}$, $k = \sum_i k_i$ and $\lambda = \sum_i \lambda_i$ for $i = 1, ..., n$. This is intuitive as $p_{Poisson}(k; \lambda)$ represents the probability of encapsulating k particles regardless of identity and $p_{Multinomial}(k_1, ..., k_n; k, p_1, ..., p_n)$ is the probability of obtaining a particular composition for a given k.

Now, organelles could be encapsulated with an approach based on close-packed particle encapsulation as described by Abate et al. [279], which allows to load droplets with a uniform number of particles. In this case, the probability distribution would be described by the multinomial term only, and the fraction of droplets with the desired composition would increase to $p_{Multinomial}(1,1,1,1; 4, 1/4, 1/4, 1/4, 1/4) = 4!/4^4 \approx 0.094 = 9.4\%$. Even higher yields of specific compositions may be possible with custom techniques, e.g. by sequentially loading droplets with one organelle at a time [179].

3.3 Transcription and RNA Localization

We now proceeded to characterize transcription (Section 3.3.1) and capture (Section 3.3.2) of fluorescent RNA aptamers from immobilized genelets in an IVT reaction.

3.3.1 Transcription

We first studied transcription from gel organelles equipped with DNA templates encoding for the fluorescent dBroccoli (dB) aptamer [280] (Figure 3.3a,b). Accumulation of the reporter RNA can be observed only in droplets containing organelles (Figure 3.3c,d). Image analysis (Section 2.2.6, [37])

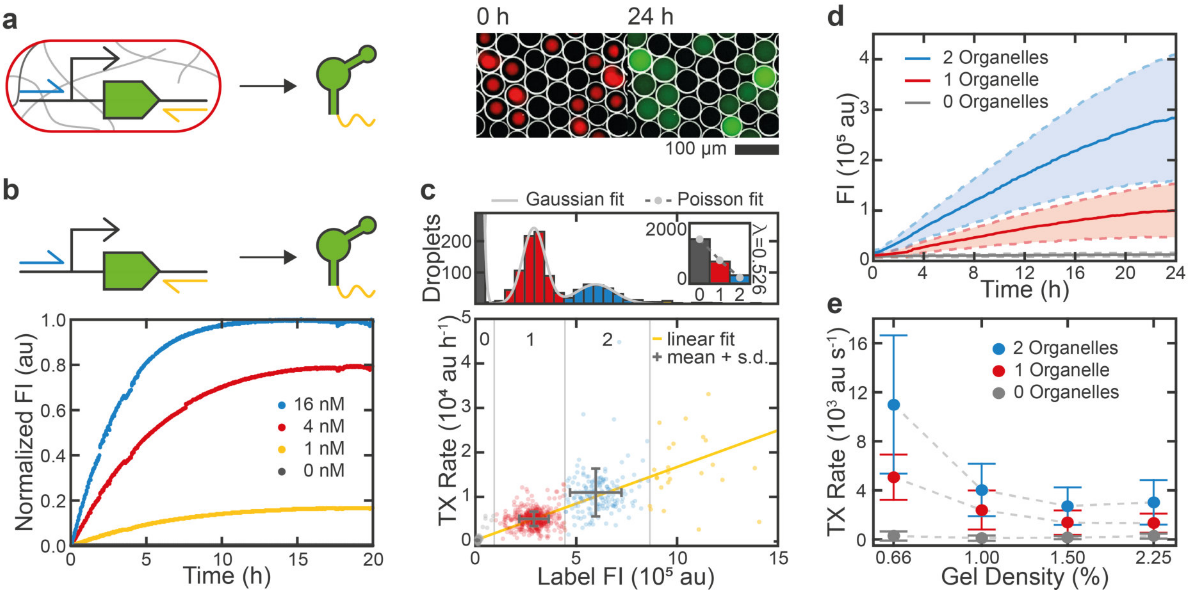


Figure 3.3: Transcription a) Schematic and fluorescence false color, inverted bright field (BF) overlay images of droplets containing 1-3 Cy5 labelled transcription organelles (red) that produce dB RNA (green) which distributes in solution. **b)** In bulk experiments the initial transcription rate increases with increasing reporter template concentration. **c)** Droplets were clustered by fitting a sum of Gaussians to the distribution of the label intensity (top). The number of organelles per droplet roughly follows a Poisson distribution (inset). The rate is proportional to the number of sender spheres. **d)** Average time traces showing the total dB fluorescence corresponding to the populations in (c). **e)** Transcription rate increases with decreasing gel density. Data shown in (c) and (d) corresponds to 0.66% gel density. All error bars and shaded areas represent standard deviations across a droplet population.

of droplets with different organelle stoichiometry reveals that the transcription rate depends linearly on organelle content (Figure 3.3c). The reaction is found to proceed over 24 hours (Figure 3.3d), which is considerably longer than in typical bulk experiments (Figure 3.3b). As partitioning of T7 RNA polymerase into the gel is expected to depend on the gel mesh size, we produced organelles with varying gel densities (Figure 3.3e). Indeed, transcription was faster for lower gel densities, indicating that transcription occurs inside the organelles.

3.3.2 Transcription and Capture

For the realization of spatially distributed reaction networks, it is desirable to target the RNA products to different locations in a programmable manner. To this end, we extended the aptamers with a 20 bp long linker sequence that is complementary to a single stranded DNA capture sequence immobilized in specialized receiver organelles. In Figure 3.4a we tested the simultaneous transcription of the fluorescent dB and the malachite green (MG) aptamer [281] from separate organelles and their subsequent sorting into dedicated ‘capture organelles’. As designed, the localization of reporter RNA species systematically depends on the stoichiometry of the organelles (Figure 3.4b).

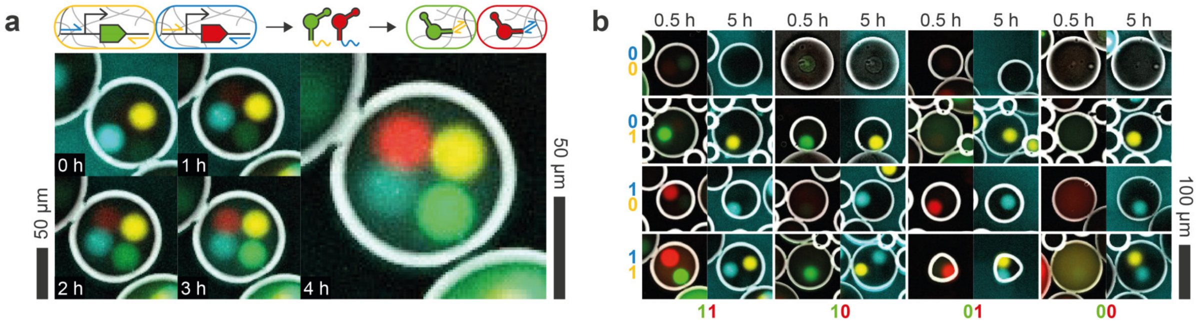


Figure 3.4: Transcription and RNA capturing. a) Fluorescence overlay time series showing the transcription of RNA from dB (yellow) and MG (cyan) transcription organelles and orthogonal RNA localization in dB (green) and MG (red) capture organelles. **b)** Overview of all binary combinations of organelles with color scheme as illustrated in (a). Fluorescence channels for transcription organelles (cyan, yellow) are shown separately from the RNA reporter channels (red, green) for clarity. In presence of both the transcription organelle and the corresponding capture organelle, the reporter RNA is localized, while in absence of the capture organelle, the reporter spreads across the entire droplet volume. Depending on the stoichiometry, the reporter RNA may exceed the capacity of the capture organelles, so that the excess reporter distributes in the solution.

3.4 Spatially Separated Transcription and Translation

Finally, we realized ACs with specialized transcription and translation organelles by implementing the individual components step by step. First, we tested the toehold switch mechanism in a bulk PURE reaction (Section 3.4.1). Then, we individually verified the functionality of transcription (Section 3.4.2) and translation organelles (Section 3.4.3) and finally combined them (Section 3.4.4).

3.4.1 Toeholdswitch Circuit in Solution

As shown in Figure 3.5a, we first verified that the toehold switch mechanism is functional in a PURE reaction. The on/ off ratio of translation in the presence/ absence of trigger is $\frac{\alpha+\alpha_0}{\alpha_0} \approx 30$.

3.4.2 Transcription in Gel and Translation in Solution

Next, we tested the case when the transcription reaction is confined in the gel, but translation occurs in the surrounding solution (Figure 3.5b-d). In contrast to the IVT reaction, translation activity ceased after 3-4 hours, which is typical for bulk CFPE reactions [20, 54] (Figure 3.5a). Clustering of droplets with the same organelle content shows a linear correlation of organelle number and expression level as for the transcription experiments (Figure 3.5d).

3.4.3 Translation in Gel

We also investigated cell-free protein expression in gel organelles by capturing mRNA molecules encoding mVenus as a reporter protein (Figure 3.5e,f). To ensure that translation can only occur when the mRNA is localized in the gel, we added a toehold switch riboregulator to the 5’ untranslated region of the mRNA, which suppresses translation in the absence of appropriate RNA or DNA trigger

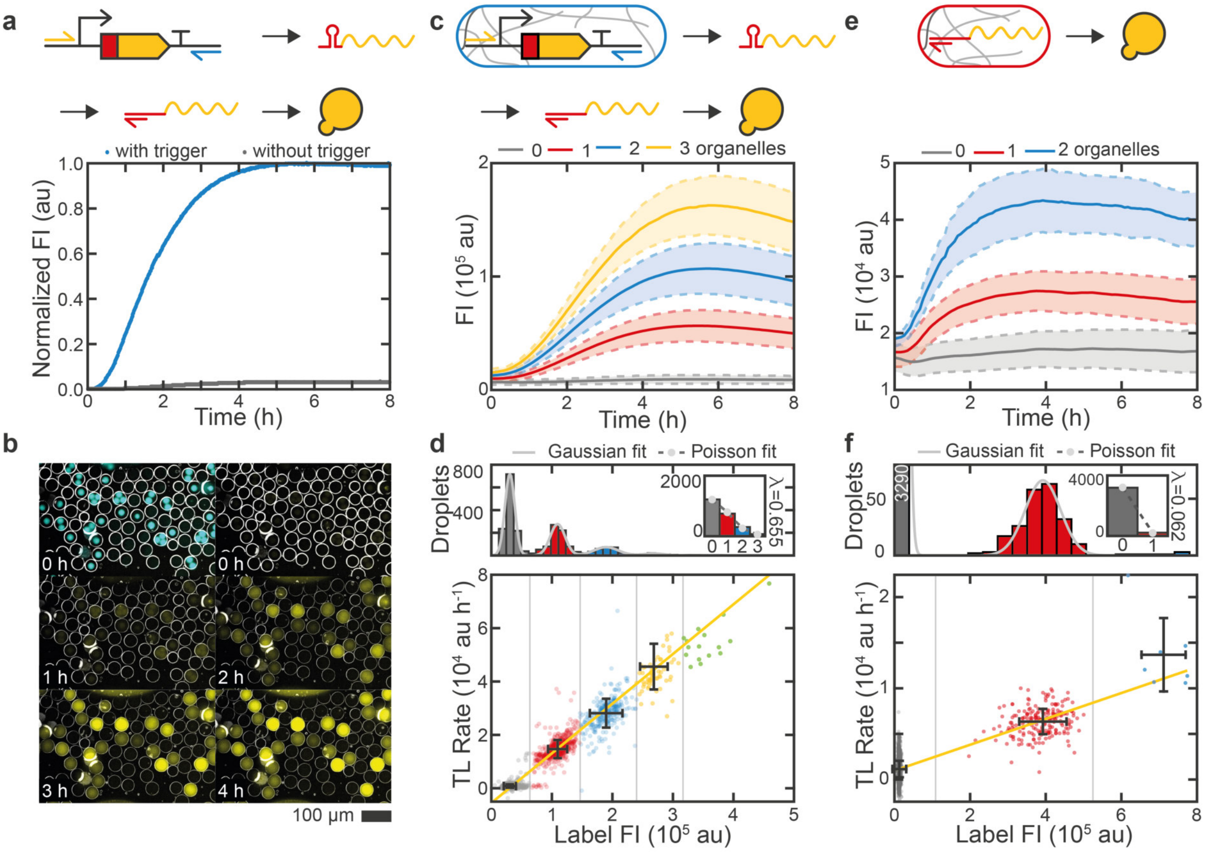


Figure 3.5: Translation in gel organelles. **a)** Bulk translation shows that translation is repressed ≈ 30-fold by the toehold switch in absence of trigger. **b)** Inverted BF, fluorescence overlay time series images, corresponding to the data shown in (c) and (d) showing transcription organelles (cyan) and expression of the reporter protein (yellow) in solution at the time indicated. **c)** Average time traces showing the total fluorescence in droplets containing a varying number of translation organelles. **d)** Clustering based on label intensity. All error bars and shaded areas represent standard deviations. **e,f)** As in (c,d), but for translation directly from the gel.

molecules [103]. We immobilized the corresponding trigger DNA in the gel, which served both to localize the mRNA and to activate its translation. Here, the apparent expression rate was ≈ 2-fold lower than with transcription organelles, which is consistent with the lower reaction volume of the organelle compared to the surrounding droplet.

3.4.4 Transcription in Gel and Translation in Gel

We finally combined transcription and translation organelles in one compartment. As illustrated in Figure 3.6a, mRNA is first transcribed in the transcription organelle, from which it diffuses into the surrounding solution. Blocked by a toehold switch, it is translationally inactive. Once the mRNA diffuses into a translation organelle it is captured and activated by the trigger DNA, resulting in protein expression. Consistent with this process, we observe a significantly increased reporter signal only in droplets containing both gel organelles (Figure 3.6b). The small background signals observed in droplets with only transcription or translation organelles are likely caused by the leakiness of the toehold switch and mRNA transcribed prior to compartmentalization into droplets, respectively. We also note that the coefficient of variation in the expression rate in droplets with both transcription

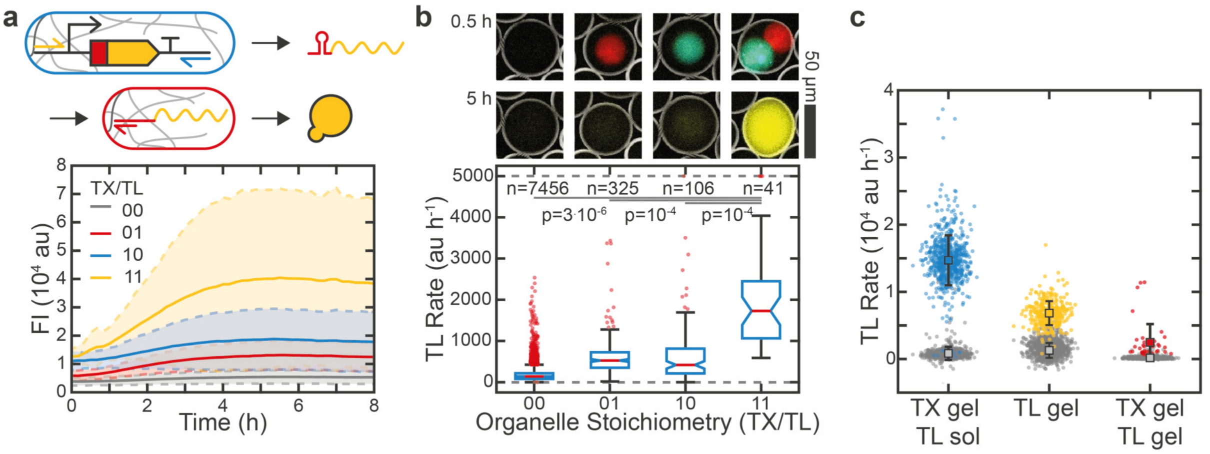


Figure 3.6: Spatial decoupling of transcription and translation. a) Average fluorescence time traces for all binary combinations of transcription (TX) and translation (TL) organelles. **b)** Inverted BF, fluorescence overlay images showing transcription (cyan) and translation organelles (red) after 0.5 h and the reporter protein (yellow) after 5 h. Boxplots illustrating the reporter expression rate for different organelle content. Boxes are 25% and 75% percentiles, notches are 95% confidence intervals, outliers were clipped at 5000 $\mathrm{au\,h^{-1}}$. **c)** Comparison of absolute fluorescence values from different experiments with different restriction on the TXTL process indicates a decreased efficiency for the more confined reactions.

and translation organelles is higher compared to experiments with a single type of organelles (Figure 3.6c). This may be attributed to small variations in organelle size and noise propagation, and highlights the need for feedback mechanisms in more complicated spatial reaction networks.

Quantitatively, the expression rate is 6-fold reduced compared to the case, where mRNA is transcribed in the gel and translated in solution and 3-fold reduced compared to translation in the gel (Figure 3.6c). It has been previously reported that separating the transcription and translation process reduces the efficiency of CFPE reactions [282]. However, already the use of toehold switches decouples the two processes – ribosomes cannot directly bind to the nascent mRNA - and therefore the reduction in expression rate in our case is probably simply due to the reduced reaction volume.

3.5 Summary

We have developed hydrogel-based reaction compartments as organelles for artificial cells, which facilitate localization and spatial organization of transcription and translation reactions and also allow sequence-addressable exchange and sorting of RNA signals or cargoes. Using different compositions of functionally distinct gel-based organelles allows programming of biochemical reactions networks at a higher hierarchical level, i.e., at the level of functional modules [31]. An important aspect of gel-based organelles is their nature as open reaction compartments, which allows exchange of reactants and therefore the sustained maintenance of chemical non-equilibrium conditions. This should be of great interest for the generation of more sophisticated reaction dynamics and spatially extended dynamical systems.

4 Synthetic Cell-based Materials Extract Positional Information from Morphogen Gradients

The contents of this chapter are based on the publication:

A. Dupin*, **L. Aufinger***, I. Styazhkin, F. Rothfischer, B. Kaufmann, S. Schwarz, N. Galensowske, H. Clausen-Schaumann, and F. C. Simmel. “Synthetic cell-based materials extract positional information from morphogen gradients”. *bioRxiv* (preprint). DOI: 10.1101/2021.04.25.441320.

I wish to acknowledge the contributions by Aurore Dupin, who conceptualized the project and performed the droplet experiments and simulations. I analyzed the droplet experiments and calculated positional information. We both interpreted and visualized the results and conceptualized the simulations.

Self-differentiating synthetic cell-based materials are an exciting platform to explore bio-inspired engineering paradigms. In analogy to systems in developmental biology, form and structure may be generated by simple genetic programs, rather than by assembly of parts. To gauge the reliability of such processes, it is crucial to quantify the robustness of the differentiation mechanism. Here, we applied the theoretical frameworks of positional information (PI) and positional error (PE) that were developed in the context of developmental biology [253, 254, 259, 283, 284] (Section 1.3.2), to the engineering problem of robust patterning of an artificial tissue. As an experimental platform, we utilize linear arrays of droplets that form droplet interface bilayers (DIBs) [29, 169] (Section 1.2.2).

As shown in Figure 4.1a, a typical sample consists of one sender droplet that acts as a morphogen source and 5 identical receiver droplets that contain a TXTL-based gene regulatory network (GRN). We evaluated three GRN topologies, based on the transcription factors LacI and TetR that mutually repress each other, with variations in the feedback interactions and degradation rate (Figure 4.1b). As the morphogen (IPTG) diffuses into the receiver droplets, a gradient establishes, that is interpreted by the GRN, leading to differentiation of the receiver droplets (Figure 4.1c,d). By analyzing sufficiently large numbers of samples ($N = 9-37$), we are able to estimate probability distributions (Figure 4.1e) to quantify the variability in the gene expression patterns in terms of PI (Figure 4.1f) and PE (Figure 4.1g).

In the following, we first meticulously describe how PI and PE can be estimated from experimental data, which requires a correction of measurement uncertainties (Section 4.1), a sufficient number of samples N, and a correction of statistical biases (Section 4.2). We then present our key results that, supported by modeling, explain how PI is generated in our system (Section 4.3) and discuss these by comparing them to the gap gene network of *Drosophila* (Section 4.4).

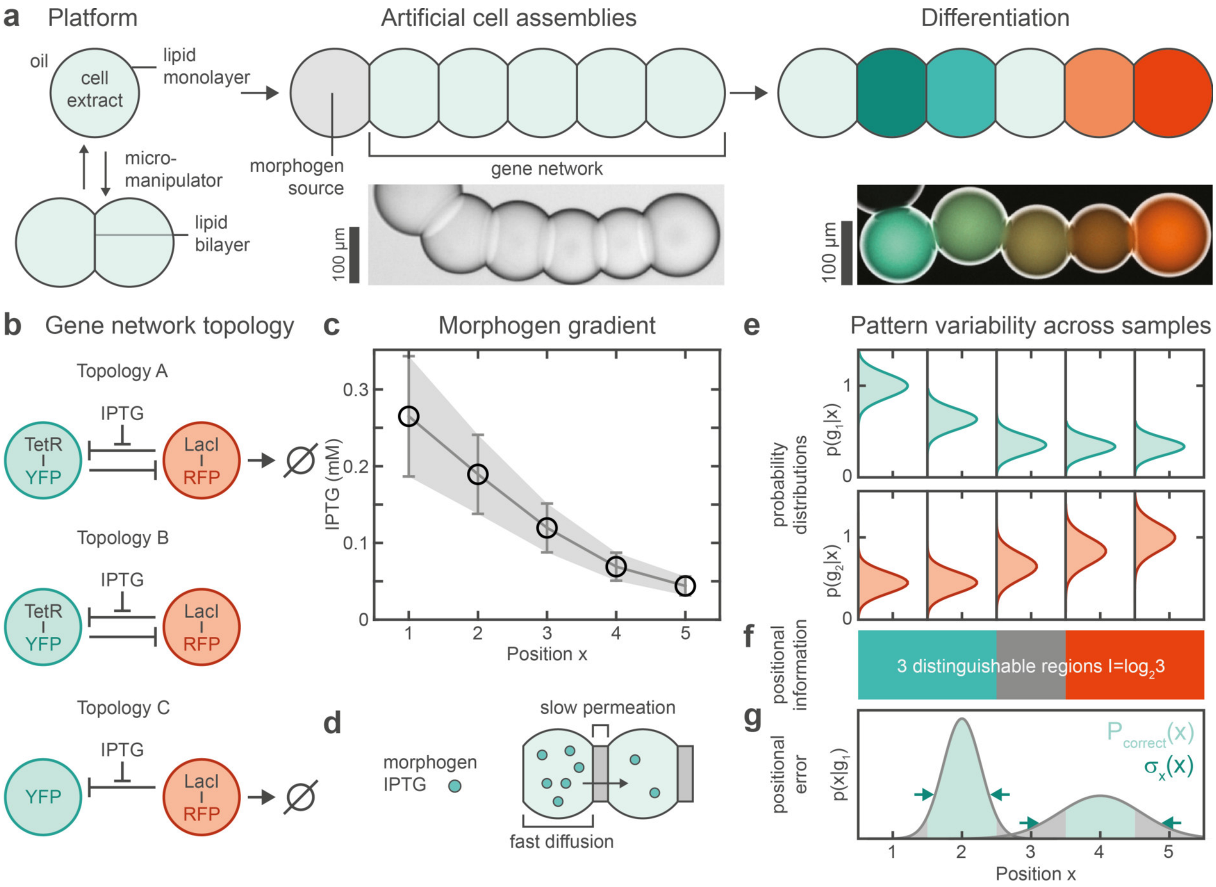


Figure 4.1: Investigation of morphogen-based differentiation in synthetic cell assemblies. The response of gene networks to morphogen gradients is used to quantify positional information and positional error, to assess the accuracy of patterning in this system. **a)** Artificial cell assemblies consist of nanoliter-sized water-in-oil droplets where oil-dispersed lipids form a monolayer at the water-oil interface. Droplets are then brought in contact using a micromanipulator to form bilayer interfaces. Assemblies consist of a source droplet containing the morphogen (IPTG) and identical receivers containing the GRN. The morphogen diffuses from its source and forms a dynamic gradient along the main axis of the system as indicated in (c). The induction of the GRN in a gradient of the morphogen results in a differentiation of the artificial cells (microscopy image: red and blue represent two fluorescent reporters of gene activity). **b)** Three GRN topologies investigated in this work that involve the transcriptional repressors LacI (fluorescence co-expressed with RFP) and TetR (co-expressed with YFP). Topology A: mutual repression with degradation of one of the repressors, Topology B: mutual repression without degradation, Topology C: Repression of YFP by LacI-RFP with degradation of the repressor. **c)** Simulated morphogen gradient for 1 mM IPTG in the sender droplet. The variation in the gradient was simulated by assuming a variability in droplet volumes (Section 4.3.3). **d)** Proposed diffusion mechanism according to Dupin and Simmel [29]. **e)** To estimate the positional error and positional information the expression of the tagged repressors is measured in each droplet with high accuracy for a large collection of droplet assemblies, resulting in position dependent distributions of gene expression levels. **f)** The capability of a circuit to differentiate distinct regions in the presence of noise can be measured by the positional information I. **g)** The local uncertainty of a position estimate based on a measurement of the gene expression levels can be quantified by the positional error $\sigma_x(x)$ and the deduced probability that a position estimate is correct $P_{corr}(x)$.

4.1 Minimization of Measurement Uncertainty

As outlined in reference [38], we captured data in the form of time-lapse epifluorescence videos (Section 2.1.2), which we analyzed to better understand which parameters influence the quality of the developing gradients. In this context, the positional error $\sigma_x(x)$ at position x within a droplet network can be understood as the precision with which x can be estimated by the genetic circuit it contains, typically by 'measuring' the local concentration of potentially multiple gene expression gradients g_i, where $i = 1, ..., J$. The positional information $I(x; g_i)$ is measured in units of bits and can be interpreted as the base-2 logarithm of the number of regions X that can be globally distinguished by the genetic circuit, i.e. $I(x; g_i) = \log_2 X$.

In our experimental setup we use a maximum of $J = 2$ gene expression gradients, TetR-YFP (mVenus) and LacI-RFP (mScarlet-I). In a DIB network we can assume that diffusive mixing within one droplet is fast (protein: $t = l^2/2D \approx (200\ \mu\mathrm{m})^2/(2 \cdot 100\ \mu\mathrm{m}^2\mathrm{s}^{-1}) = 200\ \mathrm{s}$), while the diffusion of IPTG through the network is limited by the permeation through the bilayer (Section 1.2.2). We therefore assume a discretized space with one unit length defined as one droplet (i.e. one bilayer to pass) and a total length of $X = 5$. This means that the theoretical maximum for the positional information in our setup is $I(x; g_i) = \log_2 5 \approx 2.32$ bit. Similarly, a positional error of $\sigma_x(2) = 0.5$ droplets, means that with the given gene expression level the probability that the droplet is at position 2 is about 68%. Both, PE and PI, quantify how well positions can be inferred by a given genetic circuit with respect to variations in the gene expression gradients. Hence, we first need to estimate these variations, represented as a joint probability distribution function (pdf) $p(g_i, x)$, by measuring g_i for N distinct samples.

Experimentally, we can infer the gene expression level in a droplet by extracting an appropriate measure of reporter fluorescence $f \propto g$. The experimentally observed variability σ_f, however, is a combination of the actual gene expression variability σ_g and measurement uncertainties σ_m, i.e., $\sigma_f^2 = \sigma_g^2 + \sigma_m^2$. As discussed in Section 4.3.3, the intrinsic variability σ_g is either the actual noise that the genetic circuit 'sees' as a consequence of variations in droplet or bilayer size resulting in variations in the IPTG input gradient, or output noise produced by variations in gene expression strength across droplets. Note, that we do not consider stochastic effects due to the large number of molecules in one droplet ($1\mathrm{nM} \cdot 4\mathrm{nL} \cdot N_A \approx 10^6$). In contrast, the measurement uncertainty σ_m is related to image acquisition and analysis, i.e. it is extrinsic variability that is irrelevant to the genetic network or this analysis. In order to obtain a meaningful estimate of σ_g, it is therefore necessary to reduce the measurement noise as much as possible so that $\sigma_m/\sigma_f \ll 1$ (typically at least < 0.2) and hence $\sigma_g \approx \sigma_f$ [253]. In the following, we first use a calibration data set to estimate σ_m and test alternative methods to reduce σ_m from an initial 35% to about $2-3\%$. Compared to a typical σ_f of about $15-30\%$ observed in developmental experiments this gives at least $\sigma_m/\sigma_f < 0.2$.

The overall optimized image processing workflow is summarized in Section 4.1.1. The three key aspects that reduced measurement uncertainty – data normalization by reference dye, 'trimmed' segmentation, and robust background and flatfield correction – are explained in detail in Section 4.1.2, Section 4.1.3, and Section 4.1.4, respectively. All automated image processing routines and schemes for calculation of PE and PI were implemented in a modular fashion using either FIJI or MATLAB[1].

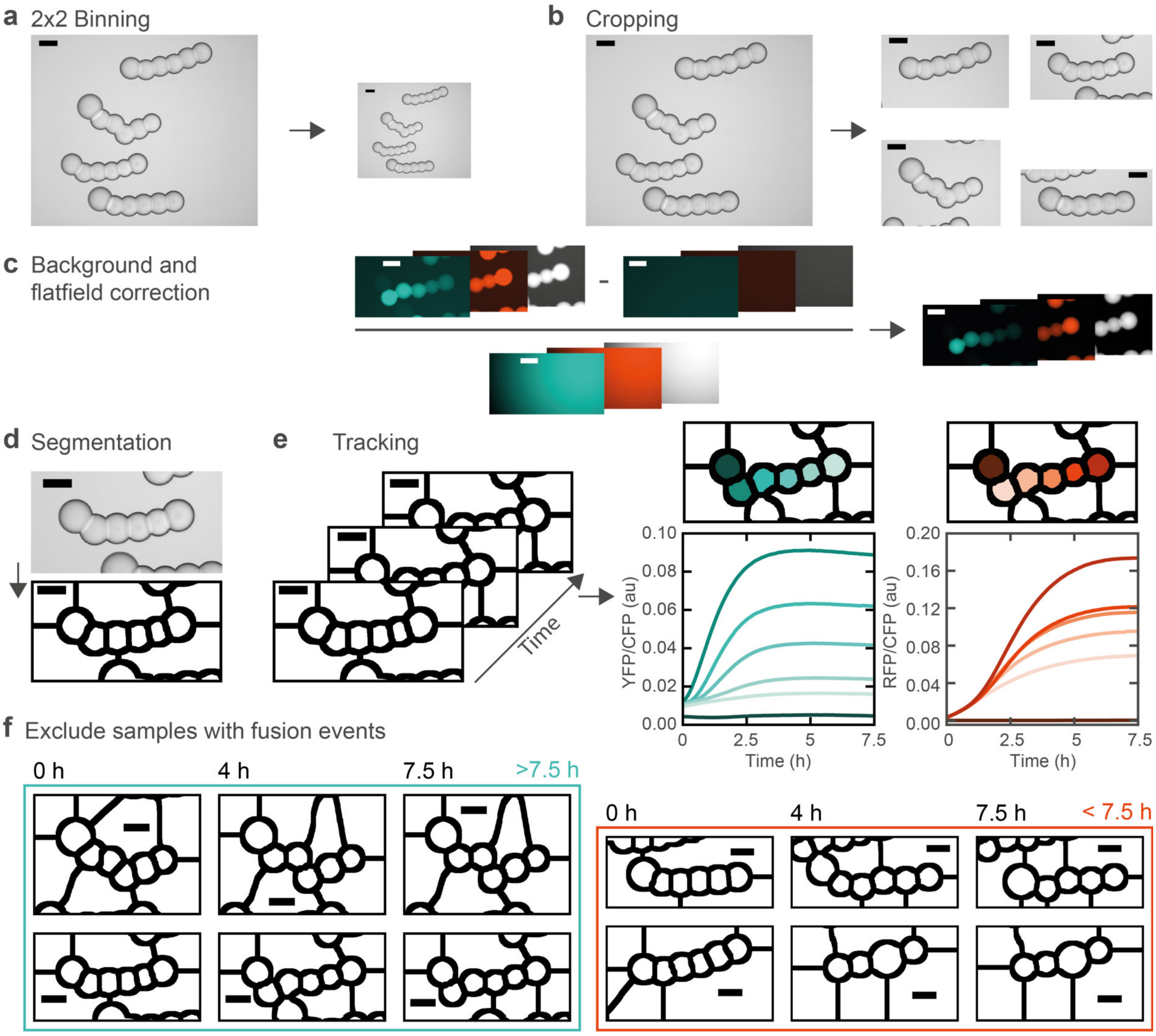


Figure 4.2: Overview of the optimized image processing routines. a) 2 × 2 binning to reduce memory and computation time. **b)** Sample cropping to reduce computation time and facilitate allocation of samples. **c)** Background and flatfield correction. For visualization the brightness of corrected, uncorrected, and background images was scaled between 0 and ≈ 5 times the average background intensity. **d)** Segmentation as described in Section 4.1.4. **e)** Droplets were tracked over time based on the segmented masks to extract the fluorescence time traces for the droplets for the YFP, RFP, and CFP channel. YFP and RFP time traces were then divided by the measured CFP intensities to yield normalized time traces. **f)** Schematic illustrating the filtering of samples with fusion events earlier than 7.5 h after starting image acquisition. All scale bars, 100 µm.

4.1.1 Image Processing Workflow

We precede the image processing with a conversion from .nd2 to .tiff, combined with a 2x2 binning (Figure 4.2a) of all images to reduce the size and computation time of any downstream processes, without affecting accuracy. In the next step we use an automated cropping routine to create stacks with a single sample (Figure 4.2b). This improves speed and facilitates allocation by assigning unique identifiers to the cropped stacks.

---

[1] https://github.com/laufluIu/EMB, https://doi.org/10.5281/zenodo.5763562

Next, we run the background and flatfield correction (Figure 4.2c, Section 4.1.3), as well as the segmentation (Figure 4.2d, Section 4.1.4). To obtain fluorescence time traces, we use a MATLAB tracking plugin[2] developed in-house by Korbinian Kapsner [269] (Figure 4.2e), after which we manually selected the segmented areas that represent droplets for all samples that had 6 droplets in the first image. The sender droplet was automatically identified by having the lowest ratio of YFP to reference fluorescence in the first frame and receiver droplets were indexed according to their Euclidean distance from the sender.

Finally, we exclude samples in which two droplets fuse at any time before 7.5 h (after which fluorescence signals saturate) after starting the image acquisition (Figure 4.2f). The occurrence of fusion events corrupts time traces due to a change in the number of samples N and is easily detected by the ending of the tracked time trace of the corresponding droplet. An overview of all samples is provided in reference [38]. Excluded samples that did not consist of 6 droplets in the first frame or that underwent a fusion event comprise about 45% of all samples across all datasets.

4.1.2 Normalization

To optimize our image processing routines, we first need to quantify σ_m. We therefore acquired a calibration data set consisting of images of several droplets with varying size that were filled with a constant concentration of reporter protein (YFP and RFP). Hence, the variation in protein concentration $\sigma_g \approx 0$ and the measured variability is an estimate of the measurement uncertainty $\sigma_f \approx \sigma_m$. The fluorescence measure f_i for the segmented droplet i that we extract from our video microscopy data must be normalized such that it is proportional to protein concentration. In Figure 4.3 we compare two normalization procedures we called 'normalization by area' and 'normalization by reference'. For each normalization scheme, we compared different image correction (Section 4.1.3) and segmentation procedures (Section 4.1.4) to determine which combination yields the lowest σ_m.

For normalization by area we first segment the droplets based on brightfield images (to avoid bias due to differences in fluorescence brightness) and then normalize the sum of fluorescence intensity F_i within a segmented area by its size A_i. The functional relation between F_i and A_i depends on the optical setup, as well as on the droplet geometry. We hence fit the calibration function

$$\mathcal{F}(A) = a \cdot A^b \ . \tag{4.1}$$

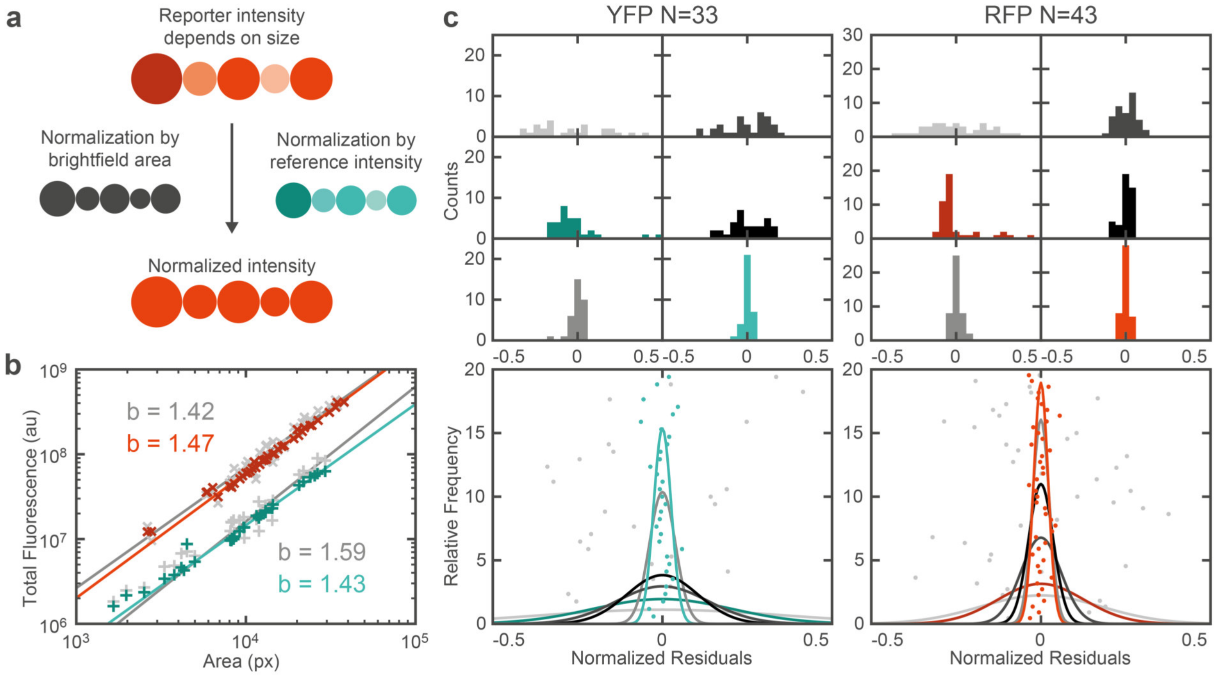
We would typically expect that $1 < b < 3/2$, where $b = 1$ corresponds to a disk-like geometry (high NA objective) and $b = 3/2$ corresponds to a sphere-like geometry (low NA). Experimentally, we find values close to $b = 1.5$ (Figure 4.3b), consistent with a low NA of 0.45. We can then calculate our observable by dividing F_i by the calibrated function

$$f_i = \frac{F_i}{\mathcal{F}(A_i)} \propto F_i \cdot A_i^{-b} \ . \tag{4.2}$$

For normalization by reference, we add an internal reference dye (CFP, mTurquoise2) that is present in all droplets with a constant concentration. Because the reference fluorescence density $r_i = const.$ and hence $r_i \propto R_i \cdot A_i^{-b} = const.$, the total reference fluorescence R_i is directly proportional to A_i^b

---

[2] https://github.com/kkapsner/Matlab



Normalization	Segmentation	BG	FFC	CV YFP	CV RFP	Color
Area	Precise	-	-	35.7%	17.8%	⬜
Area	Precise	+	+	20.4%	12.7%	◩
Reference	Precise	+	+	3.8%	2.5%	■
Reference	Trimmed	-	-	13.5%	5.9%	■
Reference	Trimmed	+	-	10.4%	3.6%	■
Reference	Trimmed	+	+	2.6%	2.1%	◩

Figure 4.3: Measurement uncertainty of the calibration data set comparing several normalization, segmentation and image correction procedures. a) Schematic illustrating normalization by area vs. normalization by reference dye. The intensity sum in each segmented area was either divided by a calibrated measure of the droplets size (b), or by the intensity sum of a reference dye. **b)** Fits to determine the exponent b required for the normalization by area (Equation (4.1)). Shown are fits for both YFP and RFP channels, each for the original images (gray) and for background (BG) and flatfield corrected (FFC) images (colored). The normalized residuals (Equation (4.4)) of the fits are shown in (c). **c)** Comparison of the measurement uncertainty for different image analysis procedures as indicated by the legend (table). Top: Histograms with bin width 0.02. Middle: Inferred Gaussian pdfs. Individual data points correspond to the normalized residuals. Bottom: Legend listing the coefficient of variation (CV) for the different procedures.

(assuming b is similar for reporter and reference fluorescence). Hence, we can simply normalize by dividing F_i by R_i

$$f_i \propto \frac{F_i \cdot A_i^{-b}}{R_i \cdot A_i^{-b}} = \frac{F_i}{R_i} \,. \tag{4.3}$$

Importantly, the geometric factor vanishes, suggesting that this method is insensitive with respect to segmentation uncertainties.

For both normalizations, the relative measurement uncertainty is then given by the coefficient of variation (CV) of f

$$\sigma_m \approx \frac{\sigma_f}{\mu_f} = \frac{\sqrt{\langle (f_i - \mu_f)^2 \rangle}}{\mu_f} = \sqrt{\left\langle \left(\frac{f_i}{\mu_f} - 1 \right)^2 \right\rangle}\,, \tag{4.4}$$

where μ_f and $\langle ... \rangle$ denote the mean and we call the quantity $\epsilon_i = f_i/\mu_f - 1$ the 'normalized residuals' (Figure 4.3c).

We note that our calibration data set includes duplicates from droplets that were imaged in different overlapping microscope positions, but in different relative positions within the field of view. The dominant source of uncertainty is related to the flatfield illumination which depends on the relative position of imaged droplets (Figure 4.4b,c). We hence purposely kept these duplicates, as keeping them should not bias the σ_m estimates, while systematically removing them might.

4.1.3 Illumination Uncertainties

The two major sources of illumination uncertainty are the (varying) background intensity and the typically Gaussian illumination profile characteristic to epifluorescence microscopy [285]. It is common to account for these uncertainties with a flatfield correction (Figure 4.4)

$$\hat{I}_c = \frac{\hat{I} - \hat{B}}{\hat{F}}\ . \tag{4.5}$$

The corrected image $\hat{I}_c$ is obtained from the original image $\hat{I}$ by first subtracting the background image $\hat{B}$ and then dividing pixel-wisely by the normalized flatfield image $\hat{F}$.

Besides camera dark noise and background autofluorescence (due to the oil bath), one main component of background light in our setup stems from the sample chamber boundaries. These consist of O-rings grafted onto a glass slide using rather strongly autofluorescent epoxide glue, which causes scattering light. The background is therefore uneven and has to be determined individually for each sample. Therefore, we initially employed the FIJI 'Subtract Background' routine which uses a rolling ball algorithm (typical radius of 200 pixels). However, we found that this method was not sufficiently robust and quantitative for our purpose and therefore implemented an alternative routine based on robust surface fitting of a 2^{nd} degree 2D polynomial in MATLAB (Figure 4.4a). To improve the robustness, the areas that contained droplets were excluded from the fit. Surface fitting is computationally expensive and had to be performed on about 3 channels × 193 time points × 50 sampels × 10 data sets ≈ 300,000 images. We therefore first scaled the images down by an empirically determined factor of 0.3 for fitting and used the fitted parameters to obtain the original-sized background image $\hat{B}$.

For flatfield correction (Figure 4.4b,c) we first recorded 10 flatfield images for each channel using a plain chamber filled with the respective fluorophore and the corresponding darkfield images for the same illumination settings. To mitigate any inhomogeneities, we next took the median intensity of each pixel, subtracted the darkfield image, fitted a 2D Gaussian surface and normalized by dividing by the mean intensity. This yields the normalized flatfield images $\hat{F}$. Note, that already the normalization by reference dye partly amounts to a flatfield correction, but does not account for differences

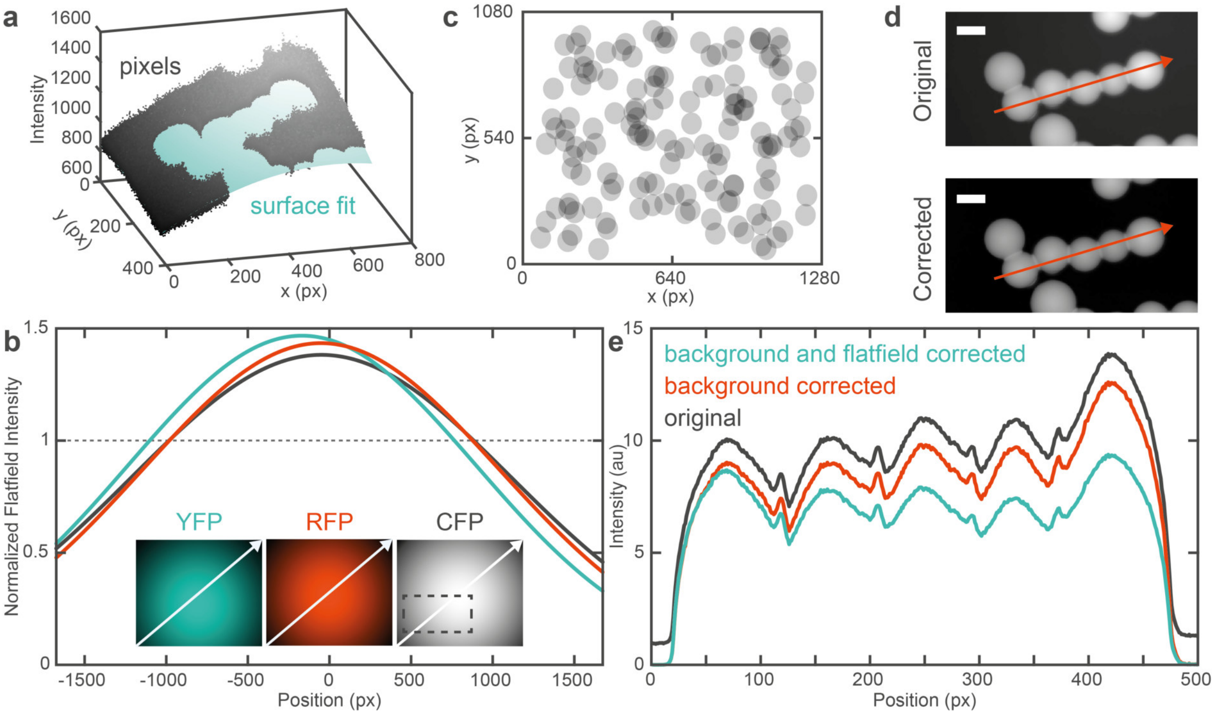


Figure 4.4: Illumination uncertainties. **a)** 3D representation illustrating the surface fit used to estimate the background for background subtraction. Data points correspond to the intensity of individual pixels and are shown as gray dots. Data points in areas belonging to the sample were excluded from the fit to improve robustness and are not displayed for clarity. Shown is the CFP channel of sample 19.3 from topology A, at 1 mM IPTG, after 4 hours (frame 49). **b)** Flatfield images for the YFP (cyan), RFP (red) and CFP (gray) channel. The arrows indicate the direction of the profile plots. The box indicates the position of the representative images in (a,d). **c)** Distribution of droplets in the calibration data set within the field of view (FOV). The approximately uniform distribution over the FOV highlights the necessity and effectiveness of flatfield correction. **d)** Original and background corrected image (CFP channel) of the same sample as in (a). Scale bars, 100 µm. **e)** Profiles along the long axis of the sample in (c) illustrating the effect of background and flatfield correction.

among different channels. Additional flatfield correction therefore decreases the measurement uncertainty significantly (Figure 4.3d,e).

4.1.4 Segmentation Uncertainties

Segmentation of droplets based on brightfield images (Figure 4.5 a) is commonly achieved by contrast thresholding, or edge detection. We here developed two custom routines, tailored specifically to the two normalization procedures, to thoroughly eradicate potential segmentation related measurement uncertainties. For normalization by area an ‘accurate segmentation’ procedure is of critical importance, while we found that for normalization by reference dye a technically simpler ‘trimmed segmentation’ routine shows a superior performance.

Investigating a typical brightfield image of a droplet assembly (Figure 4.5a) we can easily discern the outer edges as a clear contrast minimum. The DIBs, however, are less well-defined and may consist of multiple contrast minima or maxima. It is therefore challenging to find a thresholding based method that accurately locates the outer edges and DIBs simultaneously. We hence decided to first generate images with accurate outer edges and DIBs, separately.

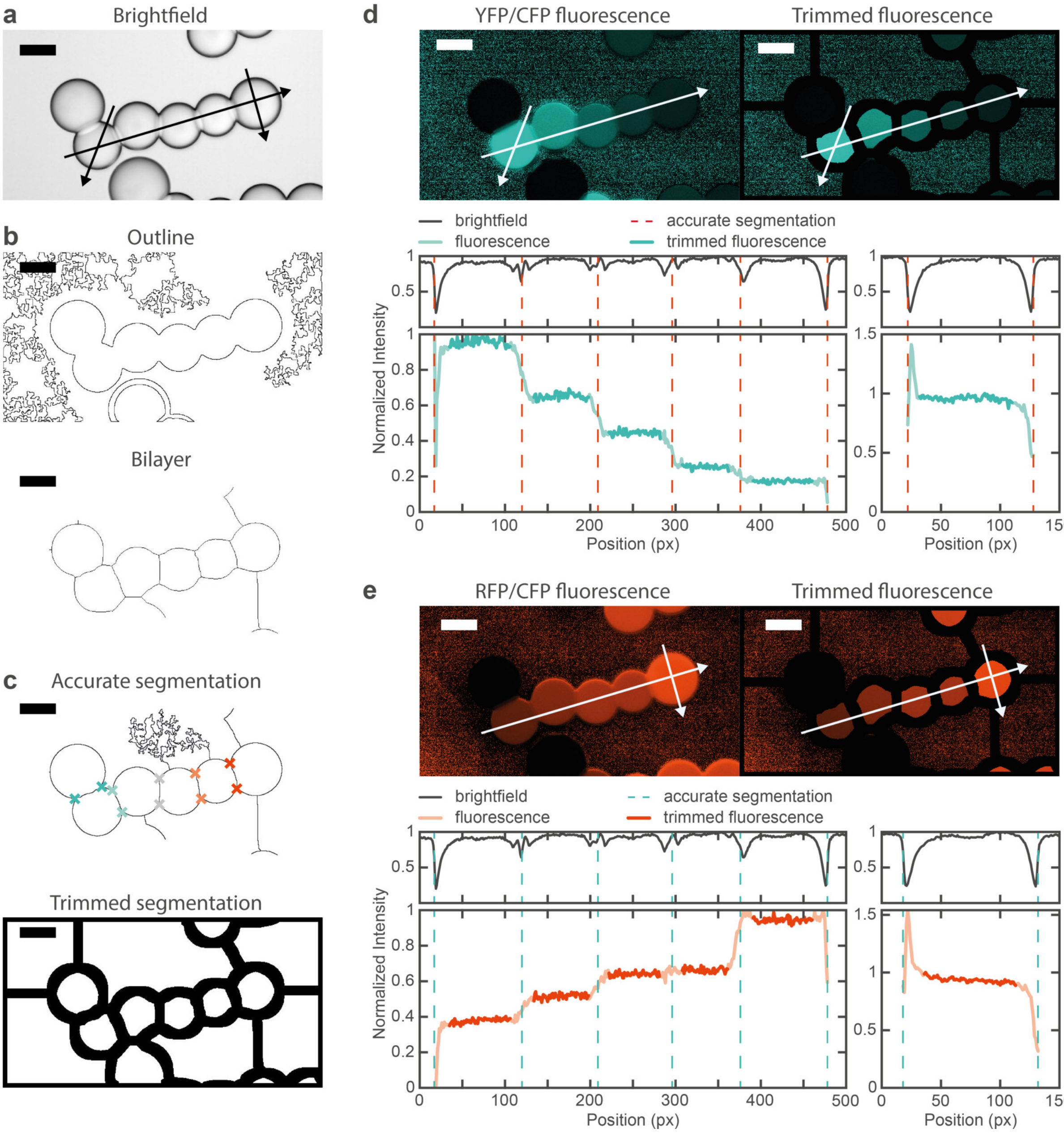


Figure 4.5: Segmentation uncertainties. a) In the brightfield image, outer edges appear as defined contrast minima, while positions of the DIBs are less clearly defined. **b)** Binary images generated by thresholding the BF image (a) to find the positions of the outline and DIB. **c)** Segmentation images generated from (b) using binary operations. Accurate segmentation was used with normalization by area, trimmed segmentation with normalization by reference dye. **d,e)** Segmentation uncertainties for the (d) YFP and (e) RFP channel. The top images show a corrected fluorescence image divided by the reference image and the same image multiplied with the trimmed segmentation image (c). Arrows indicate positions of the profile plots below. As indicated by the relatively constant fluorescence within one droplet, using trimmed segmentation avoids measurement uncertainties due to the undefined bilayer positions, pixel shifts and size estimates. All images are of sample 19.3, 4 h, topology A, 1 mM IPTG. Scale bars, 100 µm.

The outline image in Figure 4.5b (top) was generated using the auto-threshold function (FIJI) with the method ‘mean’ combined with some filtering and binary operations. Note, the fringes around the assembly can easily be sorted out at a later stage. The bilayer image in Figure 4.5b (bottom) was generated using Sobel filtering (FIJI function 'find edges'), followed by a Gaussian filter (radius 20 pixels) to blur the image, auto-thresholding (Otsu method), binary operations and a watershed transformation (FIJI plugin 'Adjustable Watershed'[3], radius 5). This procedure robustly yields images where individual droplets are segmented roughly at the bilayer positions, but is inaccurate in finding the exact locations of the outer edges.

We therefore combined these two images using binary and logic operations to get an accurate segmentation regarding both, outer edges and the bilayer positions (Figure 4.5c, top). As indicated in the figure, we used these accurate segmentation images to estimate the variablilty of droplet volumes and bilayer areas in Figure 4.10. However, accurate segmentation combined with normalization by area has three fundamental weaknesses regarding measurement uncertainty. These are visualized in Figure 4.5d,e that shows fluorescence images divided by the reference image.

First, because the size of the segmented areas is used to calculate fluorescence, any uncertainty in determining the area (due to varying contrast etc.) propagates to our final fluorescence measure. A related problem is that due to the deformation caused by the bilayers, the exact geometry of each droplet differs, causing a deviation from the calibrated exponent b in Equation (4.1). Second, an ≈ 1 px shift between fluorescence channels is common in epifluorescence microscopy [285], and can prevent the accurate location of edges. In Figure 4.5d,e (right panel) this shows as a pronounced peak near the edges of the droplet. Third, as indicated by the gradually increasing fluorescence between droplets (Figure 4.5d,e, left panel), the DIB positions are not well defined in the fluorescence channels either. This may be caused by the bilayer not being parallel to the observation axis [171], or by refraction at the bilayer.

We hence use normalization by reference dye in combination with the 'trimmed segmentation' scheme (Figure 4.5c, bottom) to effectively avoid all these potential sources of uncertainty. A 'trimmed segmentation' image is simply generated by eroding the bilayer image (Figure 4.5b, bottom) 10 times to ensure that the segmented areas do not include any pixels close to the edge or a bilayer. Due to the normalization by reference dye, the normalized fluorescence within the segmented areas is approximately constant (Figure 4.5d,e). This property renders this method insensitive to the exact size and location of the segmented area and additionally partly corrects for illumination uncertainties.

4.2 Estimation of Positional Information and Positional Error from Data

We now describe the estimation of pdfs from our data (Section 4.2.1), calculation of PI (Section 4.2.2), and PE (Section 4.2.3) and discuss the underlying assumptions.

4.2.1 Estimation of Probability Distribution Functions

For each of our 10 data sets, the image processing yields $N = 9-37$ samples that are drawn from an underlying joint pdf $p(\{g_i\}, x, t)$ (Figure 4.6), which needs to be estimated before calculating PI and PE. One set of fluorescence time trace data can be represented as a 4-dimensional array consisting of N samples, $J = 2$ genes, $X = 5$ positions and $T = 91$ time points (Figure 4.6a). For consistency, the data is normalized to the maximum intensity across all x and t of the mean intensity over n for each

---

[3] https://imagejdocu.tudor.lu/doku.php?id=plugin:segmentation:adjustable_watershed:start

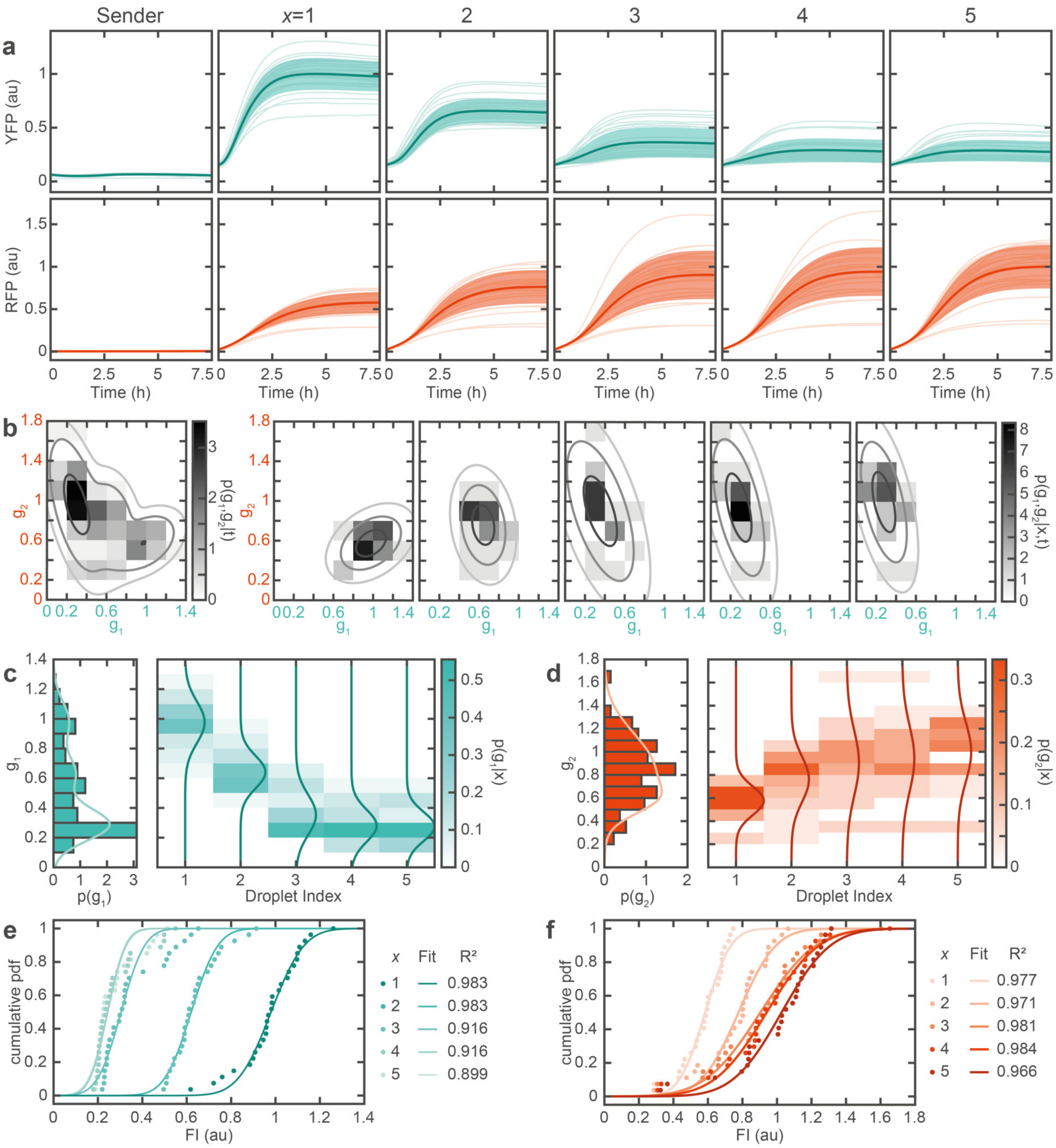


Figure 4.6: Representing fluorescence data by probability distribution functions. a) Individual fluorescence time traces (faint) with mean and standard deviation for the respective droplet. **b)** Joint marginal $p(g_1, g_2|t = 7.5\text{ h})$ and conditional pdf $p(g_1, g_2|x, t = 7.5\text{ h})$ represented as binned histograms ($B = 5$) as indicated by color. Overlaid are contour lines of Gaussian distributions inferred from the sample mean and standard deviation (SD) at 1,2, and 3 SDs. **c,d)** Single gene marginal $p(g)$ and conditional pdf $p(g|x)$ for gene g_1 and g_2, represented as a histogram ($B = 10$) as indicated by color. Overlaid are Gaussian distributions inferred from the mean and SD at the respective position. **e,f)** Cumulative pdf for $p(g|x)$. The cumulative Gaussian fits were performed to justify the Gaussian approximation used to calculate PE and PI.

gene i. Since positions and time points are uniformly distributed with marginal distributions $p_x(x) = 1/X$ and $p_t(t) = 1/T$, respectively, the joint conditional pdf is $p(\{g_i\}|x, t) = X \cdot T \cdot p(\{g_i\}, x, t)$.

As illustrated in Figure 4.6b, $p(\{g_i\}|x,t)$ can be estimated by binning the data or by assuming an underlying distribution. For a multivariate Gaussian distribution

$$p(\{g_i\}|x,t) = (2\pi)^{-J/2}|C(x,t)|^{-1/2}\exp\left[-\frac{1}{2}\sum_{i,j=1}^{J}(g_i-\bar{g}_i(x,t))[C^{-1}(x,t)]_{ij}(g_j-\bar{g}_j(x,t))\right], \quad (4.6)$$

the mean gene expression gradients $\bar{g}_i(x)$ and covariance matrix $C_{ij}(x)$ can be obtained from the fluorescence data as

$$\bar{g}_i(x,t) = \frac{\sum_{n=1}^{N} g_i^{(n)}(x,t)}{N}, \quad (4.7)$$

$$C_{ij}(x,t) = \frac{\sum_{n=1}^{N} g_i^{(n)}(x,t) g_j^{(n)}(x,t)}{N-1} - \bar{g}_i(x,t)\bar{g}_j(x,t)\,. \quad (4.8)$$

The same procedure, here defined in general for the multiple gene joint conditional pdf $p(\{g_i\}|x,t)$, also applies for the estimation of single gene conditional pdfs $p(g_i|x,t)$ (Figure 4.6c,d). While estimating PI from binned pdfs requires relatively large sample sizes to correct for binning bias (Section 4.7), the approximation of the underlying distribution by a Gaussian should be properly justified. Besides the verification schemes discussed in Section 4.7, additional support for assuming a Gaussian distribution can be gained by fitting a representative set of cumulative data (Figure 4.6e,f). Good fits are obtained with a cumulative Gaussian pdf $c(g|x) = \frac{1}{2}\mathrm{erf}\left(\frac{g-b}{a}+1\right)$, with deviations increasing in the droplets distant from the sender $x = 3, 4, 5$.

4.2.2 Calculation of Positional Information

In the following, we briefly introduce the formal definitions for PI and PE and discuss the methods presented by Tkačik et al. [253] regarding the calculation of PI and PE estimates from real data. Positional information is defined as the mutual information that the (joint) gene expression levels $\{g_i\}$ contain about position x and vice versa

$$I(x;\{g_i\}) = \int dx\, p_x(x) \int d^J\mathbf{g}\, p(\{g_i\}|x)\log_2\frac{p(\{g_i\}|x)}{p_g(\{g_i\})} = I(\{g_i\};x)\,. \quad (4.9)$$

Here $p_x(x) = 1/5$ and $p_g(\{g_i\}) = \int dx\, p_x(x)p(\{g_i\}|x)$ are the marginal distributions. Note, that for conciseness we here have omitted to explicitly write down the time dependence of PI and the pdfs, but all calculations can simply be performed for each given time point individually to obtain the temporal evolution of PI. Equation (4.9) can be rewritten as the difference of the Shannon entropy $S[p(y)] = -\int dy\, p(y)\log_2 p(y)$ of the marginal and conditional pdf

$$I(\{g_i\};x) = S[p_g(\{g_i\})] - \langle S[p(\{g_i\}|x)]\rangle_x\,. \quad (4.10)$$

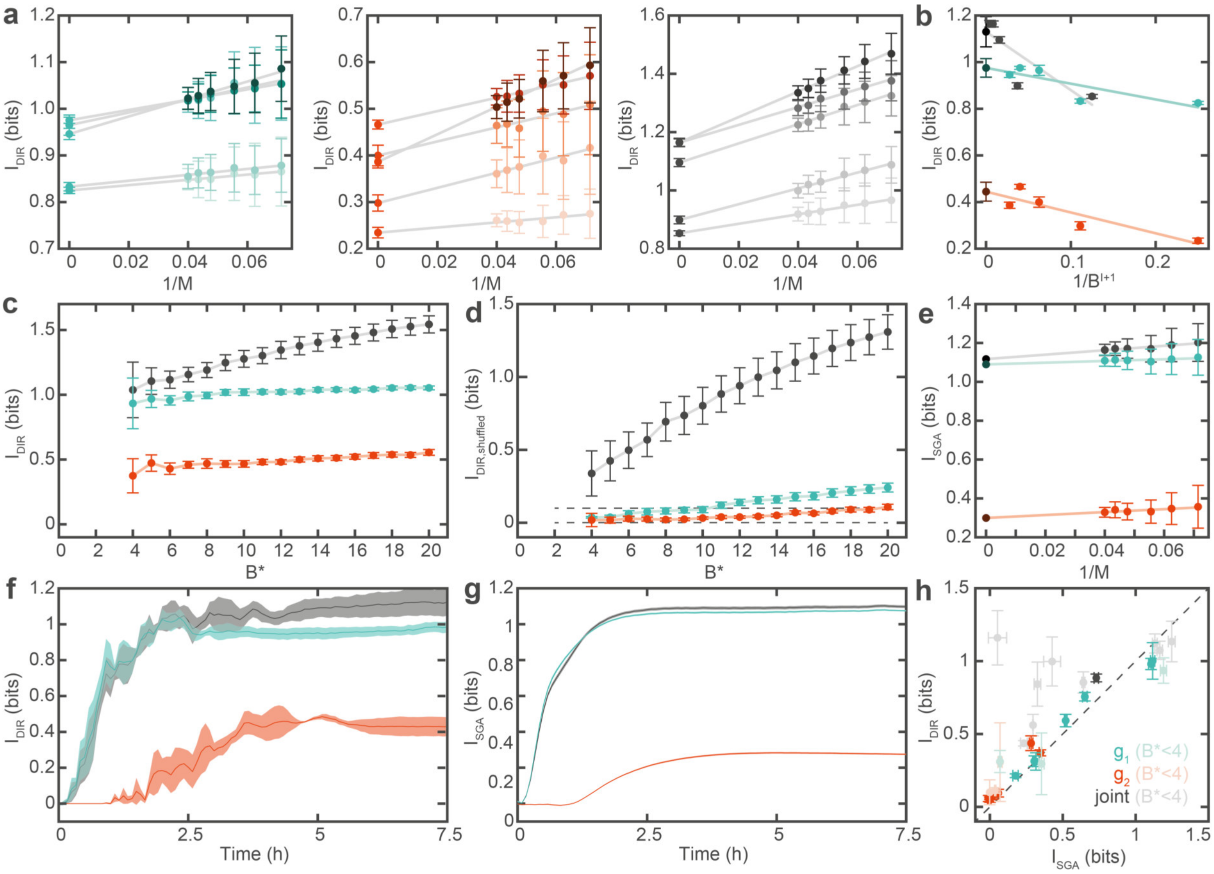


Figure 4.7: Positional Information. a,b) Extrapolation of positional information using the direct method (DIR), to (a) infinite sample size and (b) infinite bin number. **c)** Extrapolated DIR estimates for varying maximum bin numbers B^*. For single gene PI the estimates are approximately constant over the range tested, while for the joint PI the estimates systematically increase with increasing B^*. **d)** Extrapolated shuffled DIR estimates for varying B^*. Values above zero indicate a bias due to under-sampling. Horizontal dashed line indicates the tolerance of 0.1 bit. **e)** Extrapolation of naive SGA estimates to correct for the sample size bias. **f,g)** Extrapolated DIR and SGA estimates, respectively, over time. **h)** Comparison of DIR and SGA estimates for all data sets. Estimates with insufficient data for application of the DIR method ($B^* < 4$) are grayed out. (a-h) Error bars and shaded areas represent 68% confidence intervals of the extrapolated values (t-corrected standard errors). Error bars on non-extrapolated values in (a) represent the SD of 100 randomly sampled datasets. In (g) error bars are in the order of the line width. All data shown here is from topology A at 1 mM and is representative for all data sets.

The first term in Equation (4.10) is called 'total entropy', the second term 'noise entropy'. Total entropy measures the range of gene expression levels that are available to the droplets, while noise entropy quantifies the 'loss' of information due to the variability in the gene expression levels at a given position. Using Equation (4.10) to estimate PI has a subtle practical advantage over Equation (4.9) as discussed in the following.

Tkačik et al. [253] present several alternative methods to estimate PI from real data, called the direct (DIR) method, the first and second Gaussian approximation (FGA and SGA), and a Monte Carlo integration scheme (MCI). In the following discussion we first describe the DIR and SGA estimation methods, which are well suited to the specifics of our data sets. We then briefly discuss FGA and MCI, which in our case provide only minor advantages and were therefore not considered in detail.

The DIR method is attractive as it does not build on any prior assumptions about the distribution $p(\{g_i\}|x)$. However, the sample sizes required to estimate PI with a binned pdf are relatively large and grow exponentially with the number of genes J. A binned pdf can be obtained by dividing the fluorescence data at each location x into B bins of size $\Delta_g = 1/B$ (Figure 4.6b-d). Intuitively, if B is chosen too low, data is grouped into large bins and information is inevitably lost. Conversely, we can find a critical value B^* above which the pdfs become sparse. Naive PI estimates obtained by simply inserting a binned pdf into Equation (4.10) or Equation (4.9) therefore suffer from an estimation bias.

As for $B < B^*$, the estimation bias scales as $1/N$ and $1/B^{J+1}$ [286, 287], a DIR estimate can be obtained by extrapolating naive estimates for a series of fractions of the whole data with size $M < N$ and varying bin number $B < B^*$ towards infinite sample size $M \rightarrow \infty$ and zero bin size $\Delta_g \rightarrow 0$ via linear regression (Figure 4.7a,b). In practice, we first generate $K = 100$ sub-samples for each sample size $M = (6, 7, ..., 11) \cdot N/12 < N$ (ceiled to the next digit) randomly drawn from the full data set without replacement. Note, that for data sets with $N \leq 12$ we instead used $M = (7, 8, ..., N)$.

Then, we compute naive PI estimates for each of the sub-samples for a series of bin sizes $B = (2, 3, 4, 5, 6)$ (As shown below, $B^* = 6$ was a reasonable compromise for all our data sets). Note that per convention, we define B relative to the maximum mean gene expression profile, which we have normalized to 1. The actual bin number is chosen such that it spans the entire range of gene expression profiles and is thus usually larger than B. Also note that because some elements of a binned pdf will be 0, we need to (informally) set $0 \cdot \log(0) = 0 \cdot \log(0/0) = 0$ which follows from continuity [288, p.31].

For each (B, M) we then compute the mean and standard deviation of the K subsamples and linearly regress $I_{B,M} = I_B + c_M \frac{1}{M}$ (Figure 4.7a). As $1/M \rightarrow 0$, $M \rightarrow \infty$ and hence the intersects I_B are the corrected estimates in the infinite data limit. In a second extrapolation we can correct for the finite bin size by linearly regressing these I_B for the different bin numbers via $I_B = I_0 + c_B \frac{1}{B^{J+1}}$ (Figure 4.7b). Again, as $\frac{1}{B^{J+1}} \rightarrow 0$, we approach the limit of infinitely small bin size and I_0 is the final PI estimate. To gauge the precision of this estimation procedure, we take the 68% confidence interval of the last extrapolation as statistical uncertainty.

To find B^* for a given data set, we can perform a series of PI estimates with varying B^* [287]. As shown in Figure 4.7c an increase in the PI estimate, as noticeable for the joint PI estimate ($J = 2$), indicates an overestimation due to undersampling. Additionally, we can find B^* by randomly shuffling positions and gene expression levels [287] (Figure 4.7d), which should destroy any mutual

Table 4.1: Data set overview. S: data set number, T: topology, IPTG: inducer concentration in sender droplet, N: sample size. B^* are the critical bin numbers determined by the shuffling method. '<': $B^* < 4$, i.e. DIR is not applicable, '>': $B^* > 20$, i.e. higher than the tested range. All DIR estimates are for $B^* = 6$. DIR/ SGA estimates are after 7.5 h.

S	T	IPTG (mM)	N	$I_{DIR,1}$ (B^*) (bit)	$I_{SGA,1}$ (bit)	$I_{DIR,2}$ (B^*) (bit)	$I_{SGA,2}$ (bit)	$I_{DIR,1,2}$ (B^*) (bit)	$I_{SGA,1,2}$ (bit)
1	A	100	9	0.10 ± 0.19 (<)	0.06 ± 0.02	0.29 ± 0.29 (<)	0.06 ± 0.01	0.77 ± 0.39 (<)	0.00 ± 0.11
2	A	10	35	0.64 ± 0.03 (>)	0.56 ± 0.01	0.38 ± 0.02 (13)	0.36 ± 0.01	0.88 ± 0.07 (<)	0.67 ± 0.01
3	A	1	27	1.01 ± 0.03 (7)	1.08 ± 0.02	0.44 ± 0.04 (17)	0.29 ± 0.02	1.14 ± 0.07 (<)	1.11 ± 0.02
4	A	0.1	18	0.31 ± 0.07 (7)	0.36 ± 0.03	0.07 ± 0.03 (7)	0.02 ± 0.01	0.87 ± 0.14 (<)	0.37 ± 0.03
5	B	100	25	0.17 ± 0.02 (>)	0.17 ± 0.01	0.06 ± 0.01 (12)	0.00 ± 0.01	0.39 ± 0.01 (<)	0.24 ± 0.02
6	B	10	37	0.70 ± 0.02 (>)	0.67 ± 0.02	0.08 ± 0.03 (>)	0.07 ± 0.01	0.78 ± 0.02 (4)	0.76 ± 0.01
7	B	1	11	0.40 ± 0.13 (<)	0.35 ± 0.02	0.19 ± 0.12 (<)	0.00 ± 0.00	1.07 ± 0.18 (<)	0.35 ± 0.05
8	B	0.1	17	0.95 ± 0.08 (<)	1.15 ± 0.02	0.09 ± 0.01 (<)	0.03 ± 0.01	1.16 ± 0.12 (<)	1.19 ± 0.03
9	C	100	18	0.29 ± 0.05 (9)	0.27 ± 0.01	0.07 ± 0.01 (17)	0.00 ± 0.01	0.56 ± 0.09 (<)	0.26 ± 0.01
10	C	10	22	0.94 ± 0.17 (10)	1.13 ± 0.01	0.05 ± 0.02 (>)	0.00 ± 0.01	1.03 ± 0.15 (<)	1.18 ± 0.01

information $I_{shuffled} > 0$, unless B^* is chosen too large. We hence define $B^* = \max(B|I_{shuffled} + \sigma_{I_{shuffled}} < \tau)$, where $\tau = 0.1$ is a tolerance we allow for robustness. The minimum requirements to ensure at least 3-4 data points for each extrapolation fit are roughly $B^* \geq 4$ and $N \geq 10$. However, since B^* decreases with decreasing N, increasing σ_g, and increasing J, N may need to be considerably higher to ensure $B^* \geq 4$. Our data sets have $N = 9-37$ and for single genes, the conditions for the extrapolation are mostly met (Table 4.1). However, for calculating the joint information ($J = 2$), we hardly reach $B^* \geq 4$. As DIR is hence not applicable in these cases, we additionally use SGA.

To obtain an SGA estimate, we first approximate $p(\{g_i\}|x)$ by a Gaussian distribution (Equation (4.6)), where the mean and covariance can be calculated from the data (Equation (4.7) and (4.8)). This eliminates the binning problem, which aids PI estimation for multiple genes. We can then calculate the total and the noise entropy by inserting an inferred pdf into Equation (4.10). This is done numerically using a discretized multivariate normal pdf with a grid spacing of 0.01 (in the order of the smallest $\sigma_{g_i}(x,t)$). Then, we again extrapolate a series of naive estimates to infinite sample size (Figure 4.7e). Importantly, the Gaussian distribution is the maximum entropy distribution meaning that the SGA estimates can be understood as a lower bound [253]. To see how close the SGA estimates approximate the real PI value, we can compare them to DIR estimates. For the data sets for which DIR is justified, DIR and SGA estimates agree well both over time (Figure 4.7f,g) and across all data sets (Figure 4.7h).

FGA is a hybrid approach that exploits the fact that $p_g(\{g_i\})$ is sampled better than $p(\{g_i\}|x)$ because it includes the g measurements from every x. Hence, the total entropy in Equation (4.10) can be estimated directly, while the noise entropy can be calculated using the Gaussian approximation. However, our system only has 5 naturally discrete positions, whereas Tkačik et al. [253] chose to bin the continuous expression profiles of the *Drosophila* embryos into 1000 spatial bins. Hence, in our case the sampling of $p_g(\{g_i\})$ is not significantly better than $p(\{g_i\}|x)$ and consequently FGA provided only minor advantages over DIR. MC integration becomes relevant when estimating high dimensional joint PI carried by $J \geq 3$ genes. Then the fraction of the 'interesting' phase space becomes small and computations inefficient due to the curse of dimension. As we are interested in $J \leq 2$, we did not consider MC integration.

4.2.3 Calculation of Positional Error

Positional error $\sigma_x(x)$, here defined in units of droplets, can be calculated from the Fisher information

$$\mathcal{I}(x) = \int d^J\mathbf{g}\, p(\{g_i\}|x) \cdot (\partial_x \ln p(\{g_i\}|x))^2 \,. \tag{4.11}$$

In combination with the Cramér-Rao bound we get a lower limit for the positional error

$$\sigma_x^2(x) \geq \frac{1}{\mathcal{I}(x)} \,. \tag{4.12}$$

We can again approximate $p(\{g_i\}|x)$ with a Gaussian distribution Equation (4.6), which inserted into Equation (4.11) gives [253]

$$\mathcal{I}(x) = (\partial_x \bar{g}(x))^T C(x)^{-1}(\partial_x \bar{g}(x)) + \frac{1}{2}\mathrm{Tr}[C(x)^{-1}(\partial_x C(x))C(x)^{-1}(\partial_x C(x))] \,, \tag{4.13}$$

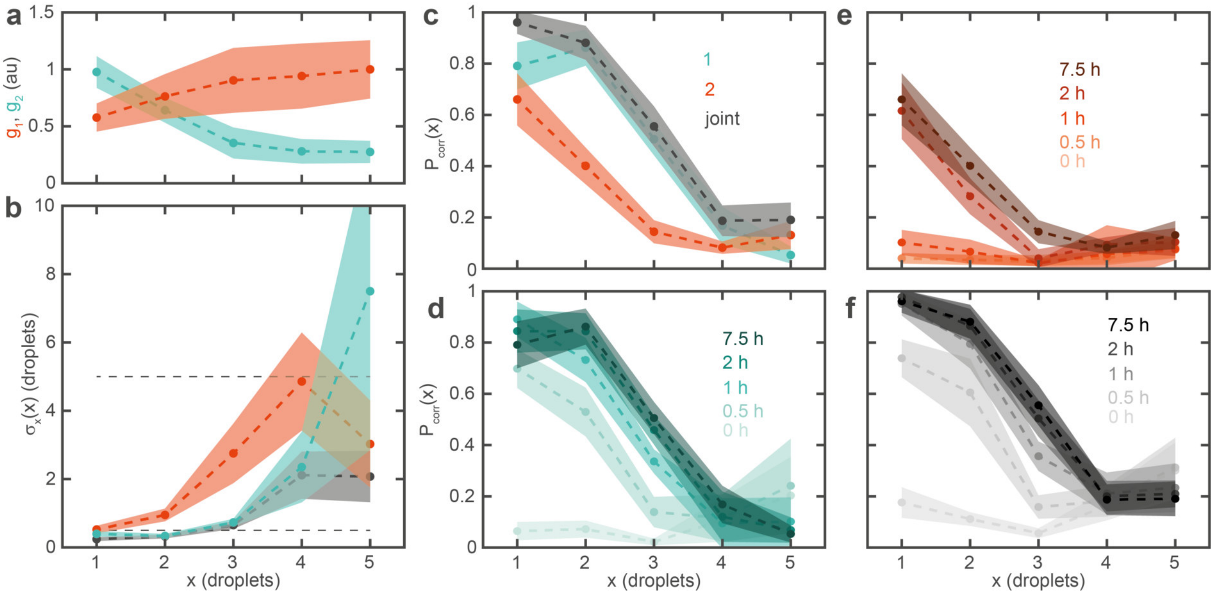


Figure 4.8: Positional Error. a) Mean fluorescence profiles with standard deviation. **b)** PE at the respective droplet positions. Dashed lines indicate a positional error of 1 and 5 droplets (i.e. the length of the entire sample). Lines are a guide to the eye. **c)** Probability of correctly identifying droplet position after 7.5 h. **d-f)** Temporal evolution of P_{corr}.

from which a lower bound for PE can be calculated from the mean and covariance (Equation (4.7) and (4.8)). Equation (4.11) is more intuitive to understand in the case of a single gene with the mean gene expression gradient $\bar{g}(x) = \bar{g}_i(x)$ and variance $\sigma_g^2(x) = C_{ii}(x)$. Then,

$$\mathcal{I}(x) = \frac{(\partial_x \bar{g}(x))^2}{\sigma_g^2(x)} + 2\frac{(\partial_x \sigma_g(x))^2}{\sigma_g^2(x)} , \tag{4.14}$$

where the first term represents the information that can be retrieved from the gradient of the mean profile and the second term represents information that can be retrieved from spatial variations in the noise (Figure 4.8a).

In practice, we first calculate PE by extrapolating a series of naive PE estimates (Equation (4.11) with (4.12)) to the infinite sample size limit ($K = 100, M = (6, 7, ..., 11) \cdot N/12 < N$, as for PI). The mean and SD of PE are then estimated by naively bootstrapping 500 times. $\sigma_x(x) \in R^+$ diverges for small $\mathcal{I}(x)$, which is inconvenient for visualization (Figure 4.8b). We therefore consider the probability that the position of a given droplet can be correctly identified $P_{corr}(x) : R^+ \to [0, 1]$ (Figure 4.8c-f)

$$P_{corr}(x) = \frac{1}{\sqrt{2\pi\sigma_x^2(x)}} \int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \exp\left(-\frac{(x')^2}{2\sigma_x^2(x)}\right) dx' , \tag{4.15}$$

where $\Delta x = 1$ is the spacing between two droplets and $P_{corr}(x) \to 0$ as $\sigma_x(x) \to \infty$ and $P_{corr}(x) \to 1$ as $\sigma_x(x) \to 0$. After a coordinate transformation $x'' = x' + 0.5$, we can evaluate the integral analytically

$$P_{corr}(x) = \frac{1}{\sqrt{2\pi\sigma_x^2(x)}} \int_0^1 \exp\left(-\frac{(x''-0.5)^2}{2\sigma_x^2(x)}\right) dx'' \approx \mathrm{erf}\left(\frac{0.353553}{\sigma_x(x)}\right) . \tag{4.16}$$

The uncertainty $\Delta P_{corr}(x)$ can be found via Gaussian error propagation

$$\Delta P_{corr}(x) = \left|\frac{\partial P_{corr}(x)}{\partial \sigma_x(x)} \Delta\sigma_x(x)\right| \approx \frac{0.398942}{\sigma_x^2(x)} \exp\left(-\frac{1}{8\sigma_x^2(x)}\right) \Delta\sigma_x(x) . \tag{4.17}$$

We note that one important consideration specific to our system is the question of how to calculate the discrete derivative $\partial_x f(x)$ of function $f(x)$ (mean or covariance matrix). As our system represents a discrete space of length $X = 5$, we cannot choose any arbitrary bin width smaller than $\Delta x = 1$ to improve accuracy. Hence, to get a suitable estimate of $\partial_x f(x)$ for all 5 positions, we take the forward derivative at the first droplet, the centered derivative at the inner droplets ($x = 2, 3, 4$), and the backward derivative at the last droplet, i.e.

$$\partial_x f(1) = \frac{f(2) - f(1)}{1} , \tag{4.18}$$

$$\partial_x f(x = 2,3,4) = \frac{f(x+1) - f(x-1)}{2} , \tag{4.19}$$

$$\partial_x f(5) = \frac{f(5) - f(4)}{1} . \tag{4.20}$$

4.3 Positional Error and Positional Information in Droplet Assemblies

We now can utilize this data analysis pipeline to make quantitative experimental observations (Section 4.3.1). Supported by modeling (Section 4.3.2), we gain qualitative insight into the reaction mechanism and discuss the relative impact of different noise sources in our system (Section 4.3.3). Finally, we demonstrate how parameter tuning can be used to optimize the PI that a GRN can extract (Section 4.3.4).

4.3.1 Experimental Observations

The calculation of PE in Figure 4.8 allows to evaluate the capability of the GRN to interpret a morphogen gradient quantitatively. Unless otherwise noted, the following analysis refers to topology A with 1 mM IPTG ($N = 27$). We can first notice that the noise in the LacI-RFP gradient is considerably larger than in the TetR-YFP gradient (Figure 4.8a). Consequently, the PE is systematically lower (Figure 4.8b), and P_{corr} is higher for the TetR-YFP gradient (Figure 4.8c). As discussed in Section 4.3.3, this may in part be attributed to propagation of noise from the TetR-YFP gradient, which senses the morphogen IPTG through a LacI repressible promoter, to the LacI-RFP gradient. Second, it is apparent that the performance of the GRN depends on the distance of the receiver droplet from the sender droplet. While, for TetR-YFP, P_{corr} in the two anterior droplets is above 80%, the accuracy drops from ≈ 50% in the third droplet to below 20% in the two posterior droplets. This is due to the nature of the diffusive morphogen gradient, which presumably does not reach concentrations in the order of the induction threshold of the GRN, before the TXTL reaction ceases. As shown below,

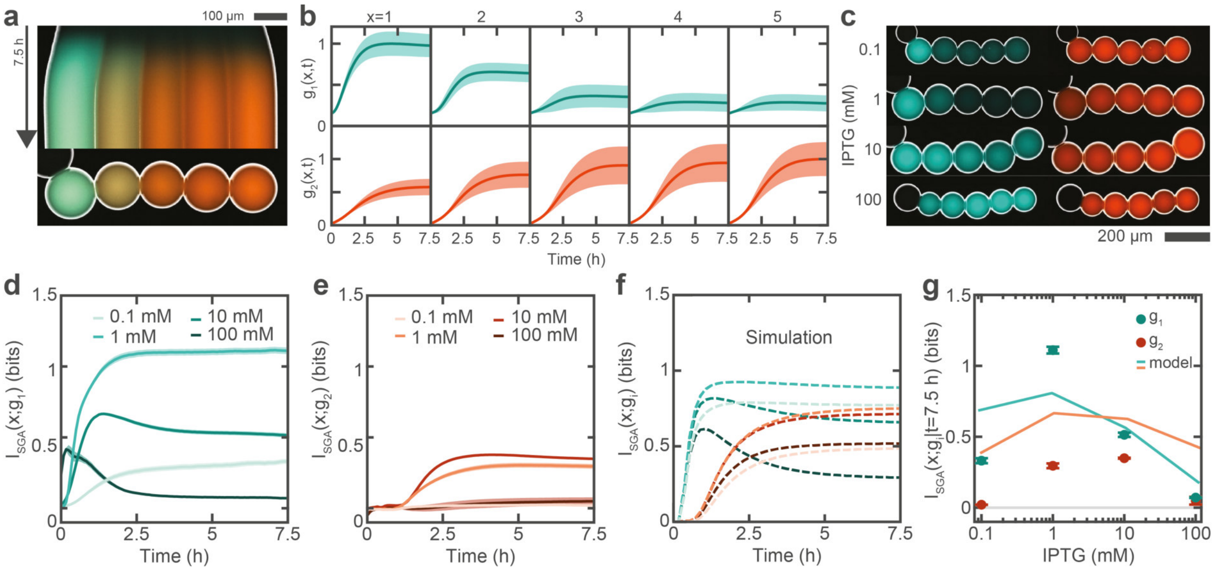


Figure 4.9: Experimental observation of gene expression gradients and determination of positional information. a) Kymograph of an overlay of two fluorescence channels and an inverted brightfield image showing the evolution of the two genes expressed in a single assembly for topology A at 1 mM IPTG. The IPTG sender droplet is on the left. **b)** Temporal evolution of normalized fluorescence intensity of the two genes (g_1: TetR-YFP, blue, top, g_2: LacI-RFP, red, bottom) for the full dataset grouped by droplet position relative to the sender droplet. Solid lines represent the mean and shaded area represents the standard deviation for $N = 27$ samples. **c)** Example images for fully developed assemblies (after 7.5 h) for varying concentrations. **d,e)** Temporal evolution of the estimated positional information in (d) the TetR-YFP and (e) the LacI-RFP gradient at different morphogen concentrations. The corresponding simulation **f)** reproduces the experimental observations qualitatively. Colors correspond to (d) and (e). **g)** Quantitative comparison of the final PI levels in simulation and experiment. Shaded areas in (d,e) and error bars in (g) are statistical uncertainties of the experimental PI and PE estimates.

naively increasing the morphogen concentration does not lead to an overall improvement of PE, or PI. Third, considering the joint information from both gradients does not significantly improve P_{corr} compared to considering the TetR-YFP gradient alone. Of course, this is partly because P_{corr} for the LacI-RFP gradient is comparably low in the first place. However, it is also in the nature of our GRN that the two gradients are strongly correlated, which means that they carry redundant information. Lastly, P_{corr} increases faster for the TetR-YFP gradient than for the LacI-RFP gradient (Figure 4.8d-f), but the kinetics are better visualized using the more compact PI measure presented in the following.

The temporal evolution of the gene expression patterns is visualized in an example kymograph in Figure 4.9a. The corresponding average fluorescence time traces with standard deviation of the full data set are shown in Figure 4.9b. Note, how repression of TetR-YFP in droplets 4 and 5 is relatively strong from the beginning, whereas repression of LacI-RFP in droplets 1 and 2 only becomes apparent after $1-2$ h. This makes sense, as LacI-RFP is only repressed once a sufficient amount of TetR has been produced. Because protein degradation in our closed system is relatively slow, even in the presence of a degradation tag in topology A, this delay leads to elevated base levels in the LacI-RFP gradient which contributes to the lower overall PI values.

Next, we investigated the influence of varying IPTG concentrations in the sender droplet. As shown by representative example images in Figure 4.9c, 0.1 mM IPTG mainly induces droplet 1, 1 mM IPTG produces a steep gradient across the first 3 droplets, 10 mM IPTG produces a more shallow gradient that extends across all 5 droplets, and 100 mM IPTG leads to homogeneous induction of all 5 droplets.

To quantify these impressions, we can calculate the PI for each time point. Considering the TetR-YFP gradient first (Figure 4.9d), the highest PI of 1.08 ± 0.02 bit is indeed reached for 1 mM IPTG, while lower concentrations lead to lower PI values. Interestingly, for 10 mM IPTG, the PI first increases and then peaks at ≈ 2 h, after which it decreases to a lower final value of 0.56 ± 0.01 bit. At 100 mM IPTG, the PI peaks between $0.5-1$ h and then even drops to 0.06 ± 0.02 bit, which is below the threshold τ used to find B^* in Figure 4.7d. What we can see here is an effect of the system's dynamics, mainly the diffusion of the morphogen gradient. For the high morphogen concentrations, the gradient initially induces the anterior droplets, which produces PI, but soon 'floods' the whole system. When the GRN in all droplets is fully induced, gene expression noise starts to decrease any transiently built up PI.

Next, we can analyze the temporal evolution of PI in the LacI-RFP gradient for the same morphogen concentrations (Figure 4.9e). As expected, the maximum PI values are considerably lower than for the TetR-YFP gradient and reach a maximum of 0.29 ± 0.02 bit and 0.36 ± 0.01 bit for 1 mM and 10 mM IPTG, respectively. What is interesting though, is that, in contrast to the TetR-YFP gradient, which produces PI instantly, essentially no PI is produced during the first ≈ 1 h. This delay is a direct consequence of the fact that LacI-RFP is controlled by a TetR repressible promoter and thus has to 'wait' until threshold level concentrations of TetR have been produced. As the TXTL reaction is most active during the initial hours [96], it is unsurprising that the overall values are comparably low.

4.3.2 ODE Model Reproduces Observations Qualitatively

To validate our observations and gain further insights, we developed an 8 ODE model, which yields good qualitative agreement with the experimental data (Figure 4.9f). The full model explicitly considers binding of IPTG to LacI, as well as maturation of fluorescent proteins, and is presented together with parameter values in the supplementary information of reference [38]. For illustrative purposes, we here only present the 3 key equations

$$\begin{aligned}
\frac{d[\mathrm{IPTG}]_x}{dt} &= \frac{DP}{V_x \cdot (PL_x + D)} \\
&\quad \cdot \left(A_{x,x+1}([\mathrm{IPTG}]_x - [\mathrm{IPTG}]_{x+1}) + A_{x,x-1}([\mathrm{IPTG}]_x - [\mathrm{IPTG}]_{x-1})\right) , && (4.21) \\
\frac{d[\mathrm{TetR}]_x}{dt} &= \alpha_T \cdot \left(\frac{K_L^{n_L}}{K_L^{n_L} + \left([\mathrm{LacI}]_x \frac{K_I}{K_I + [\mathrm{IPTG}]_x}\right)^{n_L}} + \alpha_0 \right) \cdot \left(1 - e^{t/\tau_{RNA}}\right) \cdot a(t) , && (4.22) \\
\frac{d[\mathrm{LacI}]_x}{dt} &= \alpha_L \cdot \left(\frac{K_T^{n_T}}{K_T^{n_T} + [\mathrm{TetR}]_x^{n_T}} + \alpha_0 \right) \cdot \left(1 - e^{t/\tau_{RNA}}\right) \cdot a(t) \\
&\quad - k_{deg} \cdot \frac{[\mathrm{LacI}]_x}{K_{deg} + [\mathrm{LacI}]_x} \cdot a(t) . && (4.23)
\end{aligned}$$

The ODE system is solved for each droplet x, individually. Only IPTG diffuses through the droplet network according to Equation (1.37), with an aqueous diffusion coefficient D, permeability P, droplet volume V_x, droplet length L_x and bilayer area $A_{x,x\pm1}$. TetR and LacI repress each other via Hill equations (Equation (1.16)) with apparent threshold constants K_T and K_L, apparent Hill coefficients n_T

and n_L, and maximum expression rates α_T and α_L. Note, that we here defined the leaky expression α_0 differently than in Equation (1.16). LacI is induced with IPTG with threshold constant K_I and, in presence of a degradation tag, may be degraded by ClpXP with zero-order degradation rate k_{deg} and threshold constant K_{deg} [96]. The term $\left(1 - e^{t/\tau_{RNA}}\right)$ implicitly considers transcription of mRNA with a lifetime of $\tau_{RNA} = 15$ min [96]. Finally, we account for the finite lifetime of the TXTL reaction with a step-wise activity function $a(t) = e^{-(t-t_1)/t_2}$, for $t > t_1$ and $a(t) = 1$, for $t \leq t_1$. We found the best agreement with our data for a reaction lifetime $t_2 = 1.5$ h without delay $t_1 = 0$ h.

As shown in Figure 4.9f, the model reproduces the observations that i) there is an optimum inducer concentration, ii) for high inducer concentrations, PI is transiently higher than the final value, iii) PI-RFP is generaly lower than PI-YFP, and iv) there is a delay before PI-RFP is generated. Quantitatively, the model predictions are in the correct order of ≈ 1 bit, however, the model tends to overestimate PI at non-optimal inducer concentrations and PI-RFP in general (Figure 4.9g). Despite the fact that our ODE model is of course a drastic simplification of a complicated TXTL reaction, this discrepancy may be partly resolved by parameter fine-tuning, as apparent parameters in the full system may differ from apparent parameters measured in isolation [11]. In addition, the absolute PI values are quite dependent on the particular sources of noise that are considered, as discussed in the following.

4.3.3 Simulated Impact of Different Noise Sources

Noise is a crucial component that affects PI. Two potential sources of the observed variability are geometric variability and gene expression noise. The gene expression noise can be estimated for instance from the RFP expression levels in topology C, or the YFP expression levels in topology A at saturating inducer concentrations (100 mM IPTG) and is ≈ 20%. Potential sources of gene expression noise are partitioning noise [74, 289] and macromolecular crowding [290]. The geometric variability can be estimated from microscopy videos using the accurate segmentation method (Figure 4.5c). By measuring the droplet radii and center distances, all other geometric parameters can be estimated via Equation (1.31) and (1.32). Relevant to the diffusion of IPTG are the bilayer areas A, which control permeation, and the droplet volumes V, into which the permeated morphogen is diluted. As shown in Figure 4.10a, droplets have an average volume of 1.2 ± 0.4 nL (≈ 35%), which slightly decreases over 7.5 h, probably due to evaporation. In contrast, the bilayer areas initially are $9 \pm 2 \cdot 10^3$ μm^2 large, but their size decreases drastically to $3.5 \pm 0.5 \cdot 10^3$ μm^2 after 7.5 h (Figure 4.10b). We assume that this happens as a result of adsorption of TXTL components to the lipid layer, which alters the bilayer and monolayer surface tensions and consequently the equilibrium contact angle. As this phenomenon is not included in our model, this is one possible source of quantitative discrepancies between simulations and experiments.

To test the relative influence of these variabilities, we next simulated gene expression gradients with our ODE model and a Monte Carlo method. To this end, we multiplied either the droplet volumes V_x, or the maximum expression rates α_T and α_L (independently) with a noise constant that was randomly drawn from a normal distribution with $\mu = 1$ and $\sigma_V = 0.35$, or $\sigma_\alpha = 0.20$, respectively, rejecting negative samples. Assuming a constant contact angle of $\theta_b = 42°$, bilayer areas can be calculated using via Equation (1.31) and (1.32). Repeating this sampling $N = 500$ times yields simulated gene expression profiles with noise as shown in Figure 4.10c,d.

Surprisingly, we observe almost no noise when varying the droplet volumes, but varying the gene expression noise quite closely reproduces the experimental gradients (Figure 4.8a), including the larger variability in the LacI-RFP gradient. While varying droplet volumes lead to a noisy morphogen gradient (Figure 4.1c), that noise only propagates in regions where the gradient is in a concentration

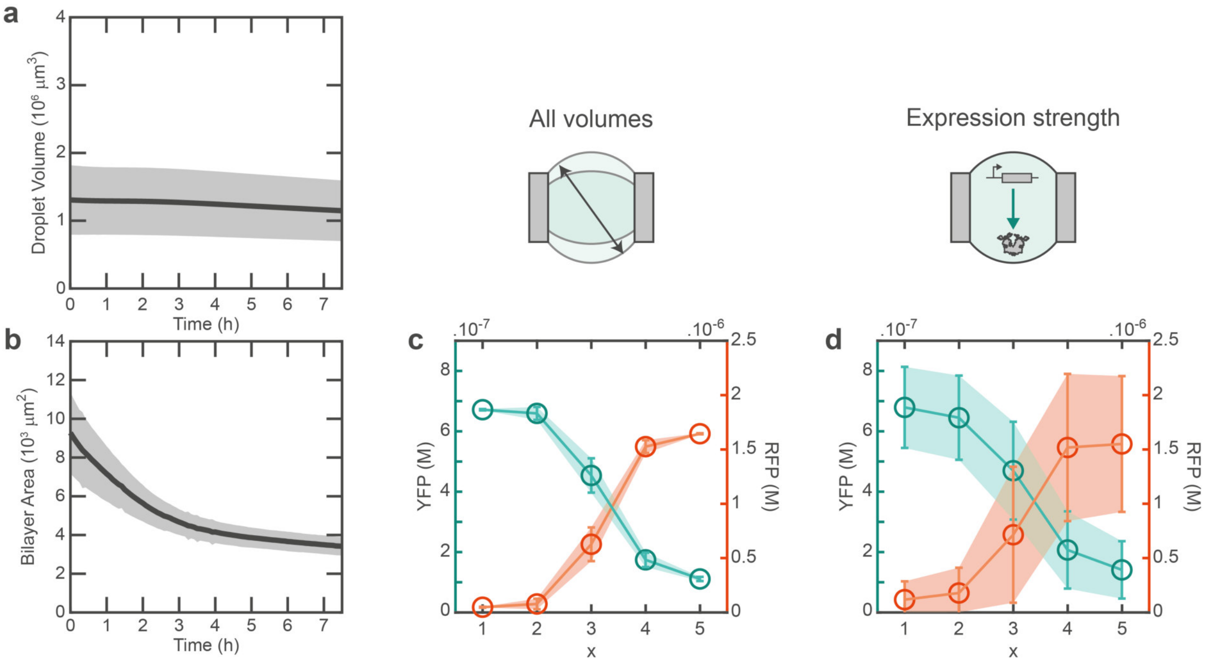


Figure 4.10: Simulated relative impact of geometrical noise and gene expression noise on PI. a) Experimentally estimated droplet volumes over time. The solid line represents the mean and the shaded area the standard deviation across the full data set with $N \cdot (X+1) = 27 \cdot 6 = 162$ droplets. **b)** Mean bilayer area with standard deviation for the same data set. **c,d)** Using Monte Carlo simulations, we consider two potential sources of noise to explain the observed variability in the gene expression gradients. First, geometrical noise, as caused for instance by variations in droplet volumes and consequently bilayer areas, leads to variability in the morphogen gradient (Figure 4.1c), but does not strongly propagate to a variability in the gene expression profiles (c). On the contrary, variability in gene expression strength does not affect the morphogen gradient, but leads to a variability of protein profiles (d) closely reproducing the observed variability in the gene expression profiles. Error bars are standard deviations from 500 simulations, where the droplet volumes, or gene expression strengths were drawn from a normal distribution with a CV of 35% and 20%, respectively.

range around the induction threshold K_I, where the transfer function is steep. Well above K_I, the system is simply fully induced and well below K_I, gene expression is off. Here, the morphogen concentrations are around the K_I in droplet 3, which reaches approximately half the maximum induction level. However, the variability in the morphogen gradient has here already averaged out over the first two droplets, which leads to the relatively low observed variability in the gene expression level.

For calculations of simulated PI in Figure 4.9, we considered both sources of variability, but this result suggests, that to optimize PI in our patterning system it is more important to reduce gene expression noise. One potential source of noise that we have not considered is related to the timing during manufacturing of the DIB samples. The droplet assemblies are constructed by first depositing individual receiver droplets, then individual sender droplets, and finally assembling everything by bringing the droplets into contact. This process takes $\approx 30-60$ min in total and $\approx 5-10$ min for connecting the receiver droplet, which initiates the diffusion of the morphogen. TetR-YFP is initially repressed by the LacI that is natively present in the TXTL solution. However, because the assembly is performed at room temperature, it is possible that LacI-RFP is already expressed before the sender droplet is connected, which could partly explain the higher LacI-RFP base levels and is one possible explanation for the lower experimentally measured PI-RFP compared to simulations (Figure 4.8a, Figure 4.10d).

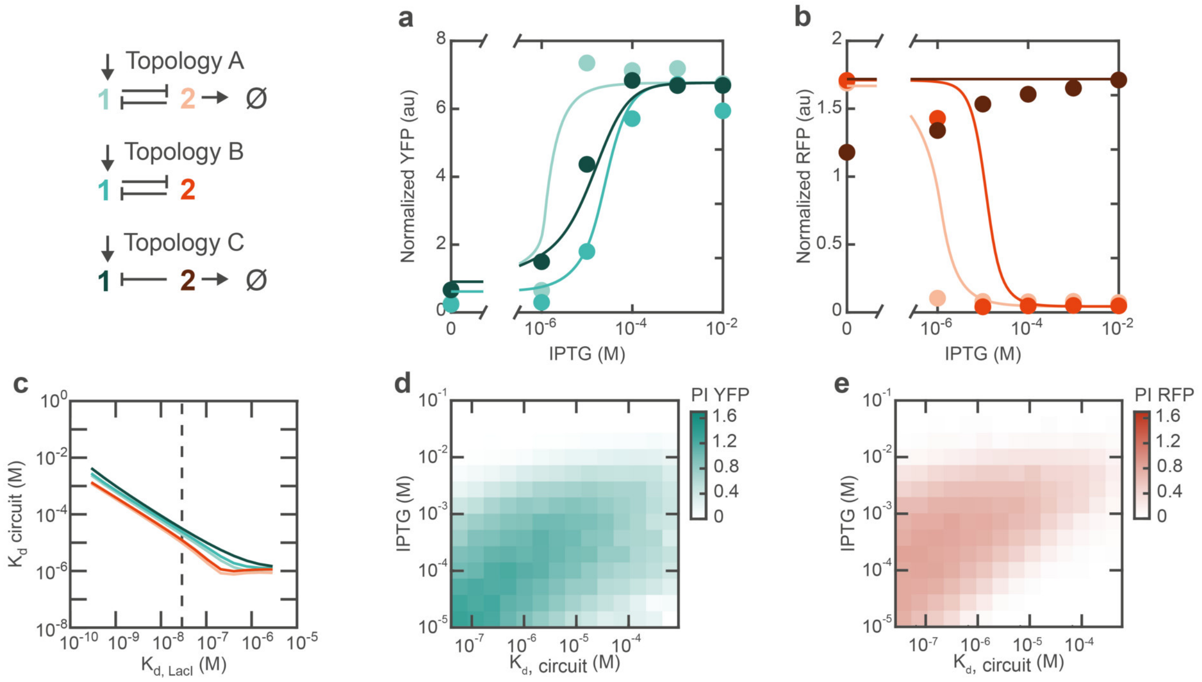


Figure 4.11: Optimizing PI extraction with parameter tuning. a,b) GRN topology changes the apparent circuit K_d. Dots show maximum expression data from bulk experiments for topology A, B and C at varying IPTG concentrations. Solid lines are simulations using the ODE model (Equation (4.22) and (4.23)), that was adapted for the corresponding GRN topology (topology B: $k_{deg} = 0$, topology C: $K_T = \infty$) without fitting parameters. The combination of a degradation tag and feedback reduces the apparent circuit K_d for both gradients. From topology A to C, the apparent circuit K_ds are (a) 1.51 µM, 26.2 µM, and 18.6 µM and (b) 1.01 µM, 11.7 µM, and ∞ µM, respectively. **c)** In the model, the apparent circuit K_d can be tuned by varying K_L, which is related to the circuit K_d by a power law for all topologies. Dashed line indicates the default value. **d,e)** Simulated PI in the YFP (d) and RFP (e) gradient, respectively, for varying IPTG concentration and circuit K_d. The simulations were generated for topology A, but the general idea is independent of topology.

4.3.4 Tuning the Apparent Threshold Constants

One strategy to improve the PI the GRN can extract, that does not depend on reducing the gene expression noise, is to tune reaction parameters. For instance, Figure 4.11a,b shows that the apparent threshold constants (here called 'circuit K_ds'), i.e. the IPTG concentrations that lead to half maximum expression of YFP or RFP, change based on the network topology. Note, that the circuit K_d is a phenomenological constant and is distinct from K_T, K_L or K_I, which are constants independent of the topology. In agreement with experiments, simulations predict that the combination of a degradation tag and feedback in topology A reduces the apparent circuit K_d of both gradients $\approx$ 10-fold compared to the other topologies.

To investigate the effect of a lower circuit K_d on PI, we calculated simulated PI values for different apparent threshold constants and morphogen concentrations. To this end, we varied K_L, which changes the apparent circuit K_d according to a power law (Figure 4.11c). As shown in Figure 4.11d,e, at a given circuit K_d there is an optimal IPTG concentration which leads to the maximum PI, in agreement with experiments (Figure 4.9g). Importantly, the maximum PI increases with decreasing circuit K_d. In conclusion, this suggests that lowering the circuit K_d is an effective method to increase the maximum PI without reducing gene expression noise, or improving the manufacturing process.

4.4 Comparison with the Gap Gene Network of *Drosophila*

To summarize, we have created assemblies of droplets that differentiate based on a diffusing morphogen gradient that is interpreted by a simple GRN. By generating a large number of samples and developing an image processing routine that corrects most of the measurement uncertainties, we were able to estimate the statistical measures PE and PI. This analysis, in combination with modeling, then allowed us to identify the key parameters to optimize the robustness of our synthetic patterning system. The capability of the GRN to extract PI depends on the kinetics of the morphogen gradient, the timing of reactions, the reaction lifetime, the sources of noise, and specific parameter values of the GRN that can be tuned by changing the topology. The manufacturing can potentially be scaled and adapted to more complicated structures using 3D printing techniques [38], which we here omitted for conciseness.

To assess the relative importance of these factors and identify potential routes for future improvement, it is interesting to compare our system with biological developmental systems, such as the well-studied gap gene network in the *Drosophila* embryo [40] (Section 1.3.1, Figure 1.13). With about 500 µm, our assemblies and the *Drosophila* egg are of comparable size. Our assemblies consist of $X = 5$ droplets that are separated by membranes, whereas *Drosophila* embryos display about $X = 59 \pm 4$ nuclei along the anterior-posterior (AP) axis (nuclear cycle 14) [291], that share a common cytoplasm [40]. Both systems as a whole are closed, which means that they rely on a finite stock of energy to produce the pattern.

Under optimized conditions, our system was able to extract 1.08 ± 0.02 bit of information with one gene (TetR-YFP), falling short of the theoretical maximum of $\log_2 5 = 2.32$ bit. With 1.11 ± 0.02 bit, the joint positional information from both gene expression gradients is not significantly higher. The gap gene network in *Drosophila* consists of four major genes, *hunchback* (*hb*), *Krüppel* (*Kr*), *knirps* (*kni*), and *giant* (*gt*), each of which carries 1.75 ± 0.05 bit to 2.26 ± 0.04 bit of positional information along the AP axis [291]. The theoretical maximum information is 5.9 ± 0.1 bit [291], if each of the $X = 59 \pm 4$ nuclei had their own identity. The experimentally determined joint information is at least 4.3 ± 0.1 bit, which implies that the information carried by the gap genes is partly redundant and partly complementary [253]. Hence, one approach to improve our system is to increase the complementarity of the gene expression gradients to improve the joint PI, possibly by utilizing GRN-level interactions across droplets [43].

Another component that may be optimized is the morphogen gradient and the mechanism that creates it. Our morphogen gradient consists of a fast diffusing molecule, with a diffusion constant $D \approx 1000\ \mu\text{m}^2\,\text{s}^{-1}$, that is slowly released from a passive sender droplet (Table 1.3). We have shown that when the molecule is released slowly, passive senders can emulate active sources of morphogens [36], but for longer reaction times, a passive sender might deplete. Permeability limited transport across droplets reduces the apparent diffusion constant to $D_{app} \approx 1\ \mu\text{m}^2\,\text{s}^{-1}$, resulting in a discrete and dynamic exponential profile. In contrast, the Bcd gradient in *Drosophila* is produced from maternal *bicoid* (*Bcd*) mRNA localized at the anterior pole of the embryo [40]. From this source, Bcd diffuses through the embryo to establish an exponential gradient that is stable during nuclear cycles 10-14 [292] and scales with the size of the embryo [293]. This is achieved by degradation of Bcd, which is supposedly enhanced inside nuclei [292]. In addition, many natural systems utilize two opposing gradients. In *Drosophila*, the Bcd gradient is complemented with a Cad gradient that is generated from maternal *caudal* (*cad*) mRNA localized at the posterior pole [40]. Another example for stable antiparallel morphogen gradients, which provides information during first 30 h of development, can be found in the developing vertebrate neural tube [284]. In our system, a complementary gradient may help to improve the relatively poor posterior PE (Figure 4.8).

Finally, improving energy supply and efficiency to prolong the reaction lifetime, as well as further enhancing the degradation rate to reduce the system timescale, may allow for more complicated, multi-step regulatory interactions. For example, the information encoded in the gap gene expression levels is decoded by the pair-rule genes in an optimal process that allows to specify the position of pair-rule expression patterns with 1% accuracy [253, 259, 291]. In contrast, if we consider our system as an information transfer cascade from TetR-YFP to LacI-RFP, the information drops from 1.08 ± 0.02 bit to 0.29 ± 0.02 bit and we have attributed this primarily to the limited reaction lifetime. Natural systems seem to have evolved optimal and efficient solutions to utilize their finite energy resources. For instance, the absolute concentration of Bcd at the anterior pole is 55 ± 3 nM, which means that a central nucleus has to be able to measure differences of ≈ 70 molecules, which appears to be at the stochastic limit to ensure the experimentally measured positional error [254]. Our system with expression in the order of 1 μM seems comparatively wasteful. Unfortunately, decay of the TXTL reaction is usually not dependent on the amount of produced protein [20]. A potential workaround is to implement a continuous energy supply [23], which also helps to enhance degradation rates and allows to implement additional regulatory interactions. In contrast to an egg that needs to contain enough energy to fully develop an animal that can eat, continuous energy supply is superficially similar to development in mammals where the mother continuously supplies the embryo.

5 Period Doubling Bifurcations in a Forced Cell-Free Genetic Oscillator

The contents of this chapter are based on the publication:

L. Aufinger, J. Brenner, and F. C. Simmel. “Period Doubling Bifurcations in a Forced Cell-Free Genetic Oscillator”. *bioRxiv* (preprint). DOI: 10.1101/2021.12.15.472802.

Across biological kingdoms, organisms including bacteria, fungi, plants, insects, and humans, regulate their day-night cycle with a circadian clock [294, 295]. The ability to measure time is presumed to have concrete evolutionary advantages [295]. In humans, malfunctions of the circadian clock are associated with diseases such as sleep disorders [296], or cancer [297]. At the molecular level, circadian clocks are often comprised of coupled genetic oscillators that are synchronized to external zeitgeber signals [298]. Theoretical studies of various circadian oscillators predict that both forced [299, 300] and freely coupled systems [301, 302] exhibit non-linear phenomena such as splitting, period-doubling, and chaos. While there is experimental evidence for de-synchronization [303] and splitting [304], observation of period-doubling and chaos in circadian clocks have remained elusive due to the experimental challenges associated with long-term observations of such systems [302].

Experimental investigation of period-doubling and chaos in a biological organism would require the accurate measurement of amplitude over many oscillation periods in a potentially fluctuating environment and in the presence of homeostatic regulation processes. One strategy to circumvent these challenges is to study minimal synthetic gene oscillators that can be operated in a controlled and isolated context. Synthetic oscillators have been previously implemented in bacteria [10], mammalian cells [305], and in cell-free batch [53, 160] or continuous reactions [22, 23]. Such systems have been used to study synchronization between communicating cells [14, 28] and among coupled oscillators [90], but also the effects of partitioning [74] and gene expression noise [45]. Transient oscillations have been found close to bifurcations [306].

Engineered gene oscillators can provide molecular rhythms or act as biochemical clocks in other contexts than their circadian counterparts. For instance, the oscillation period of a synthetic oscillator has been used as an accurate measure of bacterial growth rate [45, 307]. Cell-free gene oscillators have been utilized to drive autonomous molecular devices [72], control self-assembly processes [73] or spatio-temporal pattern formation [28]. Previously established synthetic oscillators were operated without periodic synchronization to an external signal, however, and thus provided only an intrinsic measure of time, which lost synchrony with ‘universal time’ after a few periods [308].

Here, we investigate the synchronization of a cell-free genetic oscillator [28, 90] to an external zeitgeber signal using a microfluidic reactor [23] that was previously employed for rapid prototyping of gene circuits [89]. We first verified that the dynamics of the free-running oscillator are well described by a simple model comprised of only four ordinary differential equations (ODEs) (Section 5.1). We then tested the effects of periodic forcing on the oscillator within the model, and found that the system displays period doubling bifurcations when varying the ratio of the input period to the period of the free oscillator $\lambda = T_{in}/T$ in the simulations.

Experimentally, we realized the external forcing by periodically adding either a transcriptional repressor (TetR) or an inducer (aTc), and recording experimental time traces for up to 48 hours (Section 5.2). For input periods close to the intrinsic period of the oscillator ($\lambda \approx 1$), we indeed find evidence of period doubling and even quadrupling in the forced system. Larger values of λ result in stable 1-cycles 'entrained' on the external zeitgeber. Further analysis, aided by simulations, suggests that with increasing non-linearity in the biochemical feedback loop, similar driven systems could display increasingly complex dynamics, including chaos. We conclude by discussing the implications of our results for natural and synthetic circadian rhythms (Section 5.3).

5.1 Free Oscillator

To study the dynamics of the free-running oscillator, we first introduce an ODE model (Section 5.1.1) and then compare the simulation results with experimental data (Section 5.1.2). As shown in Section 5.1.3, the period of the free oscillator defines an intrinsic timescale.

5.1.1 ODE Model of the Oscillator Circuit

As shown in Figure 5.1a, our oscillator circuit consists of two regulatory proteins. Sigma factor σ^{28} activates the expression of TetR, which in turn represses the expression of the activator, thereby forming a negative feedback loop. In the experiment, the dynamics of the system is monitored by co-expression of the fluorescent reporter proteins mVenus and mTurquoise2 for the activator and repressor, respectively. To synchronize the oscillator to an external clock signal, the system can be perturbed by either adding purified TetR from the outside, or by inactivating intrinsic TetR via induction with anhydrotetracycline (aTc). Already in this coarse-grained picture the system is constituted of three coupled dynamical variables - activator, inhibitor, and external signal -, which is one of the requirements for a system to exhibit complex non-linear dynamics [6].

To properly describe the dynamics of the genetic oscillator, however, it is necessary to explicitly consider the dynamics of the mRNA molecules (Figure 5.1b), which effectively generates the time delay that is required for sustained oscillations [33]. The dynamics of the free oscillator is then described by the following set of four ordinary differential equations,

$$\dot{r_a} = \alpha_a \frac{1}{1+(h/K_h)^{n_h}} - \left(\delta + \frac{1}{\tau_{m,a}}\right) \cdot r_a \ , \tag{5.1}$$

$$\dot{a} = r_a \cdot k_{TL,a} - \delta \cdot a \ , \tag{5.2}$$

$$\dot{r_h} = \alpha_h \frac{1}{1+(K_a/a)^{n_a}} - \left(\delta + \frac{1}{\tau_{m,h}}\right) \cdot r_h \ , \tag{5.3}$$

$$\dot{h} = r_h \cdot k_{TL,h} - \delta \cdot h \ , \tag{5.4}$$

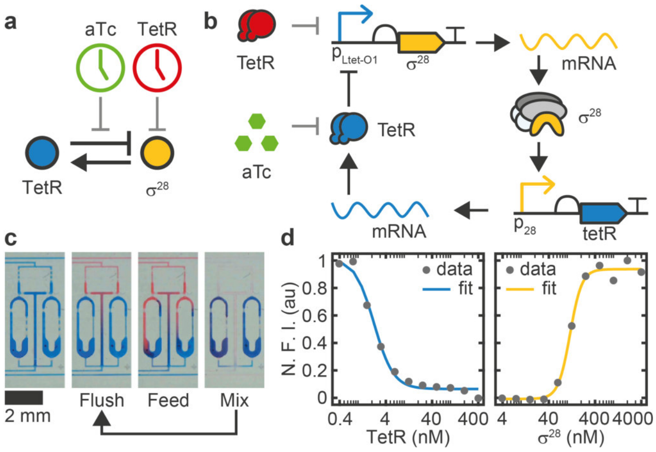


Figure 5.1: Genetic oscillator circuit and experimental setup. a) Simple and **b)** detailed circuit diagram of the synthetic genetic oscillator consisting of a negative feedback loop where σ^{28} acts as activator and TetR acts as repressor. Our four-variable ODE model (Equation (5.1))-(5.4)) considers the concentrations of the two proteins and their respective mRNAs. Protein expression is monitored via co-expression of fluorescent reporters mVenus and mTurquoise2, respectively. We can perturb the system either by inactivating existing TetR by induction with aTc, or by adding purified TetR. **c)** A microfluidic ring reactor [23] was used to maintain the reaction out of equilibrium by periodically exchanging a fraction of the reactor volume with fresh reagents. By switching between different input reagents, the reactions can be exposed to an arbitrary series of inputs. **d)** Transfer functions of the two promoters determined by titrating the regulator protein in bulk. Fitted Hill parameters with 68% confidence intervals are $K_h = 2.2 \pm 0.2$ nM, $n_h = 2.1 \pm 0.3$, $K_a = 115 \pm 6$ nM, and $n_a = 3.4 \pm 0.6$. N.F.I.: Normalized Fluorescence Intensity.

which has been used previously for the analysis of similar systems [28, 53]. The variables r_a, a, r_h, and h denote the concentrations of the activator and inhibitor mRNA and protein species, respectively. To aid the following discussion, we conceptually distinguish between *system parameters* that are essentially fixed properties depending on molecular details, and *control parameters* that can be experimentally tuned relatively easily.

The transcription rates α_a and α_h can be tuned linearly by adjusting the gene template concentrations [23, 57] and will therefore be considered as control parameters. A third control parameter is given by the dilution rate δ, which, as shown below, defines the timescale of the system. Experimentally, the reaction is kept out of equilibrium using a semi-continuously operated microfluidic ring reactor [23] (Section 2.2). As shown in Figure 5.1c, the reactions are maintained inside ring-like reaction chambers, whose volume is periodically replaced by a fraction R $(0 < R < 1)$ of fresh reagents, called the 'refresh ratio'. With a fixed time interval t_{int} between dilution cycles, the dilution rate can be precisely tuned by varying R according to

$$\delta(R) = -\frac{\ln(1-R)}{t_{int}} . \tag{5.5}$$

It had been previously shown that a reactor operated in semi-continuous mode can be used to emulate reactions in a continuously operated flow reactor, provided that t_{int} is sufficiently small compared to the system's timescale [23, 89].

The system parameters are the mRNA lifetimes $\tau_{m,a}$ and $\tau_{m,h}$, and translation rates $k_{TL,a}$ and $k_{TL,h}$, whose values have been determined previously [96], and the threshold constants K_a and K_h, and

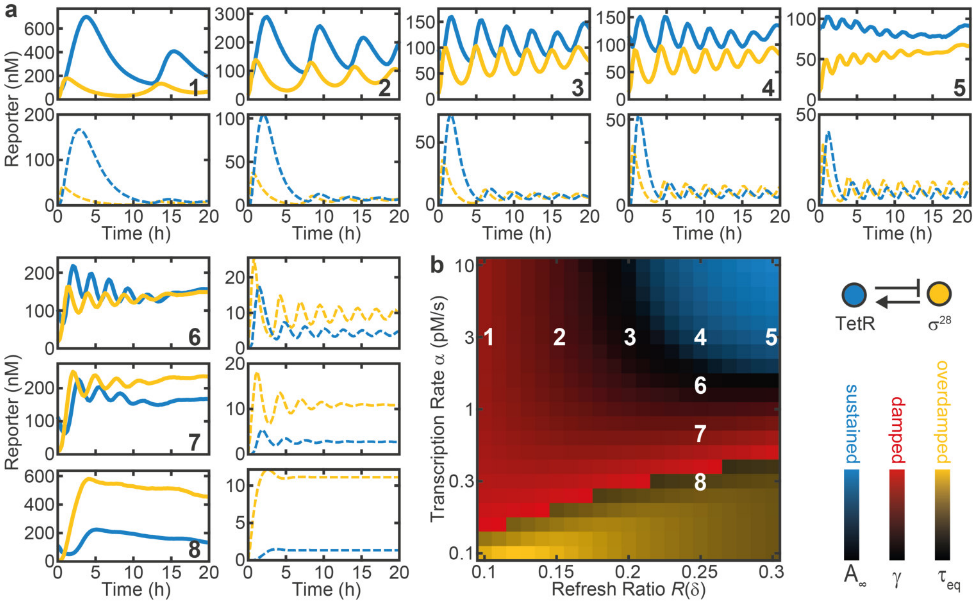


Figure 5.2: Dynamics and phase diagram of the free oscillator. a) Experimental (full) and simulated (dashed) time traces of the free oscillator for different refresh ratios $R(\delta)$ and DNA concentrations as indicated in (b). **b)** Simulated phase diagram of the free oscillator. Color overlay of different metrics as indicated in the legend reveals regions of sustained, damped, and overdamped oscillations. A_∞, γ, and τ_{eq} are the normalized equilibrium amplitude, damping ratio, and equilibration time constant for trajectories with less than one detectable maximum, respectively. Experimental data was mapped onto the diagram using α as a fitting parameter proportional to the DNA concentration. System parameters: $n = 3$, $K_a = 20$ nM, $K_h = 2$ nM, $k_{TL} = 0.02\ \mathrm{s}^{-1}$, $\tau_m = 12$ min. DNA concentrations are $0.1-1$ nM ($0.3-3\ \mathrm{pMs}^{-1}$) for the circuit plasmids and 2 nM for reporter plasmids.

Hill coefficients n_a and n_h, which can be estimated from bulk titrations (Figure 5.1d). As a caveat, one has to consider that parameters measured in isolation do not necessarily match their apparent (effective) values in the coupled system [11] - for instance, we do not explicitly account for reactions such as the competition between σ^{28} and σ^{70} for the RNAP core enzyme [90, 93].

To illustrate the effect of system and control parameters on the dynamics of the free oscillator, we can consider the nullclines ($\dot{r_a} = 0$, $\dot{r_h} = 0$, with $\dot{a} = 0$, $\dot{h} = 0$, and assuming $\delta \ll 1/\tau_m$)

$$a = \frac{\alpha_a}{\delta} \cdot \frac{k_a \cdot \tau_{m,a}}{1 + (h/K_h)^{n_h}}\ , \tag{5.6}$$

$$h = \frac{\alpha_h}{\delta} \cdot \frac{k_h \cdot \tau_{m,h}}{1 + (K_a/a)^{n_a}}\ . \tag{5.7}$$

Hence, the system parameters describe the *shape* of the gene transfer functions, whereas the control parameters define their *scale*. As the stability of the fixed point at the intersection of the nullclines depends on the local shape of the nullclines (cf. the thorough linear stability analysis described in Ref. [53]), the control parameters can be used as bifurcation parameters to tune the qualitative behavior of the system, whereas the system parameters define the relative sizes of regions in parameter space corresponding to qualitatively different dynamics. For instance, increasing n will increase the parameter range that supports sustained oscillations. In the following, we assume that the system parameters are uniform for the activator and inhibitor, i.e., $\alpha := \alpha_a = \alpha_h$, $n := n_a = n_h$, $k_{TL} := k_a = k_h$,

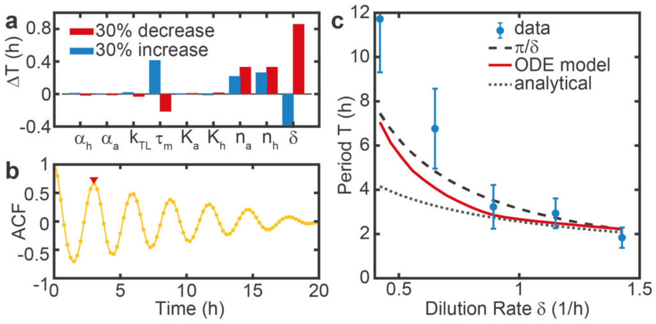


Figure 5.3: The period of the free oscillator is determined by the dilution rate. a) A sensitivity analysis was performed by analyzing the change in the period ΔT in response to a 30% change in individual model parameters [302]. **b)** Example of an auto-correlation function (ACF) that was used to estimate experimental periods for both reporters, corresponding to sample 4 in Figure 5.2a. The red triangle indicates the first maximum. **c)** Period as a function of the dilution rate $T(\delta)$. The experimentally measured periods are compared to predictions by our ODE model, an analytical solution by [309], both with an mRNA lifetime of 12 min [96], and to the phenomenological equation $T(\delta) = C\pi/\delta$ with $C = 1$. While the predictions are in good agreement with the data at higher dilution rates, the measured periods are systematically higher at lower dilution rates. Error bars are standard deviations of 2-4 measurements, plus a systematic uncertainty that scales inversely with the number of maxima in the recorded time trace. System parameters are as in Figure 5.2, with $\alpha = 3$ pMs^{-1}.

$\tau_m := \tau_{m,a} = \tau_{m,h}$, which is a standard approach to simplify the analysis while preserving the main qualitative features [10, 309].

5.1.2 Operation of the Free Oscillator

To experimentally verify the predictions of the model, we tested the free oscillator for a wide range of dilution rates δ and transcription rates α (Figure 5.2a). In good qualitative agreement with the model, we find regimes that display sustained, damped and strongly damped oscillations with varying periods. For the simulations, we used α as a global fitting parameter with a fixed ratio between samples. Reduction of α leads to a transition from sustained to damped oscillations, whereas δ mainly affects the period of the oscillations.

We also mapped the oscillator dynamics onto a simulated phase diagram, as shown in Figure 5.2b. To this end, we characterized the time traces of numerically simulated oscillations by their equilibrium amplitude A_∞ and mean damping ratio $\gamma = \left\langle \frac{y_i}{y_{i+1}} \right\rangle$, calculated from successive maxima y_i. Strongly damped oscillations with less than two maxima were characterized by the exponential equilibration time τ_{eq}. This illustrates that higher gene expression strength α and higher dilution rates δ favor sustained oscillations.

5.1.3 Intrinsic Timescale of the Oscillator

We next investigated the dependence of the period T of the free oscillator on the model parameters using a simple form of sensitivity analysis (Figure 5.3a). To this end, we tested the change of the period ΔT in response to a 30% change in each of the parameters individually [302]. In agreement with our naive expectation, the dilution rate δ is found to be the dominant control parameter determining the period of the oscillator T. The only other relevant parameters are the Hill coefficients n_a and n_h, and the mRNA lifetime τ_m, which are fixed system parameters.

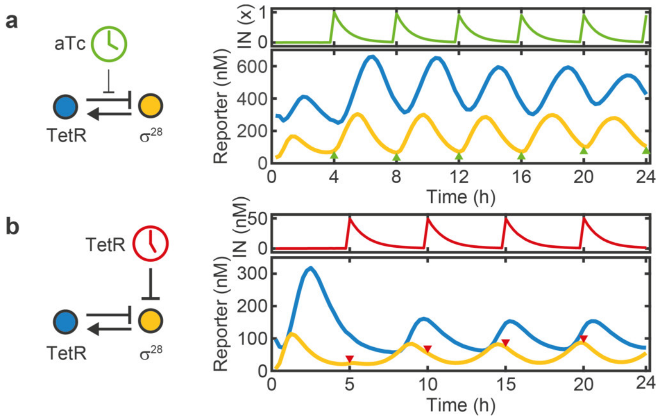


Figure 5.4: Dynamics of the forced oscillator. Here, we either added **a)** $1\times = 214$ nM aTc every 4 h ($\lambda = 1.43$), or **b)** 50 nM TetR every 5 h ($\lambda = 1.48$). Once added, the input signals are removed by the dilution with fresh reagent (without input), which results in an exponential decay of the input signal. The oscillator responds by increasing, or decreasing the production of σ^{28}, respectively, and synchronizes to the input period.

We then estimated the oscillator periods from experimental data using the first maximum of the auto-correlation function (ACF). Figure 5.3b shows an example ACF corresponding to sample 4 in Figure 5.2a. The experimental data agrees well with the predictions from our ODE model for different δ, as shown in Figure 5.3c. The discrepancies at lower dilution rates are likely a result of the low number of complete cycles in the experimental time traces due to the long periods. This leads to an overestimation of the experimental periods, as the system initially has to approach the limit cycle. In contrast, an analytical solution (Equation (24) in Ref. [309]), appears to more strongly underestimate the periods at lower dilution rate, probably as a result of the assumption that time traces are sinusoidal. Phenomenologically, we find that the ODE model predictions and the experimental data can be well approximated by the simple equation

$$T(\delta) \approx \frac{C\pi}{\delta}\,, \tag{5.8}$$

where C has a value close to 1. Importantly, this allows us to define the systems timescale based on the dilution rate δ.

5.2 Forced Oscillator

We now establish a forcing mechanism (Section 5.2.1), and find experimental proof of period doubling (Section 5.2.2). We then simulate a bifurcation diagram that shows close correspondence with the experimental data (Section 5.2.3). Finally, we extend the analysis to suggest parameter ranges for which the model produces more complex dynamics, including chaos (Section 5.2.4).

5.2.1 External Forcing of the Genetic Oscillator

First, we investigated whether we can force the oscillator to adapt to a certain period by externally supplying a periodic input signal. To this end, we replaced either the cell extract or the buffer supplied in every k-th dilution step with extract or buffer supplemented with TetR or aTc, to repress or activate the expression of σ^{28}, respectively (Figure 5.4). This generates a periodic input signal with a period $T_{in} = k \cdot t_{int}$ and amplitude A_{in}, that rises instantly and decays exponentially with rate δ, as monitored with a fluorescent reference signal (mScarlet-I).

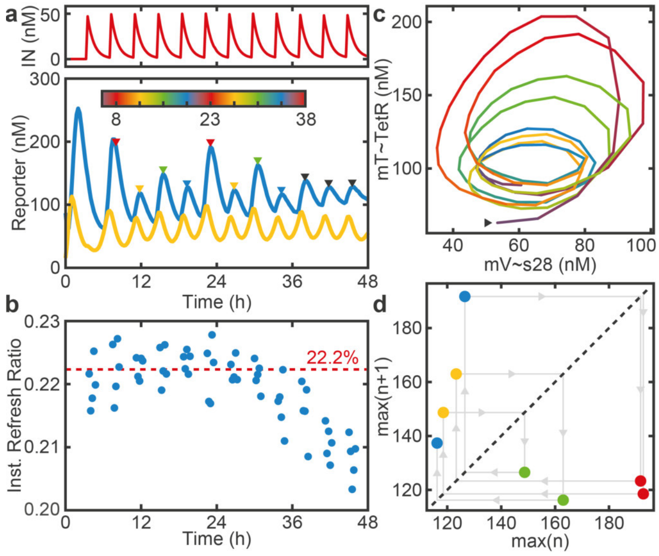


Figure 5.5: Period Doubling. a) Time traces of a 4-cycle ($\lambda = 1.20$). **b)** The instantaneous refresh ratio drops after 36 h due to fatigue. The dashed line indicates the average R_t over the first 24 h. **c)** Phase space trajectory. Color indicates time and is scaled to span $4 \cdot T_{in}$, as indicated by the colorbar in (a). The arrow marks the starting point of the trajectory, which follows a counter-clockwise path. **d)** Maximum return map for the maxima indicated in (a). After the last point, the system transitions to a 1-cycle.

The amplitude of the input signal must be chosen such that it triggers a sufficiently strong response by the oscillator, but is also diluted to levels well below the induction threshold sufficiently fast. For instance, an input signal with amplitude 1 will drop to $(1-R)^k = 0.01$ after $k = 16$ dilution steps with a refresh ratio $R = 0.25$. Hence, in practice there is a minimum input period, typically ≈ 2 h, below which effective forcing becomes challenging due to the low attainable dynamic range of the input signal.

As shown in Figure 5.4, the forced oscillators quickly adapt to T_{in} within $\approx 1-2$ cycles for both methods of external driving. While the amplitude is enhanced for positively forced oscillations, it decreases for negatively forced oscillations and the phase of the σ^{28} signal is shifted by $\pm\pi/2$ relative to the input signal for positive and negative forcing, respectively. In both cases the system displays regular 1-cycle oscillations.

As described below, the forced oscillator system can exhibit more complex dynamics, which can be described with a single dimensionless bifurcation parameter

$$\lambda = \frac{T_{in}}{T} \approx \frac{\delta \cdot T_{in}}{\pi} = \frac{\ln(1-R)^{-k}}{\pi}\ , \tag{5.9}$$

which is obtained as the ratio of input period and intrinsic period and using Equation (5.5) and (5.8) (with $C = 1$). Note that $\ln(1-R)^{-k}$ is the logarithm of the total dilution after one input period.

5.2.2 Period Doubling

In the following, we focus our analysis on the negatively forced oscillator that uses TetR as the periodically varying external input. When simulating the forced oscillator for different natural periods T and input periods T_{in}, we found that under certain conditions the resulting oscillations displayed a sequence of maxima with varying height that repeated every two or four maxima. Such a period doubling phenomenon commonly appears in non-linear systems of coupled or forced oscillators [6] and is a well studied route towards chaos [310]. Even though being investigated in great theoretical detail, period doubling has not been experimentally demonstrated in the context of synthetic

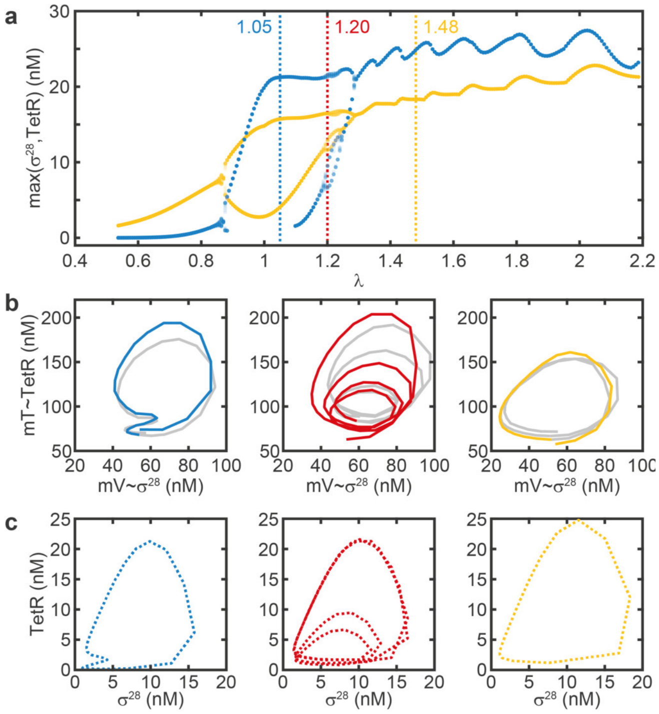


Figure 5.6: Bifurcation analysis for varying λ. **a)** Simulated bifurcation diagram showing the maxima for σ^{28} (yellow) and TetR (blue). The values of λ that correspond to the experimental data in (b) and simulations in (c) are indicated by dashed vertical lines. **b)** Experimental phase portraits exhibit a 2-, 4-, and 1-cycle. A second (third) m-cycle is grayed out for clarity. The 2-cycle is degenerate in the TetR-axis, i.e. the second maximum does not show in the time traces. Data for the 1-cycle and 4-cycle is the same as shown in Figure 5.4b and Figure 5.5, respectively. **c)** Simulated phase portraits. Parameter values are as in Figure 5.2, except $\alpha = 5$ pMs^{-1}, $n = 4$, $T_{in} = 4$ h, $A_{in} = 50$ nM.

biochemical oscillators so far [302]. Two of the experimental challenges in this context are that to record an m-cycle, the oscillator has to run reliably for $t > T_{in} \cdot 2 \cdot m = 32$ h (for $T_{in} = 4$ h and $m = 4$), and that for increasing m the bifurcation parameter λ has to be tuned with an exponentially increasing accuracy [311].

As shown in Figure 5.5a, we indeed find experimental evidence of period doubling in our system (here for $\lambda = 1.20$). After the typical large first maximum that occurs during the initial transient, the forced oscillator approaches a 4-cycle and stays there for two full revolutions. In the experiment, period doubling is more evident in the TetR dynamics than in the σ^{28} dynamics. After about 36 h the system appears to 'drop back' to a regular 1-cycle. As 36 h is close to the longest time span for which such a reactor was reportedly operated [89], this behavior is likely a sign of fatigue, which is also consistent with an observed drop of the refresh ratio towards the end of the recording.

We computed an 'instantaneous refresh ratio' $R_t = 1 - I_{t+1}/I_t$ (Figure 5.5b) using the reference time traces I_t for all time points t where $I_t > 0.3 \cdot I_{max}$. For $t \gtrsim 36$ h, R_t drops by about 1%, leading to a decrease in δ and a corresponding change in λ. Note that the instantaneous refresh ratio also slightly deviates from the nominal refresh ratio (here 20%) that was defined by calibration prior to the experiment. We hence use the more accurate instantaneous refresh ratio to calculate the control parameter λ. Similarly, a loss in activity of the supplied reagents would lead to a decrease in α over time, resulting in a stronger damping of the free oscillations.

Period doubling can further be visualized with a phase portrait (Figure 5.5c), which highlights that the trajectories return to their starting point in phase space after completion of four revolutions. Finally, we can generate a maximum return map (Figure 5.5d) by plotting the amplitude of each maximum against that of its predecessor. Again it can be observed that, within experimental variability, the system visits four distinct points in the map until it returns to its original location.

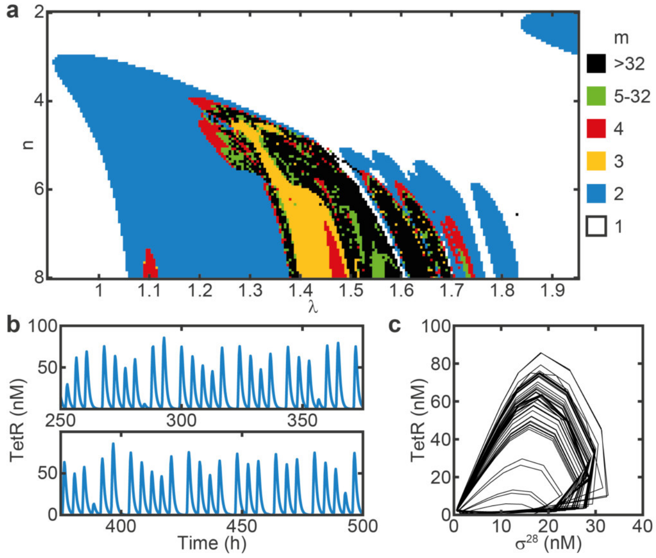


Figure 5.7: Simulations predict chaotic dynamics. a) Two-dimensional bifurcation diagram showing the rotation number m for varying n and λ. With increasing n, the dynamics become increasingly more complex, over a broader range of λ. **b)** Chaotic trajectory and **c)** phase portrait of the corresponding strange attractor for $n = 8$, $\lambda = 1.5542$. Parameter values, except n, are the same as in Figure 5.6.

5.2.3 Bifurcation Diagram

In order to gain a more complete overview of the dynamical repertoire of our biochemical oscillator, we simulated a bifurcation diagram (Figure 5.6a), for which we plotted the heights of the maxima against the parameter λ. Because successive bifurcations occur within exponentially decreasing intervals [311], we adjusted $\lambda(R)$ as smoothly as possible. Following Equation (5.9), we fixed the input period to $T_{in} = 4$ h, as it can only be adjusted in increments of $t_{int} \geq 15$ min, and instead varied the natural period T by adjusting the refresh ratio R, which can be varied, in principle, continuously.

The bifurcation analysis reveals that for low λ the system displays 1-cycles, then undergoes two period doublings to show a 4-cycle around $\lambda = 1.2$, followed by period 'halvings' and again 1-cycles at high values of λ. Qualitatively, the experimentally recorded phase space trajectories (Figure 5.6b) match the corresponding simulated trajectories for the same λ values (Figure 5.6c) remarkably well. One interesting feature is that for $\lambda = 1.05$ both experiment and simulation display a 2-cycle that is degenerate in the dynamics of TetR, i.e., the second maximum is not visible, but the period is doubled.

5.2.4 Chaotic Dynamics in the Oscillator Model

We were finally interested whether our system could, in principle, exhibit even more complex dynamics than a 4-cycle. We therefore simulated a two-dimensional bifurcation diagram, for which we varied both the Hill coefficient n and λ (Figure 5.7a). The system dynamics can then be classified by means of the rotation number m, which equals the number of periods the system undergoes before returning to the starting point. For a chaotic trajectory, $m = \infty$, but for practical reasons we classify trajectories as chaotic if $m > 32$.

As shown in Figure 5.7a, the oscillator model indeed permits chaotic solutions. Notably, chaotic regimes are interrupted by windows of mostly period 3, which is a commonly observed phenomenon [6]. The existence of $m = 3$-cycles actually implies the existence of chaotic trajectories [312], examples of which are shown in Figure 5.7b,c. Overall, this analysis reveals that higher order period doublings and chaotic behavior become increasingly prevalent for increasing non-linearity, corresponding to increasing Hill coefficients in the oscillator model.

5.3 Period Doubling in Natural and Synthetic Circadian Clocks

Inspired by the entrainment of biological circadian clocks by environmental zeitgeber cues, we have here experimentally investigated the response of a single-loop cell-free genetic oscillator to externally applied periodic perturbations. To this end, we utilized a microfluidic reactor system which allowed precisely controlled addition of components to the oscillator and dilution at regular intervals. The period of the free-running oscillator is dominated by the reactor's dilution rate δ, which defines a timescale $T \approx \frac{\pi}{\delta}$. External forcing of the oscillator is achieved by periodically supplying transcription factors or inducers with an input period T_{in}. The dynamic behavior of the forced oscillator is then determined by a single bifurcation parameter λ, which is defined as the ratio of the input period and the intrinsic timescale, i.e., $\lambda = T_{in}/T$. Supported by simulations, we experimentally found non-trivial behavior such as 2- and 4-cycles, which is indicative of period doubling. Complex dynamics were observed for λ close to 1, whereas higher values of $\lambda = 1.48$ lead to stable entrainment of the genetic oscillator to the zeitgeber timescale.

To our knowledge, period doubling has not been found in experiments with biological circadian clocks so far, but has been repeatedly predicted by theoretical models. In a theoretical study by Kurosawa and Goldbeter [300], a tendency towards complex dynamics for $\lambda \approx 1$ was found in a model of the *Neurospora* clock for strong forcing amplitudes. However, this behavior was not found in a model of the *Drosophila* circadian clock by the same authors. The authors attribute this to differences in the forcing mechanisms, which in *Neurospora* alters expression rates, whereas in *Drosophila* alters degradation rates. Also, the authors argue that the complexity observed for $\lambda \approx 1$ may explain why the free-running periods of many circadian rhythms differ substantially from a 24 h period. For instance, in humans, the free-running circadian period is 24 : 11 h ± 0 : 08 h (± SD), but in simpler organisms varies from 21.5 h ($\lambda = 1.12$) in *Neurospora* [313] to about 28 h ($\lambda = 0.86$) in *Phaseolus* [314].

Both from a biological point of view as well as for potential applications of synthetic biochemical clocks it might be desirable to actually prevent complex behavior resulting from coupled oscillator dynamics, which can be achieved in various ways. One possibility is to simply tune the free-running period away from the zeitgeber period, and thus avoid $\lambda \approx 1$.

We further notice from our simulated bifurcation diagram (Figure 5.6a) that as λ increases, the system initially undergoes two period doublings, but then does not continue to produce chaos, but follows with period halvings that eventually lead to stable 1-cycles. Similar period-halvings have been found in the study of 'single-humped' 1-dimensional maps [315], which are used as simple ecological models, when the recursion map was modified with a constant 'immigration' term that prevented the population from ever falling below a certain floor level. In the context of our biochemical oscillator, such a term would correspond to leaky/basal gene expression, potentially having a similar effect on the reversal of period-doubling transitions.

While in biological systems, biochemical rhythms are subject to evolutionary optimization, in synthetic biological systems robust behavior can be achieved by appropriately tuning system parameters such as the shape of the gene transfer functions [11]. Such engineering may be supported by *in silico* modeling, e.g., using evolutionary algorithms that intrinsically generate robust solutions [264], combined with high-throughput microfluidic reactors that enable comprehensive parameter screens [202].

In summary, we have shown that a synthetic cell-free gene circuit operated in a microfluidic reactor can be used to physically emulate the entrainment of a genetic oscillator with an external zeitgeber signal, which allowed the experimental realization of periodic doubling bifurcations, which had been previously only observed in numerical models of such systems. Apart from the fundamental interest in oscillatory biochemical systems, synthetic biochemical clocks may be of use in a wide range of applications that require intrinsic time measurements for the autonomous orchestration of downstream processes. In order to improve the accuracy of such oscillators and synchronize them to an external clock, coupling to a zeitgeber signal will be required. Our study demonstrates how the choice of system and control parameters can be used to tune the dynamics of such systems to become robust - or complex.

Conclusion

In this thesis, we utilized cell-free GRNs embedded in artificial cells to engineer synthetic biological self-organization processes. Anticipating the development of synthetic organisms, we identified and emulated three key aspects of complex biological organisms: spatial organization and hierarchy, pattern formation, and temporal synchronization.

The realization and characterization of such systems required the development of new design concepts and technologies. Programming gel organelles with dedicated functions to mimic eukaryotic transcription-translation processes required the development of compatible hydrogel, conjugate chemistry, and droplet microfluidics protocols. The quantitative analysis of pattern formation from morphogen gradients in assemblies of artificial cells required the development of accurate image processing routines, the adaptation of statistical methods to estimate positional information, and modeling to support the interpretation of our experimental results. The synchronization of a genetic oscillator to an external signal required the optimization of a microfluidic ring reactor to operate cell-free GRNs under steady-state conditions. Experiments had to be automated with custom electronic devices and software, and simulations were used to predict the conditions for complex non-linear dynamics. Hence, the main challenges that we have addressed concern the molecular implementation, the quantification, and the energy supply of self-organization phenomena.

The ultimate solution to the problem of energy supply will probably be a self-sufficient synthetic cell that is fed with sugars or light and serves as the basic building block of synthetic organisms. However, to this day even the self-regeneration of some of the transcription-translation machinery in a cell-free gene expression reaction represents a considerable challenge [21, 164, 316]. One way to circumvent this issue are systems where biochemical components are supplied continuously [22, 23]. Continuous systems allow the emulation of life-like out-of-equilibrium conditions in a simplified environment. In the case of pattern formation, continuous systems resemble mammalian embryos that are continuously supplied by their mother, whereas closed systems with a finite energy reservoir mimic eggs.

Continuous systems hence represent a bridge technology that enables the study of more complex cell-free GRNs until the problem of self-regeneration has been solved. For instance, we envision gel based microfluidic reactors that enable the emulation of pattern formation. While being open to energy supply, hydrogels provide the additional advantage that immobilization of GRN components can be used to modulate diffusion [200, 201]. This could enable Turing-like reaction-diffusion systems that actively generate positional information instead of interpreting existing morphogen gradients. Ultimately, PDMS-based microfluidics may be replaced by a synthetic vascular system that allows the supply of larger scale structures in a hierarchical manner, as is commonly found in biological structures such as muscles.

Another important direction for further development is the automated design and implementation of more complex GRNs. While automated design of logic circuits has been demonstrated [30], the design of dynamic functions still requires trial-and-error. The design-build-test cycle can be accelerated with rapid prototyping techniques [89], high-throughput methods to characterize large part libraries [202], and computational models to predict the behavior of a GRN *in silico*. Effective design requires methods to tune parameters, to deal with timing and noise issues, and to account for experimental constraints and tolerances.

To this end, we have conceptually distinguished between *tuning parameters* and (apparent) *system parameters*. We have shown that a simple GRN can generate complex non-linear dynamics by varying a single tuning parameter. In this case, the tuning parameter could be varied relatively accurately and mainly affected one system parameter. In general, however, varying a tuning parameter can change several system parameters, which, as shown in Section 1.1.4, creates experimental constraints. Additionally, the number of input nodes and the available types of input logic are limited by the molecular mechanisms. If properly identified and quantified, such constraints and uncertainties can be integrated into the design process. GRN topologies that are robust under these conditions can for instance be found with robust optimization algorithms. One such possibility are evolutionary algorithms that directly mimic the biological design process [263, 264].

Potential application areas of self-organizing artificial cells and synthetic organisms are theranostics, tissue engineering, and bio-materials. For instance, the targeted synthesis and release of a tumor drug using artificial cells has already been demonstrated in mice [156]. Implementing mechanisms with which the artificial cell could parasitize the hosts energy resources and synchronize to its circadian rhythm could result in autonomous devices that produce therapeutics at specific times. Engineering pattern formation can enable the regeneration of complex tissues, such as bone or muscle [317]. Similarly, smart functional hybrid materials could be generated that respond to environmental cues and self-organize according to a genetic program.

Another application of synthetic organisms is to study the fundamental principles that underlie biological self-organization phenomena [19]. The limited complexity of synthetic GRNs has so far forced synthetic biologists to focus on simple network motifs, which helped to identify the minimal set of core principles that is necessary to generate complex behavior [44]. For instance, quantifying the dynamic evolution of positional information in our synthetic system has highlighted the crucial importance of balancing timing, noise, and system parameters. Realizing complex dynamics in the simple forced genetic oscillator has suggested that the ratio of the free-running and the forcing period is of crucial importance for stable entrainment. The range of questions that can be studied with synthetic systems may expand as the complexity of synthetic GRNs increases. For example, this would allow to study more specific regulatory mechanisms and evaluate different GRN topologies. In any case, however, our humble attempts to create even simple systems that replicate functions of biological organisms remind us of the necessity to appreciate and conserve the beautiful complexity of life.

Bibliography

[1] E. Schrödinger. *What is Life? The Physical Aspect of the Living Cell*. Cambridge University Press, 1944.

[2] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. CRC Press, 2006.

[3] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. "Network motifs in the transcriptional regulation network of Escherichia coli". *Nature Genetics* 31.1 (2002), pp. 64–68. DOI: 10.1038/ng881.

[4] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. "Network Motifs: Simple Building Blocks of Complex Networks". *Science* 298.5594 (2002), pp. 824–827. DOI: 10.1126/science.298.5594.824.

[5] P. W. Anderson. "More Is Different". *Science* 177.4047 (1972), pp. 393–396.

[6] S. H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering.* Second Edition. Westview Press, 2014.

[7] N. Laohakunakorn, L. Grasemann, B. Lavickova, G. Michielin, A. Shahein, Z. Swank, and S. J. Maerkl. "Bottom-Up Construction of Complex Biomolecular Systems With Cell-Free Synthetic Biology". *Frontiers in Bioengineering and Biotechnology* 8 (2020), p. 213. DOI: 10.3389/fbioe.2020.00213.

[8] T. S. Gardner, C. R. Cantor, and J. J. Collins. "Construction of a genetic toggle switch in Escherichia coli". *Nature* 403.6767 (2000), pp. 339–342. DOI: 10.1038/35002131.

[9] R. Weiss and T. F. Knight Jr. "Engineered communications for microbial robotics". *International Workshop on DNA-Based Computers*. Springer, 2000, pp. 1–16. DOI: 10.1007/3-540-44992-2.

[10] M. B. Elowitz and S. Leibler. "A synthetic oscillatory network of transcriptional regulators". *Nature* 403.6767 (2000), pp. 335–338. DOI: 10.1038/35002125.

[11] J. A. N. Brophy and C. A. Voigt. "Principles of genetic circuit design". *Nature Methods* 11.5 (2014), pp. 508–520. DOI: 10.1038/nmeth.2926.

[12] G. M. Church, M. B. Elowitz, C. D. Smolke, C. A. Voigt, and R. Weiss. "Realizing the potential of synthetic biology". *Nature Reviews Molecular Cell Biology* 15.4 (2014), pp. 289–294. DOI: 10.1038/nrm3767.

[13] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss. "A synthetic multicellular system for programmed pattern formation". *Nature* 434.7037 (2005), pp. 1130–1134. DOI: 10.1038/nature03461.

[14] T. Danino, O. Mondragón-Palomino, L. Tsimring, and J. Hasty. "A synchronized quorum of genetic clocks". *Nature* 463.7279 (2010), pp. 326–330. DOI: 10.1038/nature08753.

[15] M. Weitz, A. Mückl, K. Kapsner, R. Berg, A. Meyer, and F. C. Simmel. "Communication and computation by bacteria compartmentalized within microemulsion droplets". *Journal of the American Chemical Society* 136.1 (2014), pp. 72–75. DOI: 10.1021/ja411132w.

[16] R. Lentini, N. Yeh Martín, and S. S. Mansy. "Communicating artificial cells". *Current Opinion in Chemical Biology* 34 (2016), pp. 53–61. DOI: 10.1016/j.cbpa.2016.06.013.

[17] J. W. Szostak, D. P. Bartel, and P. L. Luisi. "Synthesizing life". *Nature* 409.6818 (2001), pp. 387–390. DOI: 10.1038/35053176.

[18] P. L. Luisi, F. Ferri, and P. Stano. "Approaches to semi-synthetic minimal cells: a review". *Naturwissenschaften* 93.1 (2006), pp. 1–13. DOI: 10.1007/s00114-005-0056-z.

[19] A. Salehi-Reyhani, O. Ces, and Y. Elani. "Artificial cell mimics as simplified models for the study of cell biology". *Experimental Biology and Medicine* 242.13 (2017), pp. 1309–1317. DOI: `10.1177/1535370217711441`.

[20] V. Noireaux and A. Libchaber. "A vesicle bioreactor as a step toward an artificial cell assembly". *Proceedings of the National Academy of Sciences* 101.51 (2004), pp. 17669–17674. DOI: `10.1073/pnas.0408236101`.

[21] B. Lavickova, N. Laohakunakorn, and S. J. Maerkl. "A partially self-regenerating synthetic cell". *Nature Communications* 11.1 (2020), p. 6340. DOI: `10.1038/s41467-020-20180-6`.

[22] E. Karzbrun, A. M. Tayar, V. Noireaux, and R. H. Bar-Ziv. "Programmable on-chip DNA compartments as artificial cells". *Science* 345.6198 (2014), pp. 829–832. DOI: `10.1126/science.1255550`.

[23] H. Niederholtmeyer, V. Stepanova, and S. J. Maerkl. "Implementation of cell-free biological networks at steady state". *Proceedings of the National Academy of Sciences* 110.40 (2013), pp. 15985–15990. DOI: `10.1073/pnas.1311166110`.

[24] A. M. Tayar, E. Karzbrun, V. Noireaux, and R. H. Bar-Ziv. "Propagating gene expression fronts in a one-dimensional coupled system of artificial cells". *Nature Physics* 11.12 (2015), pp. 1037–1041. DOI: `10.1038/nphys3469`.

[25] M. Schwarz-Schilling, L. Aufinger, A. Mückl, and F. C. Simmel. "Chemical communication between bacteria and cell-free gene expression systems within linear chains of emulsion droplets". *Integrative Biology* 8.4 (2016), pp. 564–570. DOI: `10.1039/C5IB00301F`.

[26] R. Lentini, N. Y. Martín, M. Forlin, L. Belmonte, J. Fontana, M. Cornella, L. Martini, S. Tamburini, W. E. Bentley, O. Jousson, and S. S. Mansy. "Two-Way Chemical Communication between Artificial and Natural Cells". *ACS Central Science* (2017). DOI: `10.1021/acscentsci.6b00330`.

[27] M. Isalan, C. Lemerle, and L. Serrano. "Engineering Gene Networks to Emulate Drosophila Embryonic Pattern Formation". *PLoS Biology* 3.3 (2005), e64. DOI: `10.1371/journal.pbio.0030064`.

[28] A. M. Tayar, E. Karzbrun, V. Noireaux, and R. H. Bar-Ziv. "Synchrony and pattern formation of coupled genetic oscillators on a chip of artificial cells". *Proceedings of the National Academy of Sciences* 114.44 (2017), pp. 11609–11614. DOI: `10.1073/pnas.1710620114`.

[29] A. Dupin and F. C. Simmel. "Signalling and differentiation in emulsion-based multi-compartmentalized in vitro gene circuits". *Nature Chemistry* 11.1 (2018), pp. 32–39. DOI: `10.1038/s41557-018-0174-9`.

[30] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt. "Genetic circuit design automation". *Science* 352.6281 (2016), aac7341 1–11. DOI: `10.1126/science.aac7341`.

[31] S. Rollié, M. Mangold, and K. Sundmacher. "Designing biological systems: Systems engineering meets synthetic biology". *Chemical Engineering Science* 69.1 (2012), pp. 1–29. DOI: `10.1016/j.ces.2011.10.068`.

[32] J. Santos-Moreno and Y. Schaerli. "Using Synthetic Biology to Engineer Spatial Patterns". *Advanced Biosystems* 3.4 (2019), p. 1800280. DOI: `10.1002/adbi.201800280`.

[33] B. Novák and J. J. Tyson. "Design principles of biochemical oscillators". *Nature Reviews Molecular Cell Biology* 9.12 (2008), pp. 981–991. DOI: `10.1038/nrm2530`.

[34] A. M. Tayar, S. S. Daube, and R. H. Bar-Ziv. "Progress in programming spatiotemporal patterns and machine-assembly in cell-free protein expression systems". *Current Opinion in Chemical Biology* 40 (2017), pp. 37–46. DOI: `10.1016/j.cbpa.2017.05.005`.

[35] N. S. Scholes, D. Schnoerr, M. Isalan, and M. P. Stumpf. "A Comprehensive Network Atlas Reveals That Turing Patterns Are Common but Not Robust". *Cell Systems* 9.3 (2019), 243–257.e4. DOI: `10.1016/j.cels.2019.07.007`.

[36] L. Aufinger and F. C. Simmel. "Establishing communication between artificial cells". *Chemistry - A European Journal* 25.55 (2019), pp. 12659–12670. DOI: `10.1002/chem.201901726`.

[37] L. Aufinger and F. C. Simmel. “Artificial Gel-based Organelles for Spatial Organization of Cell-free Gene Expression Reactions”. *Angewandte Chemie International Edition* 57.52 (2018), pp. 17245–17248. DOI: `10.1002/anie.201809374`.

[38] A. Dupin, L. Aufinger, I. Styazhkin, F. Rothfischer, B. Kaufmann, S. Schwarz, N. Galensowske, H. Clausen-Schaumann, and F. C. Simmel. “Synthetic cell-based materials extract positional information from morphogen gradients”. *bioRxiv* (preprint). DOI: `10.1101/2021.04.25.441320`.

[39] L. Aufinger, J. Brenner, and F. C. Simmel. “Period Doubling Bifurcations in a Forced Cell-Free Genetic Oscillator”. *bioRxiv* (preprint). DOI: `10.1101/2021.12.15.472802`.

[40] J. Jaeger. “The gap gene network”. *Cellular and Molecular Life Sciences* 68.2 (2011), pp. 243–274. DOI: `10.1007/s00018-010-0536-y`.

[41] A. Eldar, R. Dorfman, D. Weiss, H. Ashe, B.-Z. Shilo, and N. Barkai. “Robustness of the BMP morphogen gradient in Drosophila embryonic patterning”. *Nature* 419.6904 (2002), pp. 304–308. DOI: `10.1038/nature01061`.

[42] N. Barkai and B.-Z. Shilo. “Robust Generation and Decoding of Morphogen Gradients”. *Cold Spring Harbor Perspectives in Biology* 1.5 (2009), a001990–a001990. DOI: `10.1101/cshperspect.a001990`.

[43] O. Barad, D. Rosin, E. Hornstein, and N. Barkai. “Error Minimization in Lateral Inhibition Circuits”. *Science Signaling* 3.129 (2010), ra51–ra51. DOI: `10.1126/scisignal.2000857`.

[44] Y. Schaerli, A. Munteanu, M. Gili, J. Cotterell, J. Sharpe, and M. Isalan. “A unified design space of synthetic stripe-forming networks”. *Nature Communications* 5.1 (2014), p. 4905. DOI: `10.1038/ncomms5905`.

[45] L. Potvin-Trottier, N. D. Lord, G. Vinnicombe, and J. Paulsson. “Synchronous long-term oscillations in a synthetic gene circuit”. *Nature* 538.7626 (2016), pp. 514–517. DOI: `10.1038/nature19841`.

[46] A. S. Zadorin, Y. Rondelez, G. Gines, V. Dilhas, G. Urtel, A. Zambrano, J.-C. Galas, and A. Estevez-Torres. “Synthesis and materialization of a reaction–diffusion French flag pattern”. *Nature Chemistry* 9.10 (2017), pp. 990–996. DOI: `10.1038/nchem.2770`.

[47] I. Barbier, R. Perez-Carrasco, and Y. Schaerli. “Controlling spatiotemporal pattern formation in a concentration gradient with a synthetic toggle switch”. *Molecular Systems Biology* 16.6 (2020), e9361. DOI: `10.15252/msb.20199361`.

[48] D. Soloveichik, G. Seelig, and E. Winfree. “DNA as a universal substrate for chemical kinetics”. *Proceedings of the National Academy of Sciences* 107.12 (2010), pp. 5393–5398. DOI: `10.1073/pnas.0909380107`.

[49] J. Kim, K. S. White, and E. Winfree. “Construction of an in vitro bistable circuit from synthetic transcriptional switches”. *Molecular Systems Biology* 2.1 (2006), p. 68. DOI: `10.1038/msb4100099`.

[50] 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 (2000), pp. 605–608. DOI: `10.1038/35020524`.

[51] D. Y. Zhang and E. Winfree. “Control of DNA Strand Displacement Kinetics Using Toehold Exchange”. *Journal of the American Chemical Society* 131.47 (2009), pp. 17303–17314. DOI: `10.1021/ja906987s`.

[52] F. C. Simmel, B. Yurke, and H. R. Singh. “Principles and Applications of Nucleic Acid Strand Displacement Reactions”. *Chemical Reviews* 119.10 (2019), pp. 6326–6369. DOI: `10.1021/acs.chemrev.8b00580`.

[53] J. Kim and E. Winfree. “Synthetic in vitro transcriptional oscillators”. *Molecular Systems Biology* 7.1 (2011), pp. 465–465. DOI: `10.1038/msb.2010.119`.

[54] Z. Z. Sun, C. A. Hayes, J. Shin, F. Caschera, R. M. Murray, and V. Noireaux. “Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology”. *Journal of Visualized Experiments* 79 (2013). DOI: `10.3791/50762`.

[55] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda. “Cell-free translation reconstituted with purified components”. *Nature biotechnology* 19.8 (2001), pp. 751–755. DOI: `10.1038/90802`.

[56] Y. Shimizu, T. Kanamori, and T. Ueda. “Protein synthesis by pure translation systems”. *Methods* 36.3 (2005), pp. 299–304. DOI: 10.1016/j.ymeth.2005.04.006.

[57] Z. Z. Sun, E. Yeung, C. A. Hayes, V. Noireaux, and R. M. Murray. “Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system”. *ACS Synthetic Biology* 3.6 (2014), pp. 387–397. DOI: 10.1021/sb400131a.

[58] B. Lavickova and S. J. Maerkl. “A Simple, Robust, and Low-Cost Method To Produce the PURE Cell-Free System”. *ACS Synthetic Biology* 8.2 (2019), pp. 455–462. DOI: 10.1021/acssynbio.8b00427.

[59] D. Foshag, E. Henrich, E. Hiller, M. Schäfer, C. Kerger, A. Burger-Kentischer, I. Diaz-Moreno, S. M. García-Mauriño, V. Dötsch, S. Rupp, and F. Bernhard. “The E. coli S30 lysate proteome: A prototype for cell-free protein production”. *New Biotechnology* 40 (2018), pp. 245–260. DOI: 10.1016/j.nbt.2017.09.005.

[60] P. van Nies, I. Westerlaken, D. Blanken, M. Salas, M. Mencía, and C. Danelon. “Self-replication of DNA by its encoded proteins in liposome-based synthetic cells”. *Nature Communications* 9.1 (2018), pp. 1–12. DOI: 10.1038/s41467-018-03926-1.

[61] J. Shin, P. Jardine, and V. Noireaux. “Genome Replication, Synthesis, and Assembly of the Bacteriophage T7 in a Single Cell-Free Reaction”. *ACS Synthetic Biology* 1.9 (2012), pp. 408–413. DOI: 10.1021/sb300049p.

[62] D.-M. Kim and J. R. Swartz. “Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis”. *Biotechnology and bioengineering* 74.4 (2001), pp. 309–316. DOI: 10.1002/bit.1121.

[63] F. Caschera and V. Noireaux. “Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription–translation system”. *Biochimie* 99 (2014), pp. 162–168. DOI: 10.1016/j.biochi.2013.11.025.

[64] F. Caschera and V. Noireaux. “A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system”. *Metabolic Engineering* 27 (2015), pp. 29–37. DOI: 10.1016/j.ymben.2014.10.007.

[65] D. C. Garcia, J. L. N. Dinglasan, H. Shrestha, P. E. Abraham, R. L. Hettich, and M. J. Doktycz. “A lysate proteome engineering strategy for enhancing cell-free metabolite production”. *Metabolic engineering communications* 12 (2021), e00162. DOI: 10.1016/j.mec.2021.e00162.

[66] J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce. “NUPACK: Analysis and design of nucleic acid systems”. *Journal of Computational Chemistry* 32.1 (2011), pp. 170–173. DOI: 10.1002/jcc.21596.

[67] G. Seelig, D. Soloveichik, D. Y. Zhang, and E. Winfree. “Enzyme-free nucleic acid logic circuits”. *Science* 314.5805 (2006), pp. 1585–1588. DOI: 10.1126/science.1132493.

[68] L. Qian, E. Winfree, and J. Bruck. “Neural network computation with DNA strand displacement cascades”. *Nature* 475.7356 (2011), pp. 368–372. DOI: 10.1038/nature10262.

[69] L. Qian and E. Winfree. “Scaling up digital circuit computation with DNA strand displacement cascades”. *Science* 332.6034 (2011), pp. 1196–1201. DOI: 10.1126/science.1200520.

[70] B. Wang, C. Thachuk, A. D. Ellington, E. Winfree, and D. Soloveichik. “Effective design principles for leakless strand displacement systems”. *Proceedings of the National Academy of Sciences* 115.52 (2018), E12182–E12191. DOI: 10.1073/pnas.1806859115.

[71] J. Kim, J. Hopfield, and E. Winfree. “Neural network computation by in vitro transcriptional circuits”. *Advances in neural information processing systems*. Vol. 17. 2004, pp. 681–688.

[72] E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree, and F. C. Simmel. “Timing molecular motion and production with a synthetic transcriptional clock”. *Proceedings of the National Academy of Sciences* 108.40 (2011), E784–E793. DOI: 10.1073/pnas.1100060108.

[73] L. N. Green, H. K. K. Subramanian, V. Mardanlou, J. Kim, R. F. Hariadi, and E. Franco. “Autonomous dynamic control of DNA nanostructure self-assembly”. *Nature Chemistry* 11.6 (2019), pp. 510–520. DOI: 10.1038/s41557-019-0251-8.

[74] M. Weitz, J. Kim, K. Kapsner, E. Winfree, E. Franco, and F. C. Simmel. “Diversity in the dynamical behaviour of a compartmentalized programmable biochemical oscillator”. *Nature Chemistry* 6.4 (2014), pp. 295–302. DOI: 10.1038/nchem.1869.

[75] A. J. Meyer, J. W. Ellefson, and A. D. Ellington. “Directed evolution of a panel of orthogonal T7 RNA polymerase variants for *in vivo* or *in vitro* synthetic circuitry”. *ACS Synthetic Biology* 4.10 (2015), pp. 1070–1076. DOI: 10.1021/sb500299c.

[76] J. Lloyd, C. H. Tran, K. Wadhwani, C. Cuba Samaniego, H. K. K. Subramanian, and E. Franco. “Dynamic Control of Aptamer–Ligand Activity Using Strand Displacement Reactions”. *ACS Synthetic Biology* 7.1 (2018), pp. 30–37. DOI: 10.1021/acssynbio.7b00277.

[77] S. W. Schaffter, L. N. Green, J. Schneider, H. K. K. Subramanian, R. Schulman, and E. Franco. “T7 RNA polymerase non-specifically transcribes and induces disassembly of DNA nanostructures”. *Nucleic Acids Research* 46.10 (2018), pp. 5332–5343. DOI: 10.1093/nar/gky283.

[78] F. J. Triana-Alonso, M. Dabrowski, J. Wadzack, and K. H. Nierhaus. “Self-coded 3’-extension of run-off transcripts produces aberrant products during in vitro transcription with T7 RNA polymerase”. *Journal of Biological Chemistry* 270.11 (1995), pp. 6298–6307. DOI: 10.1074/jbc.270.11.6298.

[79] Y. Gholamalipour, A. Karunanayake Mudiyanselage, and C. T. Martin. “3’ end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character - RNA-Seq analyses”. *Nucleic Acids Research* 46.18 (2018), pp. 9253–9263. DOI: 10.1093/nar/gky796.

[80] S. W. Schaffter and R. Schulman. “Building in vitro transcriptional regulatory networks by successively integrating multiple functional circuit modules”. *Nature Chemistry* 11.9 (2019), pp. 829–838. DOI: 10.1038/s41557-019-0292-z.

[81] H. Wu, T. Wei, B. Yu, R. Cheng, F. Huang, X. Lu, Y. Yan, X. Wang, C. Liu, and B. Zhu. “A single mutation attenuates both the transcription termination and RNA-dependent RNA polymerase activity of T7 RNA polymerase”. *RNA biology* 18.1 (2021), pp. 451–466. DOI: 10.1080/15476286.2021.1954808.

[82] K. Pardee, S. Slomovic, P. Q. Nguyen, J. W. Lee, N. Donghia, D. Burrill, T. Ferrante, F. R. McSorley, Y. Furuta, A. Vernet, M. Lewandowski, C. N. Boddy, N. S. Joshi, and J. J. Collins. “Portable, On-Demand Biomolecular Manufacturing”. *Cell* 167.1 (2016), 248–259.e12. DOI: 10.1016/j.cell.2016.09.013.

[83] K. Pardee, A. A. Green, M. K. Takahashi, D. Braff, G. Lambert, J. W. Lee, T. Ferrante, D. Ma, N. Donghia, M. Fan, N. M. Daringer, I. Bosch, D. M. Dudley, D. H. O’Connor, L. Gehrke, and J. J. Collins. “Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components”. *Cell* 165.5 (2016), pp. 1255–1266. DOI: 10.1016/j.cell.2016.04.059.

[84] L. Kai and P. Schwille. “Cell-Free Protein Synthesis and Its Perspectives for Assembling Cells from the Bottom-Up”. *Advanced Biosystems* 3.6 (2019), p. 1800322. DOI: 10.1002/adbi.201800322.

[85] R. D. Finn, E. V. Orlova, B. Gowen, M. Buck, and M. van Heel. “Escherichia coli RNA polymerase core and holoenzyme structures”. *The EMBO Journal* 19.24 (2000), pp. 6833–6844. DOI: 10.1093/emboj/19.24.6833.

[86] F. Caschera, M. A. Bedau, A. Buchanan, J. Cawse, D. de Lucrezia, G. Gazzola, M. M. Hanczyc, and N. H. Packard. “Coping with complexity: Machine learning optimization of cell-free protein synthesis”. *Biotechnology and Bioengineering* 108.9 (2011), pp. 2218–2228. DOI: 10.1002/bit.23178.

[87] V. Noireaux, R. Bar-Ziv, and A. Libchaber. “Principles of cell-free genetic circuit assembly”. *Proceedings of the National Academy of Sciences* 100.22 (2003), pp. 12672–12677. DOI: 10.1073/pnas.2135496100.

[88] M. K. Takahashi, C. A. Hayes, J. Chappell, Z. Z. Sun, R. M. Murray, V. Noireaux, and J. B. Lucks. “Characterizing and prototyping genetic networks with cell-free transcription–translation reactions”. *Methods* 86 (2015), pp. 60–72. DOI: 10.1016/j.ymeth.2015.05.020.

[89] H. Niederholtmeyer, Z. Z. Sun, Y. Hori, E. Yeung, A. Verpoorte, R. M. Murray, and S. J. Maerkl. “Rapid cell-free forward engineering of novel genetic ring oscillators”. *eLife* 4 (2015), e09771. DOI: 10.7554/eLife.09771.

[90] M. Yelleswarapu, A. J. van der Linden, B. van Sluijs, P. A. Pieters, E. Dubuc, T. F. A. de Greef, and W. T. S. Huck. “Sigma Factor-Mediated Tuning of Bacterial Cell-Free Synthetic Genetic Oscillators”. *ACS Synthetic Biology* 7.12 (2018), pp. 2879–2887. DOI: 10.1021/acssynbio.8b00300.

[91] P. A. Pieters, B. L. Nathalia, A. J. van der Linden, P. Yin, J. Kim, W. T. Huck, and T. F. A. de Greef. “Cell-Free Characterization of Coherent Feed-Forward Loop-Based Synthetic Genetic Circuits”. *ACS Synthetic Biology* 10.6 (2021), pp. 1406–1416. DOI: 10.1101/2021.01.11.426179.

[92] Z. Swank, N. Laohakunakorn, and S. J. Maerkl. “Cell-free gene regulatory network engineering with synthetic transcription factors”. *Proceedings of the National Academy of Sciences* 116.13 (2019), pp. 5892–5901. DOI: 10.1073/pnas.1816591116.

[93] J. Shin and V. Noireaux. “An *E. coli* cell-free expression toolbox: Application to synthetic gene circuits and artificial cells”. *ACS Synthetic Biology* 1.1 (2012), pp. 29–41. DOI: 10.1021/sb200016s.

[94] J. Garamella, R. Marshall, M. Rustad, and V. Noireaux. “The All *E. coli* TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology”. *ACS Synthetic Biology* 5.4 (2016), pp. 344–355. DOI: 10.1021/acssynbio.5b00296.

[95] D. Garenne, S. Thompson, A. Brisson, A. Khakimzhan, and V. Noireaux. “The all-E. coli TXTL toolbox 3.0: new capabilities of a cell-free synthetic biology platform”. *Synthetic Biology* 6.1 (2021), ysab017. DOI: 10.1093/synbio/ysab017.

[96] E. Karzbrun, J. Shin, R. H. Bar-Ziv, and V. Noireaux. “Coarse-grained dynamics of protein synthesis in a cell-free system”. *Physical Review Letters* 106.4 (2011), p. 048104. DOI: 10.1103/PhysRevLett.106.048104.

[97] R. Marshall and V. Noireaux. “Quantitative modeling of transcription and translation of an all-E. coli cell-free system”. *Scientific Reports* 9.1 (2019), p. 11980. DOI: 10.1038/s41598-019-48468-8.

[98] M. K. Takahashi, J. Chappell, C. A. Hayes, Z. Z. Sun, J. Kim, V. Singhal, K. J. Spring, S. Al-Khabouri, C. P. Fall, V. Noireaux, R. M. Murray, and J. B. Lucks. “Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription–Translation (TX-TL) Systems”. *ACS Synthetic Biology* 4.5 (2015), pp. 503–515. DOI: 10.1021/sb400206c.

[99] S. S. Yim, N. I. Johns, V. Noireaux, and H. H. Wang. “Protecting linear DNA templates in cell-free expression systems from diverse bacteria”. *ACS Synthetic Biology* 9.10 (2020), pp. 2851–2855. DOI: 10.1021/acssynbio.0c00277.

[100] R. Marshall, C. S. Maxwell, S. P. Collins, C. L. Beisel, and V. Noireaux. “Short DNA Containing *X* Sites Enhances DNA Stability and Gene Expression in E. coli Cell-Free Transcription–Translation Systems”. *Biotechnology and Bioengineering* 114.9 (2017), p. 6. DOI: 10.1002/bit.26333.

[101] D. Wigley. “Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB”. *Nature Reviews Microbiology* 11.1 (2013), pp. 9–13. DOI: 10.1038/nrmicro2917.

[102] D. Lewis, P. Le, C. Zurla, L. Finzi, and S. Adhya. “Multilevel autoregulation of repressor protein CI by DNA looping in vitro”. *Proceedings of the National Academy of Sciences* 108.36 (2011), pp. 14807–14812. DOI: 10.1073/pnas.1111221108.

[103] A. A. Green, P. A. Silver, J. J. Collins, and P. Yin. “Toehold Switches: De-Novo-Designed Regulators of Gene Expression”. *Cell* 159.4 (2014), pp. 925–939. DOI: 10.1016/j.cell.2014.10.002.

[104] B. Alberts, J. Lewis, D. Morgan, M. Raff, K. Roberts, P. Walter, A. D. Johnson, U. Schäfer, and C. Horstmann. *Molekularbiologie der Zelle*. John Wiley & Sons, Incorporated, 2017.

[105] J. L. Oakley, R. E. Strothkamp, A. H. Sarris, and J. E. Coleman. “T7 RNA polymerase: promoter structure and polymerase binding”. *Biochemistry* 18.3 (1979), pp. 528–537. DOI: 10.1021/bi00570a023.

[106] M. L. Urbanowski, C. P. Lostroh, and E. P. Greenberg. “Reversible acyl-homoserine lactone binding to purified *vibrio fischeri* LuxR protein”. *Journal of Bacteriology* 186.3 (2004), pp. 631–637. DOI: 10.1128/JB.186.3.631-637.2004.

[107] L. C. M. Antunes, R. B. R. Ferreira, C. P. Lostroh, and E. P. Greenberg. “A Mutational Analysis Defines Vibrio fischeri LuxR Binding Sites”. *Journal of Bacteriology* 190.13 (2008), pp. 4392–4397. DOI: 10.1128/JB.01443-07.

[108] E. P. Hamilton and N. Lee. “Three binding sites for AraC protein are required for autoregulation of araC in Escherichia coli.” *Proceedings of the National Academy of Sciences* 85.6 (1988), pp. 1749–1753. DOI: `10.1073/pnas.85.6.1749`.

[109] W. Hillen and C. Berens. “Mechanisms underlying expression of Tn 10 encoded tetracycline resistance”. *Annual Review of Microbiology* 48.1 (1994), pp. 345–369. DOI: `10.1146/annurev.mi.48.100194.002021`.

[110] S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, and B. Müller-Hill. “Quality and position of the three lac operators of E. coli define efficiency of repression.” *The EMBO Journal* 13.14 (1994), pp. 3348–3355. DOI: `10.1002/j.1460-2075.1994.tb06637.x`.

[111] S. L. Barnes, N. M. Belliveau, W. T. Ireland, J. B. Kinney, and R. Phillips. “Mapping DNA sequence to transcription factor binding energy in vivo”. *PLOS Computational Biology* 15.2 (2019), e1006226. DOI: `10.1371/journal.pcbi.1006226`.

[112] A. Sarai and Y. Takeda. “Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically.” *Proceedings of the National Academy of Sciences* 86.17 (1989), pp. 6513–6517. DOI: `10.1073/pnas.86.17.6513`.

[113] J. C. van Kessel, L. E. Ulrich, I. B. Zhulin, and B. L. Bassler. “Analysis of Activator and Repressor Functions Reveals the Requirements for Transcriptional Control by LuxR, the Master Regulator of Quorum Sensing in Vibrio harveyi”. *mBio* 4.4 (2013), e00378–13. DOI: `10.1128/mBio.00378-13`.

[114] B. J. Hussey and D. R. McMillen. “Programmable T7-based synthetic transcription factors”. *Nucleic Acids Research* 46.18 (2018), pp. 9842–9854. DOI: `10.1093/nar/gky785`.

[115] J. K. Jung, K. K. Alam, M. S. Verosloff, D. A. Capdevila, M. Desmau, P. R. Clauer, J. W. Lee, P. Q. Nguyen, P. A. Pastén, S. J. Matiasek, J.-F. Gaillard, D. P. Giedroc, J. J. Collins, and J. B. Lucks. “Cell-free biosensors for rapid detection of water contaminants”. *Nature Biotechnology* 38.12 (2020), pp. 1451–1459. DOI: `10.1038/s41587-020-0571-7`.

[116] R. R. Beerli and C. F. Barbas. “Engineering polydactyl zinc-finger transcription factors”. *Nature Biotechnology* 20.2 (2002), pp. 135–141. DOI: `10.1038/nbt0202-135`.

[117] M. J. Moscou and A. J. Bogdanove. “A Simple Cipher Governs DNA Recognition by TAL Effectors”. *Science* 326.5959 (2009), pp. 1501–1501. DOI: `10.1126/science.1178817`.

[118] R. Morbitzer, P. Romer, J. Boch, and T. Lahaye. “Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors”. *Proceedings of the National Academy of Sciences* 107.50 (2010), pp. 21617–21622. DOI: `10.1073/pnas.1013133107`.

[119] M. H. Larson, L. A. Gilbert, X. Wang, W. A. Lim, J. S. Weissman, and L. S. Qi. “CRISPR interference (CRISPRi) for sequence-specific control of gene expression”. *Nature Protocols* 8.11 (2013), pp. 2180–2196. DOI: `10.1038/nprot.2013.132`.

[120] W. Ji, D. Lee, E. Wong, P. Dadlani, D. Dinh, V. Huang, K. Kearns, S. Teng, S. Chen, J. Haliburton, G. Heimberg, B. Heineike, A. Ramasubramanian, T. Stevens, K. J. Helmke, V. Zepeda, L. S. Qi, and W. A. Lim. “Specific Gene Repression by CRISPRi System Transferred through Bacterial Conjugation”. *ACS Synthetic Biology* 3.12 (2014), pp. 929–931. DOI: `10.1021/sb500036q`.

[121] A. A. Nielsen and C. A. Voigt. “Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks”. *Molecular Systems Biology* 10.11 (2014), pp. 763–763. DOI: `10.15252/msb.20145735`.

[122] A. K. Brödel, R. Rodrigues, A. Jaramillo, and M. Isalan. “Accelerated evolution of a minimal 63–amino acid dual transcription factor”. *Science Advances* 6.24 (2019), eaba2728. DOI: `10.1101/725739`.

[123] R. E. Rondon, T. M. Groseclose, A. E. Short, and C. J. Wilson. “Transcriptional programming using engineered systems of transcription factors and genetic architectures”. *Nature Communications* 10.1 (2019), p. 4784. DOI: `10.1038/s41467-019-12706-4`.

[124] R. Lutz and H. Bujard. “Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements”. *Nucleic Acids Research* 25.6 (1997), pp. 1203–1210. DOI: `10.1093/nar/25.6.1203`.

[125] H. Niederholtmeyer, C. Chaggan, and N. K. Devaraj. “Communication and quorum sensing in non-living mimics of eukaryotic cells”. *Nature communications* 9.1 (2018), pp. 1–8. DOI: `10.1038/s41467-018-07473-7`.

[126] S. A. Lesley and R. R. Burgess. “Characterization of the Escherichia coli transcription factor σ^{70}: localization of a region involved in the interaction with core RNA polymerase”. *Biochemistry* 28.19 (1989), pp. 7728–7734. DOI: `10.1021/bi00445a031`.

[127] Y. Wang and P. L. deHaseth. “Sigma 32-Dependent Promoter Activity In Vivo: Sequence Determinants of the *groE* Promoter”. *Journal of Bacteriology* 185.19 (2003), pp. 5800–5806. DOI: `10.1128/JB.185.19.5800-5806.2003`.

[128] D. N. Arnosti and M. J. Chamberlin. “Secondary σ factor controls transcription of flagellar and chemotaxis genes in Escherichia coli”. *Proceedings of the National Academy of Sciences* 86.3 (1989), pp. 830–834. DOI: `10.1073/pnas.86.3.830`.

[129] M. Buck and W. Cannon. “Specific binding of the transcription factor sigma-54 to promoter DNA”. *Nature* 358.6385 (1992), pp. 422–422. DOI: `10.1038/358422a0`.

[130] H. Maeda, N. Fujita, and A. Ishihama. “Competition among seven Escherichia coli σ subunits: relative binding affinities to the core RNA polymerase”. *Nucleic Acids Research* 28.18 (2000), pp. 3497–3503. DOI: `10.1093/nar/28.18.3497`.

[131] H. M. Salis, E. A. Mirsky, and C. A. Voigt. “Automated design of synthetic ribosome binding sites to control protein expression”. *Nature Biotechnology* 27.10 (2009), pp. 946–950. DOI: `10.1038/nbt.1568`.

[132] M. Wachsmuth, S. Findeiss, N. Weissheimer, P. F. Stadler, and M. Morl. “De novo design of a synthetic riboswitch that regulates transcription termination”. *Nucleic Acids Research* 41.4 (2013), pp. 2541–2551. DOI: `10.1093/nar/gks1330`.

[133] A. Espah Borujeni, D. M. Mishler, J. Wang, W. Huso, and H. M. Salis. “Automated physics-based design of synthetic riboswitches from diverse RNA aptamers”. *Nucleic Acids Research* 44.1 (2016), pp. 1–13. DOI: `10.1093/nar/gkv1289`.

[134] G. Domin, S. Findeiß, M. Wachsmuth, S. Will, P. F. Stadler, and M. Mörl. “Applicability of a computational design approach for synthetic riboswitches”. *Nucleic Acids Research* (2016), gkw1267. DOI: `10.1093/nar/gkw1267`.

[135] J. Kim, Y. Zhou, P. D. Carlson, M. Teichmann, S. Chaudhary, F. C. Simmel, P. A. Silver, J. J. Collins, J. B. Lucks, P. Yin, and A. A. Green. “De novo-designed translation-repressing riboregulators for multi-input cellular logic”. *Nature Chemical Biology* 15.12 (2019), pp. 1173–1182. DOI: `10.1038/s41589-019-0388-1`.

[136] K. Pardee, A. A. Green, T. Ferrante, D. E. Cameron, A. DaleyKeyser, P. Yin, and J. J. Collins. “Paper-Based Synthetic Gene Networks”. *Cell* 159.4 (2014), pp. 940–954. DOI: `10.1016/j.cell.2014.10.004`.

[137] F. Hong, D. Ma, K. Wu, L. A. Mina, R. C. Luiten, Y. Liu, H. Yan, and A. A. Green. “Precise and Programmable Detection of Mutations Using Ultraspecific Riboregulators”. *Cell* 180.5 (2020), 1018–1032.e16. DOI: `10.1016/j.cell.2020.02.011`.

[138] A. Senoussi, J. Lee Tin Wah, Y. Shimizu, J. Robert, A. Jaramillo, S. Findeiss, I. M. Axmann, and A. Estevez-Torres. “Quantitative Characterization of Translational Riboregulators Using an *in Vitro* Transcription–Translation System”. *ACS Synthetic Biology* 7.5 (2018), pp. 1269–1278. DOI: `10.1021/acssynbio.7b00387`.

[139] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips. “Transcriptional regulation by the numbers: applications”. *Current Opinion in Genetics & Development* 15.2 (2005), pp. 125–135. DOI: `10.1016/j.gde.2005.02.006`.

[140] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. “Transcriptional regulation by the numbers: models”. *Current Opinion in Genetics & Development* 15.2 (2005), pp. 116–124. DOI: `10.1016/j.gde.2005.02.007`.

[141] K. Sneppen and G. Zocchi. *Physics in molecular biology*. Cambridge University Press, 2005.

[142] A. Vigouroux, E. Oldewurtel, L. Cui, D. Bikard, and S. van Teeffelen. “Tuning dCas9’s ability to block transcription enables robust, noiseless knockdown of bacterial genes”. *Molecular Systems Biology* 14.3 (2018), e7899. DOI: `10.15252/msb.20177899`.

[143] A. Tamsir, J. J. Tabor, and C. A. Voigt. “Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’”. *Nature* 469.7329 (2011), pp. 212–215. DOI: `10.1038/nature09565`.

[144] R. S. Cox, M. G. Surette, and M. B. Elowitz. “Programming gene expression with combinatorial promoters”. *Molecular Systems Biology* 3.1 (2007), p. 145. DOI: `10.1038/msb4100187`.

[145] Y. Chen, J. M. L. Ho, D. L. Shis, C. Gupta, J. Long, D. S. Wagner, W. Ott, K. Josić, and M. R. Bennett. “Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors”. *Nature Communications* 9.1 (2018), p. 64. DOI: `10.1038/s41467-017-02473-5`.

[146] R. Daniel, J. R. Rubens, R. Sarpeshkar, and T. K. Lu. “Synthetic analog computation in living cells”. *Nature* 497.7451 (2013), pp. 619–623. DOI: `10.1038/nature12148`.

[147] N. E. Buchler, U. Gerland, and T. Hwa. “On schemes of combinatorial transcription logic”. *Proceedings of the National Academy of Sciences* 100.9 (2003), pp. 5136–5141. DOI: `10.1073/pnas.0930314100`.

[148] N. E. Buchler and M. Louis. “Molecular Titration and Ultrasensitivity in Regulatory Networks”. *Journal of Molecular Biology* 384.5 (2008), pp. 1106–1119. DOI: `10.1016/j.jmb.2008.09.079`.

[149] N. E. Buchler and F. R. Cross. “Protein sequestration generates a flexible ultrasensitive response in a genetic network”. *Molecular Systems Biology* 5.1 (2009), p. 272. DOI: `10.1038/msb.2009.30`.

[150] K. P. Adamala, D. A. Martin-Alarcon, K. R. Guthrie-Honea, and E. S. Boyden. “Engineering genetic circuit interactions within and between synthetic minimal cells”. *Nature Chemistry* 9.5 (2016), pp. 431–439. DOI: `10.1038/nchem.2644`.

[151] M. Kaneda, S.-i. M. Nomura, S. Ichinose, S. Kondo, K.-i. Nakahama, K. Akiyoshi, and I. Morita. “Direct formation of proteo-liposomes by in vitro synthesis and cellular cytosolic delivery with connexin-expressing liposomes”. *Biomaterials* 30.23-24 (2009), pp. 3971–3977. DOI: `10.1016/j.biomaterials.2009.04.006`.

[152] R. Lentini, S. P. Santero, F. Chizzolini, D. Cecchi, J. Fontana, M. Marchioretto, C. Del Bianco, J. L. Terrell, A. C. Spencer, L. Martini, M. Forlin, M. Assfalg, M. D. Serra, W. E. Bentley, and S. S. Mansy. “Integrating artificial with natural cells to translate chemical messages that direct E. coli behaviour”. *Nature Communications* 5.1 (2014). DOI: `10.1038/ncomms5012`.

[153] P. M. Gardner, K. Winzer, and B. G. Davis. “Sugar synthesis in a protocellular model leads to a cell signalling response in bacteria”. *Nature Chemistry* 1.5 (2009), pp. 377–383. DOI: `10.1038/nchem.296`.

[154] G. Rampioni, F. D’Angelo, L. Leoni, and P. Stano. “Gene-Expressing Liposomes as Synthetic Cells for Molecular Communication Studies”. *Frontiers in Bioengineering and Biotechnology* 7 (2019), p. 1. DOI: `10.3389/fbioe.2019.00001`.

[155] Y. Ding, L. E. Contreras-Llano, E. Morris, M. Mao, and C. Tan. “Minimizing Context Dependency of Gene Networks Using Artificial Cells”. *ACS Applied Materials & Interfaces* 10.36 (2018), pp. 30137–30146. DOI: `10.1021/acsami.8b10029`.

[156] N. Krinsky, M. Kaduri, A. Zinger, J. Shainsky-Roitman, M. Goldfeder, I. Benhar, D. Hershkovitz, and A. Schroeder. “Synthetic Cells Synthesize Therapeutic Proteins inside Tumors”. *Advanced Healthcare Materials* 7.9 (2018), p. 1701163. DOI: `10.1002/adhm.201701163`.

[157] T.-Y. D. Tang, D. Cecchi, G. Fracasso, D. Accardi, A. Coutable-Pennarun, S. S. Mansy, A. W. Perriman, J. L. R. Anderson, and S. Mann. “Gene-Mediated Chemical Communication in Synthetic Protocell Communities”. *ACS Synthetic Biology* 7.2 (2017), pp. 339–346. DOI: `10.1021/acssynbio.7b00306`.

[158] Y. Elani, R. V. Law, and O. Ces. “Vesicle-based artificial cells as chemical microreactors with spatially segregated reaction pathways”. *Nature Communications* 5.1 (2014), pp. 1–5. DOI: `10.1038/ncomms6305`.

[159] N. Tompkins, N. Li, C. Girabawe, M. Heymann, G. B. Ermentrout, I. R. Epstein, and S. Fraden. "Testing Turing's theory of morphogenesis in chemical cells". *Proceedings of the National Academy of Sciences* 111.12 (2014), pp. 4397–4402. DOI: 10.1073/pnas.1322005111.

[160] K. Montagne, R. Plasson, Y. Sakai, T. Fujii, and Y. Rondelez. "Programming an in vitro DNA oscillator using a molecular networking strategy". *Molecular Systems Biology* 7.1 (2011), pp. 466–466. DOI: 10.1038/msb.2010.120.

[161] G. Gines, A. S. Zadorin, J.-C. Galas, T. Fujii, A. Estevez-Torres, and Y. Rondelez. "Microscopic agents programmed by DNA circuits". *Nature Nanotechnology* 12.4 (2017), pp. 351–359. DOI: 10.1038/nnano.2016.299.

[162] A. Joesaar, S. Yang, B. Bögels, A. van der Linden, P. Pieters, B. V. V. S. P. Kumar, N. Dalchau, A. Phillips, S. Mann, and T. F. A. de Greef. "DNA-based communication in populations of synthetic protocells". *Nature Nanotechnology* 14.4 (2019), pp. 369–378. DOI: 10.1038/s41565-019-0399-9.

[163] A. A. Hyman, C. A. Weber, and F. Jülicher. "Liquid-Liquid Phase Separation in Biology". *Annual Review of Cell and Developmental Biology* 30.1 (2014), pp. 39–58. DOI: 10.1146/annurev-cellbio-100913-013325.

[164] H. Jia, M. Heymann, F. Bernhard, P. Schwille, and L. Kai. "Cell-free protein synthesis in micro compartments: building a minimal cell from biobricks". *New Biotechnology* 39 (2017), pp. 199–205. DOI: 10.1016/j.nbt.2017.06.014.

[165] X. Zhou, H. Wu, M. Cui, S. N. Lai, and B. Zheng. "Long-lived protein expression in hydrogel particles: towards artificial cells". *Chemical Science* 9.18 (2018), pp. 4275–4279. DOI: 10.1039/C8SC00383A.

[166] L. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, and J. E. Gouaux. "Structure of Staphylococcal α-Hemolysin, a Heptameric Transmembrane Pore". *Science* 274.5294 (1996), pp. 1859–1865. DOI: 10.1126/science.274.5294.1859.

[167] K. Jahnke, N. Ritzmann, J. Fichtler, A. Nitschke, Y. Dreher, T. Abele, G. Hofhaus, I. Platzman, R. R. Schröder, D. J. Müller, J. P. Spatz, and K. Göpfrich. "Proton gradients from light-harvesting E. coli control DNA assemblies for synthetic cells". *Nature Communications* 12.1 (2021), p. 3967. DOI: 10.1038/s41467-021-24103-x.

[168] E. Altamura, P. Albanese, R. Marotta, F. Milano, M. Fiore, M. Trotta, P. Stano, and F. Mavelli. "Chromatophores efficiently promote light-driven ATP synthesis and DNA transcription inside hybrid multicompartment artificial cells". *Proceedings of the National Academy of Sciences* 118.7 (2021), e2012170118. DOI: 10.1073/pnas.2012170118.

[169] G. Villar, A. D. Graham, and H. Bayley. "A tissue-like printed material". *Science* 340.6128 (2013), pp. 48–52. DOI: 10.1126/science.1229495.

[170] A. Alcinesio, O. J. Meacock, R. G. Allan, C. Monico, V. Restrepo Schild, I. Cazimoglu, M. T. Cornall, R. Krishna Kumar, and H. Bayley. "Controlled packing and single-droplet resolution of 3D-printed functional synthetic tissues". *Nature Communications* 11.1 (2020), p. 2105. DOI: 10.1038/s41467-020-15953-y.

[171] S. S. Dixit, A. Pincus, B. Guo, and G. W. Faris. "Droplet shape analysis and permeability studies in droplet lipid bilayers". *Langmuir* 28.19 (2012), pp. 7442–7451. DOI: 10.1021/la3005739.

[172] H. Bayley, I. Cazimoglu, and C. E. Hoskin. "Synthetic tissues". *Emerging Topics in Life Sciences* 3.5 (2019), pp. 615–622. DOI: 10.1042/ETLS20190120.

[173] M. J. Booth, V. R. Schild, A. D. Graham, S. N. Olof, and H. Bayley. "Light-activated communication in synthetic tissues". *Science Advances* 2.4 (2016), e1600056. DOI: 10.1126/sciadv.1600056.

[174] A. R. Thiam, N. Bremond, and J. Bibette. "From stability to permeability of adhesive emulsion bilayers". *Langmuir* 28.15 (2012), pp. 6291–6298. DOI: 10.1021/la3003349.

[175] K. Funakoshi, H. Suzuki, and S. Takeuchi. "Lipid Bilayer Formation by Contacting Monolayers in a Microfluidic Device for Membrane Protein Analysis". *Analytical Chemistry* 78.24 (2006), pp. 8169–8174. DOI: 10.1021/ac0613479.

[176] J.-C. Baret. “Surfactants in droplet-based microfluidics”. *Lab on a Chip* 12.3 (2012), pp. 422–433. DOI: `10.1039/C1LC20582J`.

[177] F. Crick. “Diffusion in embryogenesis”. *Nature* 225.5231 (1970), pp. 420–422. DOI: `10.1038/225420a0`.

[178] 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 (2008), p. 1632. DOI: `10.1039/b806706f`.

[179] A. R. Abate, T. Hung, P. Mary, J. J. Agresti, and D. A. Weitz. “High-throughput injection with microfluidics using picoinjectors”. *Proceedings of the National Academy of Sciences* 107.45 (2010), pp. 19163–19166. DOI: `10.1073/pnas.1006888107`.

[180] A. R. Abate and D. A. Weitz. “Faster multiple emulsification with drop splitting”. *Lab on a Chip* 11.11 (2011), pp. 1911–1915. DOI: `10.1039/C0LC00706D`.

[181] A. Sciambi and A. R. Abate. “Accurate microfluidic sorting of droplets at 30 kHz”. *Lab on a Chip* 15.1 (2015), pp. 47–51. DOI: `10.1039/C4LC01194E`.

[182] P. Gruner, B. Riechers, B. Semin, J. Lim, A. Johnston, K. Short, and J.-C. Baret. “Controlling molecular transport in minimal emulsions”. *Nature Communications* 7.1 (2016), p. 10392. DOI: `10.1038/ncomms10392`.

[183] R. Seemann, M. Brinkmann, T. Pfohl, and S. Herminghaus. “Droplet based microfluidics”. *Reports on Progress in Physics* 75.1 (2012), p. 016601. DOI: `10.1088/0034-4885/75/1/016601`.

[184] C. Martino and A. J. deMello. “Droplet-based microfluidics for artificial cell generation: a brief review”. *Interface Focus* 6.4 (2016), p. 20160011. DOI: `10.1098/rsfs.2016.0011`.

[185] S. Majumder, N. Wubshet, and A. P. Liu. “Encapsulation of complex solutions using droplet microfluidics towards the synthesis of artificial cells”. *Journal of Micromechanics and Microengineering* 29.8 (2019), p. 083001. DOI: `10.1088/1361-6439/ab2377`.

[186] K. Göpfrich, B. Haller, O. Staufer, Y. Dreher, U. Mersdorf, I. Platzman, and J. P. Spatz. “One-Pot Assembly of Complex Giant Unilamellar Vesicle-Based Synthetic Cells”. *ACS Synthetic Biology* 8.5 (2019), pp. 937–947. DOI: `10.1021/acssynbio.9b00034`.

[187] Y. Skhiri, P. Gruner, B. Semin, Q. Brosseau, D. Pekin, L. Mazutis, V. Goust, F. Kleinschmidt, A. El Harrak, J. B. Hutchison, E. Mayot, J.-F. Bartolo, A. D. Griffiths, V. Taly, and J.-C. Baret. “Dynamics of molecular transport by surfactants in emulsions”. *Soft Matter* 8.41 (2012), p. 10618. DOI: `10.1039/c2sm25934f`.

[188] A. N. Zaikin and A. M. Zhabotinsky. “Concentration wave propagation in two-dimensional liquid-phase self-oscillating system”. *Nature* 225.5232 (1970), pp. 535–537. DOI: `10.1038/225535b0`.

[189] N. Li, J. Delgado, H. O. González-Ochoa, I. R. Epstein, and S. Fraden. “Combined excitatory and inhibitory coupling in a 1-D array of Belousov–Zhabotinsky droplets”. *Physical Chemistry Chemical Physics* 16.22 (2014), pp. 10965–10978. DOI: `10.1039/C4CP00957F`.

[190] N. Li, N. Tompkins, H. Gonzalez-Ochoa, and S. Fraden. “Tunable diffusive lateral inhibition in chemical cells”. *The European Physical Journal E* 38.3 (2015), pp. 1–12. DOI: `10.1140/epje/i2015-15018-3`.

[191] N. Tompkins, M. C. Cambria, A. L. Wang, M. Heymann, and S. Fraden. “Creation and perturbation of planar networks of chemical oscillators”. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25.6 (2015), p. 064611. DOI: `10.1063/1.4922056`.

[192] J. Thiele, Y. Ma, S. M. C. Bruekers, S. Ma, and W. T. S. Huck. “25th Anniversary Article: Designer Hydrogels for Cell Cultures: A Materials Selection Guide”. *Advanced Materials* 26.1 (2014), pp. 125–148. DOI: `10.1002/adma.201302958`.

[193] H. Zhang, E. Tumarkin, R. M. A. Sullan, G. C. Walker, and E. Kumacheva. “Exploring Microfluidic Routes to Microgels of Biological Polymers”. *Macromolecular Rapid Communications* 28.5 (2007), pp. 527–538. DOI: `10.1002/marc.200600776`.

[194] E. Tumarkin and E. Kumacheva. “Microfluidic generation of microgels from synthetic and natural polymers”. *Chemical Society Reviews* 38.8 (2009), p. 2161. DOI: `10.1039/b809915b`.

[195] D. Velasco, E. Tumarkin, and E. Kumacheva. “Microfluidic encapsulation of cells in polymer microgels”. *Small* 8.11 (2012), pp. 1633–1642. DOI: 10.1002/smll.201102464.

[196] J. Thiele, Y. Ma, D. Foschepoth, M. M. K. Hansen, C. Steffen, H. A. Heus, and W. T. S. Huck. “DNA-functionalized hydrogels for confined membrane-free *in vitro* transcription/translation”. *Lab on a Chip* 14.15 (2014), p. 2651. DOI: 10.1039/c3lc51427g.

[197] N. Fatin-Rouge, A. Milon, J. Buffle, R. R. Goulet, and A. Tessier. “Diffusion and Partitioning of Solutes in Agarose Hydrogels: The Relative Influence of Electrostatic and Specific Interactions”. *The Journal of Physical Chemistry B* 107.44 (2003), pp. 12126–12137. DOI: 10.1021/jp0303164.

[198] P. G. de Gennes. “Reptation of a polymer chain in the presence of fixed obstacles”. *The Journal of Chemical Physics* 55.2 (1971), p. 572. DOI: 10.1063/1.1675789.

[199] A. Pluen, P. A. Netti, R. K. Jain, and D. A. Berk. “Diffusion of macromolecules in agarose gels: Comparison of linear and globular configurations”. *Biophysical Journal* 77.1 (1999), pp. 542–552. DOI: 10.1016/S0006-3495(99)76911-0.

[200] P. B. Allen, X. Chen, Z. B. Simpson, and A. D. Ellington. “Modeling Scalable Pattern Generation in DNA Reaction Networks”. *Natural Computing* 13.4 (2014), pp. 583–595. DOI: 10.1007/s11047-013-9392-7.

[201] T. Rodjanapanyakul, F. Takabatake, K. Abe, I. Kawamata, S. M. Nomura, and S. Murata. “Diffusion modulation of DNA by toehold exchange”. *Physical Review E* 97.5 (2018), p. 052617. DOI: 10.1103/PhysRevE.97.052617.

[202] Z. Swank and S. J. Maerkl. “CFPU: A Cell-Free Processing Unit for High-Throughput, Automated In Vitro Circuit Characterization in Steady-State Conditions”. *BioDesign Research* 2021 (2021). DOI: 10.34133/2021/2968181.

[203] C. B. Goy, R. E. Chaile, and R. E. Madrid. “Microfluidics and hydrogel: A powerful combination”. *Reactive and Functional Polymers* 145 (2019), p. 104314. DOI: 10.1016/j.reactfunctpolym.2019.104314.

[204] J. Nie, Q. Gao, Y. Wang, J. Zeng, H. Zhao, Y. Sun, J. Shen, H. Ramezani, Z. Fu, Z. Liu, M. Xiang, J. Fu, P. Zhao, W. Chen, and Y. He. “Vessel-on-a-chip with Hydrogel-based Microfluidics”. *Small* 14.45 (2018), p. 1802368. DOI: 10.1002/smll.201802368.

[205] Y. Ling, J. Rubin, Y. Deng, C. Huang, U. Demirci, J. M. Karp, and A. Khademhosseini. “A cell-laden microfluidic hydrogel”. *Lab on a Chip* 7.6 (2007), p. 756. DOI: 10.1039/b615486g.

[206] A. Buxboim, M. Bar-Dagan, V. Frydman, D. Zbaida, M. Morpurgo, and R. Bar-Ziv. “A Single-Step Photolithographic Interface for Cell-Free Gene Expression and Active Biochips”. *Small* 3.3 (2007), pp. 500–510. DOI: 10.1002/smll.200600489.

[207] G. Pardatscher, M. Schwarz-Schilling, S. S. Daube, R. H. Bar-Ziv, and F. C. Simmel. “Gene Expression on DNA Biochips Patterned with Strand-Displacement Lithography”. *Angewandte Chemie International Edition* 57.17 (2018), pp. 4783–4786. DOI: 10.1002/anie.201800281.

[208] F. Greiss, S. S. Daube, V. Noireaux, and R. Bar-Ziv. “From deterministic to fuzzy decision-making in artificial cells”. *Nature Communications* 11.1 (2020), pp. 1–9. DOI: 10.1038/s41467-020-19395-4.

[209] M. A. Unger, H.-P. Chou, T. Thorsen, A. Scherer, and S. R. Quake. “Monolithic microfabricated valves and pumps by multilayer soft lithography”. *Science* 288.5463 (2000), pp. 113–116. DOI: 10.1126/science.288.5463.113.

[210] A. J. van der Linden, M. Yelleswarapu, P. A. Pieters, Z. Swank, W. T. S. Huck, S. J. Maerkl, and T. F. A. de Greef. “A Multilayer Microfluidic Platform for the Conduction of Prolonged Cell-Free Gene Expression”. *Journal of Visualized Experiments* 152 (2019), p. 59655. DOI: 10.3791/59655.

[211] N. Laohakunakorn, B. Lavickova, Z. Swank, J. Laurent, and S. J. Maerkl. “Steady-State Cell-Free Gene Expression with Microfluidic Chemostats”. *Synthetic Gene Circuits*. Vol. 2229. Springer US, 2021, pp. 189–203. DOI: 10.1007/978-1-0716-1032-9.

[212] S. J. Maerkl and S. R. Quake. “A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors”. *Science* 315.5809 (2007), pp. 233–237. DOI: 10.1126/science.1131007.

[213] Z. Noszticzius, W. Horsthemke, W. D. McCormick, H. L. Swinney, and W. Y. Tam. “Sustained chemical waves in an annular gel reactor: a chemical pinwheel”. *Nature* 329.6140 (1987), pp. 619–620. DOI: 10.1038/329619a0.

[214] W. Y. Tam, W. Horsthemke, Z. Noszticzius, and H. L. Swinney. “Sustained spiral waves in a continuously fed unstirred chemical reactor”. *The Journal of Chemical Physics* 88.5 (1988), pp. 3395–3396. DOI: 10.1063/1.453935.

[215] G. Kshirsagar, Z. Noszticzius, W. McCormick, and H. L. Swinney. “Spatial patterns in a uniformly fed membrane reactor”. *Physica D: Nonlinear Phenomena* 49.1-2 (1991), pp. 5–12. DOI: 10.1016/0167-2789(91)90187-E.

[216] V. Castets, E. Dulos, J. Boissonade, and P. De Kepper. “Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern”. *Physical Review Letters* 64.24 (1990), p. 2953. DOI: 10.1103/PhysRevLett.64.2953.

[217] Q. Ouyang and H. L. Swinney. “Transition from a uniform state to hexagonal and striped Turing patterns”. *Nature* 352.6336 (1991), pp. 610–612. DOI: 10.1038/352610a0.

[218] R. D. Vigil, Q. Ouyang, and H. L. Swinney. “Turing patterns in a simple gel reactor”. *Physica A: Statistical Mechanics and its Applications* 188.1-3 (1992), pp. 17–25. DOI: 10.1016/0378-4371(92)90248-O.

[219] A. P. Wong, R. Perez-Castillejos, J. Christopher Love, and G. M. Whitesides. “Partitioning microfluidic channels with hydrogel to construct tunable 3-D cellular microenvironments”. *Biomaterials* 29.12 (2008), pp. 1853–1861. DOI: 10.1016/j.biomaterials.2007.12.044.

[220] M.-H. Wu, S.-B. Huang, Z. Cui, Z. Cui, and G.-B. Lee. “Development of perfusion-based micro 3-D cell culture platform and its application for high throughput drug testing”. *Sensors and Actuators B: Chemical* 129.1 (2008), pp. 231–240. DOI: 10.1016/j.snb.2007.07.145.

[221] M. P. Cuchiara, A. C. Allen, T. M. Chen, J. S. Miller, and J. L. West. “Multilayer microfluidic PEGDA hydrogels”. *Biomaterials* 31.21 (2010), pp. 5491–5497. DOI: 10.1016/j.biomaterials.2010.03.031.

[222] K. Y. Lee and D. J. Mooney. “Hydrogels for Tissue Engineering”. *Chemical Reviews* 101.7 (2001), pp. 1869–1880. DOI: 10.1021/cr000108x.

[223] W. Saadi, S. W. Rhee, F. Lin, B. Vahidi, B. G. Chung, and N. L. Jeon. “Generation of stable concentration gradients in 2D and 3D environments using a microfluidic ladder chamber”. *Biomedical Microdevices* 9.5 (2007), pp. 627–635. DOI: 10.1007/s10544-007-9051-9.

[224] X. Shi, J. Zhou, Y. Zhao, L. Li, and H. Wu. “Gradient-Regulated Hydrogel for Interface Tissue Engineering: Steering Simultaneous Osteo/Chondrogenesis of Stem Cells on a Chip”. *Advanced Healthcare Materials* 2.6 (2013), pp. 846–853. DOI: 10.1002/adhm.201200333.

[225] P. J. Dorsey, M. Rubanov, W. Wang, and R. Schulman. “Digital Maskless Photolithographic Patterning of DNA-Functionalized Poly(ethylene glycol) Diacrylate Hydrogels with Visible Light Enabling Photodirected Release of Oligonucleotides”. *ACS Macro Letters* 8.9 (2019), pp. 1133–1140. DOI: 10.1021/acsmacrolett.9b00450.

[226] L. Tian, M. Li, A. J. Patil, B. W. Drinkwater, and S. Mann. “Artificial morphogen-mediated differentiation in synthetic protocells”. *Nature Communications* 10.1 (2019), p. 3321. DOI: 10.1038/s41467-019-11316-4.

[227] T.-Y. D. Tang, D. van Swaay, A. deMello, J. L. Ross Anderson, and S. Mann. “*In vitro* gene expression within membrane-free coacervate protocells”. *Chemical Communications* 51.57 (2015), pp. 11429–11432. DOI: 10.1039/C5CC04220H.

[228] B. V. V. S. Pavan Kumar, J. Fothergill, J. Bretherton, L. Tian, A. J. Patil, S. A. Davis, and S. Mann. “Chloroplast-containing coacervate micro-droplets as a step towards photosynthetically active membrane-free protocells”. *Chemical Communications* 54.29 (2018), pp. 3594–3597. DOI: 10.1039/C8CC01129J.

[229] R. Booth, Y. Qiao, M. Li, and S. Mann. “Spatial positioning and chemical coupling in coacervate-in-proteinosome protocells”. *Angewandte Chemie International Edition* 58.27 (2019), pp. 9120–9124. DOI: 10.1002/ange.201903756.

[230] V. Mukwaya, P. Zhang, L. Liu, A. Y. Dang-i, M. Li, S. Mann, and H. Dou. "Programmable Membrane-Mediated Attachment of Synthetic Virus-like Nanoparticles on Artificial Protocells for Enhanced Immunogenicity". *Cell Reports Physical Science* 2.1 (2021), p. 100291. DOI: `10.1016/j.xcrp.2020.100291`.

[231] S. Yang, P. A. Pieters, A. Joesaar, B. W. Bögels, R. Brouwers, I. Myrgorodska, S. Mann, and T. F. de Greef. "Light-activated signaling in DNA-encoded sender-receiver architectures". *ACS Nano* 14.11 (2020), pp. 15992–16002. DOI: `10.1101/2020.06.10.144162`.

[232] L. Rodríguez-Arco, B. V. V. S. P. Kumar, M. Li, A. J. Patil, and S. Mann. "Modulation of Higher-order Behaviour in Model Protocell Communities by Artificial Phagocytosis". *Angewandte Chemie International Edition* 58.19 (2019), pp. 6333–6337. DOI: `10.1002/anie.201901469`.

[233] P. Gobbo, A. J. Patil, M. Li, R. Harniman, W. H. Briscoe, and S. Mann. "Programmed assembly of synthetic protocells into thermoresponsive prototissues". *Nature Materials* 17.12 (2018), pp. 1145–1153. DOI: `10.1038/s41563-018-0183-5`.

[234] A. I. Oparin. *The origin of life on the earth.* 3rd Edition. Oliver & Boyd, 1957.

[235] I. Salazar-Ciudad. "Mechanisms of pattern formation in development and evolution". *Development* 130.10 (2003), pp. 2027–2037. DOI: `10.1242/dev.00425`.

[236] L. Wolpert. "Positional Information and Pattern Formation". *Current Topics in Developmental Biology*. Vol. 6. Academic Press, 1971, pp. 183–224. DOI: `10.1016/S0070-2153(08)60641-9`.

[237] A. M. Turing. "The chemical basis of morphogenesis". *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 237.641 (1952), pp. 37–72. DOI: `10.1098/rstb.1952.0012`.

[238] J. B. A. Green and J. Sharpe. "Positional information and reaction-diffusion: two big ideas in developmental biology combine". *Development* 142.7 (2015), pp. 1203–1211. DOI: `10.1242/dev.114991`.

[239] L. Wolpert. "Positional Information and the Spatial Pattern of Cellular Differentiation". *Journal of Theoretical Biology* 25.1 (1969), pp. 1–47. DOI: `10.1016/S0022-5193(69)80016-0`.

[240] A. Gierer and H. Meinhardt. "A theory of biological pattern formation". *Kybernetik* 12.1 (1972), pp. 30–39. DOI: `10.1007/BF00289234`.

[241] L. Wolpert. "Positional information revisited". *Development* 107 (1989), pp. 3–12. DOI: `10.1242/dev.107.Supplement.3`.

[242] C. Mou, B. Jackson, P. Schneider, P. A. Overbeek, and D. J. Headon. "Generation of the primary hair follicle pattern". *Proceedings of the National Academy of Sciences* 103.24 (2006), pp. 9075–9080. DOI: `10.1073/pnas.0600825103`.

[243] S. Sick, S. Reinker, J. Timmer, and T. Schlake. "WNT and DKK Determine Hair Follicle Spacing Through a Reaction-Diffusion Mechanism". *Science* 314.5804 (2006), pp. 1447–1450. DOI: `10.1126/science.1130088`.

[244] F. Michon, L. Forest, E. Collomb, J. Demongeot, and D. Dhouailly. "BMP2 and BMP7 play antagonistic roles in feather induction". *Development* 135.16 (2008), pp. 2797–2805. DOI: `10.1242/dev.018341`.

[245] H.-S. Jung, P. H. Francis-West, R. B. Widelitz, T.-X. Jiang, S. Ting-Berreth, C. Tickle, L. Wolpert, and C.-M. Chuong. "Local Inhibitory Action of BMPs and Their Relationships with Activators in Feather Formation: Implications for Periodic Patterning". *Developmental Biology* 196.1 (1998), pp. 11–23. DOI: `10.1006/dbio.1998.8850`.

[246] B. Ding, E. L. Patterson, S. V. Holalu, J. Li, G. A. Johnson, L. E. Stanley, A. B. Greenlee, F. Peng, H. Bradshaw, M. L. Blinov, B. K. Blackman, and Y.-W. Yuan. "Two MYB Proteins in a Self-Organizing Activator-Inhibitor System Produce Spotted Pigmentation Patterns". *Current Biology* 30.5 (2020), 802–814.e8. DOI: `10.1016/j.cub.2019.12.067`.

[247] A. D. Economou, A. Ohazama, T. Porntaveetus, P. T. Sharpe, S. Kondo, M. A. Basson, A. Gritli-Linde, M. T. Cobourne, and J. B. A. Green. "Periodic stripe formation by a Turing mechanism operating at growth zones in the mammalian palate". *Nature Genetics* 44.3 (2012), pp. 348–351. DOI: `10.1038/ng.1090`.

[248] R. Sheth, L. Marcon, M. F. Bastida, M. Junco, L. Quintana, R. Dahn, M. Kmita, J. Sharpe, and M. A. Ros. "Hox Genes Regulate Digit Patterning by Controlling the Wavelength of a Turing-Type Mechanism". *Science* 338.6113 (2012), pp. 1476–1480. DOI: 10.1126/science.1226804.

[249] J. Raspopovic, L. Marcon, L. Russo, and J. Sharpe. "Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients". *Science* 345.6196 (2014), pp. 566–570. DOI: 10.1126/science.1252960.

[250] V. E. Foe and B. M. Alberts. "Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis". *Journal of Cell Science* 61.1 (1983), pp. 31–70. DOI: 10.1242/jcs.61.1.31.

[251] E. Poustelnikova, A. Pisarev, M. Blagov, M. Samsonova, and J. Reinitz. "A database for management of gene expression data in situ". *Bioinformatics* 20.14 (2004), pp. 2212–2221. DOI: 10.1093/bioinformatics/bth222.

[252] A. Pisarev, E. Poustelnikova, M. Samsonova, and J. Reinitz. "FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution". *Nucleic Acids Research* 37.suppl$_1$ (2009), pp. D560–D566. DOI: 10.1093/nar/gkn717.

[253] G. Tkačik, J. O. Dubuis, M. D. Petkova, and T. Gregor. "Positional Information, Positional Error, and Readout Precision in Morphogenesis: A Mathematical Framework". *Genetics* 199.1 (2015), pp. 39–59. DOI: 10.1534/genetics.114.171850.

[254] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek. "Probing the Limits to Positional Information". *Cell* 130.1 (2007), pp. 153–164. DOI: 10.1016/j.cell.2007.05.025.

[255] H. Meinhardt. "A model of pattern formation in insect embryogenesis". *Journal of Cell Science* 23.1 (1977), pp. 117–139. DOI: 10.1242/jcs.23.1.117.

[256] S. Sen, P. Ghosh, S. S. Riaz, and D. S. Ray. "Spatial periodicity induced by a chemical wave train". *Physical Review E* 81.1 (2010), p. 017101. DOI: 10.1103/PhysRevE.81.017101.

[257] L. Manukyan, S. A. Montandon, A. Fofonjka, S. Smirnov, and M. C. Milinkovitch. "A living mesoscopic cellular automaton made of skin scales". *Nature* 544.7649 (2017), pp. 173–179. DOI: 10.1038/nature22031.

[258] S. Wolfram. "Universality and complexity in cellular automata". *Physica D: Nonlinear Phenomena* 10.1-2 (1984), pp. 1–35. DOI: 10.1016/0167-2789(84)90245-8.

[259] M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, and T. Gregor. "Optimal Decoding of Cellular Identities in a Genetic Network". *Cell* 176.4 (2019), 844–855.e15. DOI: 10.1016/j.cell.2019.01.007.

[260] I. J. Good. "Studies in the history of probability and statistics. XXXVII AM Turing's statistical work in World War II". *Biometrika* (1979), pp. 393–396. DOI: 10.2307/2335677.

[261] K. H. ten Tusscher and P. Hogeweg. "Evolution of Networks for Body Plan Patterning; Interplay of Modularity, Robustness and Evolvability". *PLoS Computational Biology* 7.10 (2011), e1002208. DOI: 10.1371/journal.pcbi.1002208.

[262] J. Schonfeld and D. Ashlock. "A comparison of the robustness of evolutionary computation and random walks". *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*. Vol. 1. IEEE, 2004, pp. 250–257. DOI: 10.1109/CEC.2004.1330864.

[263] H. Q. Dinh, N. Aubert, N. Noman, T. Fujii, Y. Rondelez, and H. Iba. "An Effective Method for Evolving Reaction Networks in Synthetic Biochemical Systems". *IEEE Transactions on Evolutionary Computation* 19.3 (2015), pp. 374–386. DOI: 10.1109/TEVC.2014.2326863.

[264] C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami. "Evolution of digital organisms at high mutation rates leads to survival of the flattest". *Nature* 412.6844 (2001), pp. 331–333. DOI: 10.1038/35085569.

[265] S. Kijani, U. Yrlid, M. Heyden, M. Levin, J. Borén, and P. Fogelstrand. "Filter-Dense Multicolor Microscopy". *PLOS ONE* 10.3 (2015), e0119499. DOI: 10.1371/journal.pone.0119499.

[266] S. Kaneda, K. Ono, T. Fukuba, T. Nojima, T. Yamamoto, and T. Fujii. "Modification of the glass surface property in PDMS-glass hybrid microfluidic devices". *Analytical Sciences* 28.1 (2012), p. 39. DOI: `10.2116/analsci.28.39`.

[267] G. F. Christopher and S. L. Anna. "Microfluidic methods for generating continuous droplet streams". *Journal of Physics D: Applied Physics* 40.19 (2007), R319–R336. DOI: `10.1088/0022-3727/40/19/R01`.

[268] S. L. Anna, N. Bontoux, and H. A. Stone. "Formation of dispersions using "flow focusing" in microchannels". *Applied Physics Letters* 82.3 (2003), p. 364. DOI: `10.1063/1.1537519`.

[269] K. Kapsner and F. C. Simmel. "Partitioning variability of a compartmentalized *in vitro* transcriptional thresholding circuit". *ACS Synthetic Biology* 4.10 (2015), pp. 1136–1143. DOI: `10.1021/acssynbio.5b00051`.

[270] E. Falgenhauer, S. Schönberg, C. Meng, A. Mückl, K. Vogele, Q. Emslander, C. Ludwig, and F. C. Simmel. "Evaluation of an *E. coli* Cell Extract Prepared by Lysozyme-Assisted Sonication via Gene Expression, Phage Assembly and Proteomics". *ChemBioChem* 22.18 (2021), pp. 2805–2813. DOI: `10.1002/cbic.202100257`.

[271] A. N. De Belder and K. Granath. "Preparation and properties of fluorescein-labelled dextrans". *Carbohydrate Research* 30.2 (1973), pp. 375–378. DOI: `10.1016/S0008-6215(00)81824-8`.

[272] A. E. Speers and B. F. Cravatt. "Profiling enzyme activities in vivo using click chemistry methods". *Chemistry & Biology* 11.4 (2004), pp. 535–546. DOI: `10.1016/j.chembiol.2004.03.012`.

[273] R. Gabler, E. W. Westhead, and N. C. Ford. "Studies of ribosomal diffusion coefficients using laser light-scattering spectroscopy". *Biophysical Journal* 14.7 (1974), pp. 528–545. DOI: `10.1016/S0006-3495(74)85933-3`.

[274] G. T. Hermanson. *Bioconjugate techniques*. 2nd edition. Academic Press, 2008.

[275] V. V. Rostovtsev, L. G. Green, V. V. Fokin, and K. B. Sharpless. "A stepwise huisgen cycloaddition process: copper (I)-catalyzed regioselective "ligation" of azides and terminal alkynes". *Angewandte Chemie* 114.14 (2002), pp. 2708–2711. DOI: `10.1002/1521-3757(20020715)114:14<2708::AID-ANGE2708>3.0.CO;2-0`.

[276] R. Axen, H. Drevin, and J. Carlsson. "Preparation of modified agarose gels containing thiol groups". *Acta Chemica Scandinavica B* 29.4 (1975), pp. 471–474. DOI: `10.3891/acta.chem.scand.29b-0471`.

[277] S. Punna, E. Kaltgrad, and M. G. Finn. ""Clickable" Agarose for Affinity Chromatography". *Bioconjugate Chemistry* 16.6 (2005), pp. 1536–1541. DOI: `10.1021/bc0501496`.

[278] P. Serwer. "Agarose gels: Properties and use for electrophoresis". *Electrophoresis* 4.6 (1983), pp. 375–382. DOI: `10.1002/elps.1150040602`.

[279] A. R. Abate, C.-H. Chen, J. J. Agresti, and D. A. Weitz. "Beating Poisson encapsulation statistics using close-packed ordering". *Lab on a Chip* 9.18 (2009), p. 2628. DOI: `10.1039/b909386a`.

[280] G. S. Filonov, J. D. Moon, N. Svensen, and S. R. Jaffrey. "Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution". *Journal of the American Chemical Society* 136.46 (2014), pp. 16299–16308. DOI: `10.1021/ja508478x`.

[281] J. R. Babendure, S. R. Adams, and R. Y. Tsien. "Aptamers Switch on Fluorescence of Triphenylmethane Dyes". *Journal of the American Chemical Society* 125.48 (2003), pp. 14716–14717. DOI: `10.1021/ja037994o`.

[282] M. M. K. Hansen, M. Ventosa Rosquelles, M. Yelleswarapu, R. J. M. Maas, A. J. van Vugt-Jonker, H. A. Heus, and W. T. S. Huck. "Protein Synthesis in Coupled and Uncoupled Cell-Free Prokaryotic Gene Expression Systems". *ACS Synthetic Biology* 5.12 (2016), pp. 1433–1440. DOI: `10.1021/acssynbio.6b00010`.

[283] G. Tkačik, C. G. Callan, and W. Bialek. "Information flow and optimization in transcriptional regulation". *Proceedings of the National Academy of Sciences* 105.34 (2008), pp. 12265–12270. DOI: `10.1073/pnas.0806077105`.

[284] M. Zagorski, Y. Tabata, N. Brandenberg, M. P. Lutolf, G. Tkačik, T. Bollenbach, J. Briscoe, and A. Kicheva. “Decoding of position in the developing neural tube from antiparallel morphogen gradients”. *Science* 356.6345 (2017), pp. 1379–1383. DOI: `10.1126/science.aam5887`.

[285] J. C. Waters. “Accuracy and precision in quantitative fluorescence microscopy”. *The Journal of Cell Biology* 185.7 (2009), pp. 1135–1148. DOI: `10.1083/jcb.200903097`.

[286] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek. “Entropy and Information in Neural Spike Trains”. *Physical Review Letters* 80.1 (1998), pp. 197–200. DOI: `10.1103/PhysRevLett.80.197`.

[287] N. Slonim, G. S. Atwal, G. Tkacik, and W. Bialek. “Estimating mutual information and multi–information in large networks”. *arXiv* preprint (2005).

[288] T. M. Cover and J. A. Thomas. *Elements of Information Theory, Second Edition*. Wiley, 2005.

[289] Y. Guan, Z. Li, S. Wang, P. M. Barnes, X. Liu, H. Xu, M. Jin, A. P. Liu, and Q. Yang. “A robust and tunable mitotic oscillator in artificial cells”. *eLife* 7 (2018), e33549. DOI: `10.7554/eLife.33549`.

[290] M. M. K. Hansen, L. H. H. Meijer, E. Spruijt, R. J. M. Maas, M. V. Rosquelles, J. Groen, H. A. Heus, and W. T. S. Huck. “Macromolecular crowding creates heterogeneous environments of gene expression in picolitre droplets”. *Nature Nanotechnology* 11.2 (2015), pp. 191–197. DOI: `10.1038/nnano.2015.243`.

[291] J. O. Dubuis, G. Tkacik, E. F. Wieschaus, T. Gregor, and W. Bialek. “Positional information, in bits”. *Proceedings of the National Academy of Sciences* 110.41 (2013), pp. 16301–16308. DOI: `10.1073/pnas.1315642110`.

[292] T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank. “Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient”. *Cell* 130.1 (2007), pp. 141–152. DOI: `10.1016/j.cell.2007.05.026`.

[293] T. Gregor, W. Bialek, R. R. d. R. van Steveninck, D. W. Tank, and E. F. Wieschaus. “Diffusion and scaling during early embryonic pattern formation”. *Proceedings of the National Academy of Sciences* 102.51 (2005), pp. 18403–18407. DOI: `10.1073/pnas.0509483102`.

[294] P. L. Lakin-Thomas. “Circadian rhythms: new functions for old clock genes”. *Trends in Genetics* 16.3 (2000), pp. 135–142. DOI: `10.1016/S0168-9525(99)01945-9`.

[295] A. N. Dodd, N. Salathia, A. Hall, E. Kévei, R. Tóth, F. Nagy, J. M. Hibberd, A. J. Millar, and A. A. R. Webb. “Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage”. *Science* 309.5734 (2005), pp. 630–633. DOI: `10.1126/science.1115581`.

[296] K. Vanselow, J. T. Vanselow, P. O. Westermark, S. Reischl, B. Maier, T. Korte, A. Herrmann, H. Herzel, A. Schlosser, and A. Kramer. “Differential effects of PER2 phosphorylation: molecular basis for the human familial advanced sleep phase syndrome (FASPS)”. *Genes & Development* 20.19 (2006), pp. 2660–2672. DOI: `10.1101/gad.397006`.

[297] L. Fu and C. C. Lee. “The circadian clock: pacemaker and tumour suppressor”. *Nature Reviews Cancer* 3.5 (2003), pp. 350–361. DOI: `10.1038/nrc1072`.

[298] J. S. Takahashi. “Transcriptional architecture of the mammalian circadian clock”. *Nature Reviews Genetics* 18.3 (2017), pp. 164–179. DOI: `10.1038/nrg.2016.150`.

[299] J.-C. Leloup and A. Goldbeter. “Chaos and Birhythmicity in a Model for Circadian Oscillations of the PER and TIM Proteins in Drosophila”. *Journal of Theoretical Biology* 198.3 (1999), pp. 445–459. DOI: `10.1006/jtbi.1999.0924`.

[300] G. Kurosawa and A. Goldbeter. “Amplitude of circadian oscillations entrained by 24-h light–dark cycles”. *Journal of Theoretical Biology* 242.2 (2006), pp. 478–488. DOI: `10.1016/j.jtbi.2006.03.016`.

[301] D. Gonze, J.-C. Leloup, and A. Goldbeter. “Theoretical models for circadian rhythms in Neurospora and Drosophila”. *Comptes Rendus de l’Académie des Sciences - Series III - Sciences de la Vie* 323.1 (2000), pp. 57–67. DOI: `10.1016/S0764-4469(00)00111-6`.

[302] I. van Soest, M. del Olmo, C. Schmal, and H. Herzel. “Nonlinear phenomena in models of the circadian clock”. *Journal of The Royal Society Interface* 17.170 (2020), p. 20200556. DOI: `10.1098/rsif.2020.0556`.

[303] R. A. Wever. “Light Effects on Human Circadian Rhythms: A Review of Recent Andechs Experiments”. *Journal of Biological Rhythms* 4.2 (1989), pp. 49–73. DOI: `10.1177/074873048900400206`.

[304] S. Daan and C. Berde. “Two coupled oscillators: Simulations of the circadian pacemaker in mammalian activity rhythms”. *Journal of Theoretical Biology* 70.3 (1978), pp. 297–313. DOI: `10.1016/0022-5193(78)90378-8`.

[305] M. Tigges, T. T. Marquez-Lago, J. Stelling, and M. Fussenegger. “A tunable synthetic mammalian oscillator”. *Nature* 457.7227 (2009), pp. 309–312. DOI: `10.1038/nature07616`.

[306] A. J. Genot, A. Baccouche, R. Sieskind, N. Aubert-Kato, N. Bredeche, J. F. Bartolo, V. Taly, T. Fujii, and Y. Rondelez. “High-resolution mapping of bifurcations in nonlinear biochemical circuits”. *Nature Chemistry* 8.8 (2016), pp. 760–767. DOI: `10.1038/nchem.2544`.

[307] D. T. Riglar, D. L. Richmond, L. Potvin-Trottier, A. A. Verdegaal, A. D. Naydich, S. Bakshi, E. Leoncini, L. G. Lyon, J. Paulsson, and P. A. Silver. “Bacterial variability in the mammalian gut captured by a single-cell synthetic oscillator”. *Nature Communications* 10.1 (2019), p. 4665. DOI: `10.1038/s41467-019-12638-z`.

[308] J. Henningsen, M. Schwarz-Schilling, A. Leibl, J. Gutiérrez, S. Sagredo, and F. C. Simmel. “Single Cell Characterization of a Synthetic Bacterial Clock with a Hybrid Feedback Loop Containing dCas9-sgRNA”. *ACS Synthetic Biology* 9.12 (2020), pp. 3377–3387. DOI: `10.1021/acssynbio.0c00438`.

[309] Y. Hori, M. Takada, and S. Hara. “Biochemical oscillations in delayed negative cyclic feedback: Existence and profiles”. *Automatica* 49.9 (2013), pp. 2581–2590. DOI: `10.1016/j.automatica.2013.04.020`.

[310] H. L. Swinney. “Observations of order and chaos in nonlinear systems”. *Physica D: Nonlinear Phenomena* 7.1-3 (1983), pp. 3–15. DOI: `10.1016/0167-2789(83)90111-2`.

[311] M. J. Feigenbaum. “Quantitative universality for a class of nonlinear transformations”. *Journal of Statistical Physics* 19.1 (1978), pp. 25–52. DOI: `10.1007/BF01020332`.

[312] T.-Y. Li and J. A. Yorke. “Period Three Implies Chaos”. *The American Mathematical Monthly* 82.10 (1975), pp. 985–992. DOI: `10.2307/2318254`.

[313] N. Y. Garceau, Y. Liu, J. J. Loros, and J. C. Dunlap. “Alternative Initiation of Translation and Time-Specific Phosphorylation Yield Multiple Forms of the Essential Clock Protein FREQUENCY”. *Cell* 89.3 (1997), pp. 469–476. DOI: `10.1016/s0092-8674(00)80227-5`.

[314] E. Bünning and I. Moser. “Influence of Valinomycin on Circadian Leaf Movements of Phaseolus”. *Proceedings of the National Academy of Sciences* 69.9 (1972), pp. 2732–2733. DOI: `10.1073/pnas.69.9.2732`.

[315] L. Stone. “Period-doubling reversals and chaos in simple ecological models”. *Nature* 365.6447 (1993), pp. 617–620. DOI: `10.1038/365617a0`.

[316] J. Li, W. Haas, K. Jackson, E. Kuru, M. C. Jewett, Z. H. Fan, S. Gygi, and G. M. Church. “Cogenerating Synthetic Parts toward a Self-Replicating System”. *ACS Synthetic Biology* 6.7 (2017), pp. 1327–1336. DOI: `10.1021/acssynbio.6b00342`.

[317] J. A. Davies and E. Cachat. “Synthetic biology meets tissue engineering”. *Biochemical Society Transactions* 44.3 (2016), pp. 696–701. DOI: `10.1042/BST20150289`.