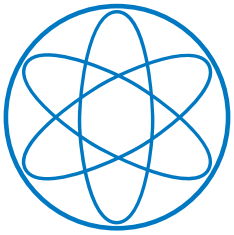

Physics Department

Cell-Free Expression Systems: a Bottom-Up Approach to Synthetic Biology

Dissertation by

Matthaeus Schwarz-Schilling

TECHNISCHE UNIVERSITÄT MÜNCHEN

TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Physik
Physik synthetischer Biosysteme

Cell-Free Expression Systems: a Bottom-Up Approach to Synthetic Biology

Matthaeus Schwarz-Schilling

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Martin Zacharias

Prüfer der Dissertation:

1. Prof. Dr. Friedrich C. Simmel

2. Prof. Dr. Ulrich Gerland

Die Dissertation wurde am 06.05.2019 bei der Technischen Universität München eingereicht und durch die Fakultät der Physik am 01.07.2019 angenommen.

Contents

Abstract

Synthetic biology focuses on the design and implementation of new tools and functions in biological systems. As an alternative to cells, cell-free expression systems provide an open and controllable platform to develop and study functional aspects of synthetic gene constructs. In this thesis, we employed cell-free expression systems to characterize synthetic gene constructs and then transferred them in living organisms.

First, we characterized a sender-receiver plasmid system which is based on a quorum sensing molecule in bacteria and in a cell-free gene expression system. To study the response of the receiver plasmid in spatial gradients of the gene inducer, we encapsulated the bacteria/cell-free gene expression system in emulsion droplets. We quantified the reaction-diffusion system based on the expression of a fluorescent protein within the receiver droplets and were able to demonstrate spatial control of gene expression. We further showed that the chemical signal can be produced enzymatically *in vitro* and *in vivo*. Finally, these findings allowed us to implement chemical communication between a cell-free expression system and bacteria.

In a second project, we investigated the expression of a RNA-protein nanostructure in a cell-free expression system. The RNA-component of the structure combined functional aspects such as scaffolding of a protein FRET-pair, binding of a fluorescent ligand and binding of the nanostructure to streptavidin coated beads. We characterized the structure in regard to complex formation efficiency and chemical stability. We then optimized the simultaneous expression of the components and monitored their assembly on streptavidin coated beads in a cell-free expression system in emulsion droplets. Finally, we transformed the plasmids into bacteria and compared the expression and assembly of the components in bacteria and in a cell-free gene expression system.

In a third project, we studied the reversibility of filamentation of *E. coli* bacteria with an inducible version of CRISPR interference (CRISPRi). In this context we tested CRISPRi and an anti-sgRNA strategy in a cell-free expression system. We then targeted the expression of the *ftsZ* gene with CRISPRi in bacteria. Targeting the gene stops cell division while still allowing for cell growth, thus resulting in filamentous cells. We showed that both static (decoy binding sites) and inducible thresholds (anti-sgRNA) can affect the characteristics of the switching process. Combining bulk data with single cell measurements, we eventually characterized the efficiency of the switching process.

Zusammenfassung

Synthetische Biologie beschäftigt sich mit dem Entwerfen und Implementieren von neuen Werkzeugen und Funktionen in biologischen Systemen. Hierbei stellen zellfreie Expressionssysteme eine Option zur Erprobung und Untersuchung synthetischer Werkzeuge in Zellen dar. Zellfreie Expressionssysteme bieten eine offene und kontrollierbare Umgebung für die Entwicklung und Beschreibung von funktionellen Aspekten synthetischer Genkonstrukte.

In einem ersten Projekt wurde ein System, welches auf Sender- und Empfänger-Plasmiden und Quorum sensing Molekülen, welche als chemische Signale fungieren, basiert, in Baktieren und in einem zellfreien Expressionsystem charakterisiert. Um die Wirkung von räumlichen Konzentrationsgradienten des chemischen Signals auf das Empfänger-Plasmid zu untersuchen, wurden die Baktierien bzw. das zellfreie Expressionssytem in Emulsionströpfchen kompartimentiert. Anhand der Expression eines fluoreszenten Proteins vom Empfänger-Plasmid wurde die Diffusion des chemischen Signalmoleküls quantifiziert. In der Folge war es möglich, Genexpression räumlich zu strukturieren. Zudem wurde gezeigt, dass es möglich ist, das chemische Signalmolekül *in vitro* und *in vivo* zu produzieren. Dies ermöglichte die Etablierung einer chemischen Kommunikation zwischen Bakterien und dem zellfreien Expressionssystem.

In einem zweiten Projekt wurde die Expression einer hybriden Nanostruktur aus RNA und Protein in einem zellfreien Expressionsystem untersucht. Der RNA-Anteil der Struktur erfüllt mehrere Funktionen wie das Bereitstellen eines Gerüsts für ein Protein-basiertes FRET-Paar, das Anbinden der Nanostruktur an Streptavidin-bedeckte Mikrokugeln oder das Binden eines fluoreszenten Liganden. Zudem wurde die Struktur in Bezug auf ihre chemische Stabilität und auf die Effizienz der Komplexbildung charakterisiert. Daraufhin wurde die simultane Expression der Komponenten der RNA-Protein-Nanostruktur optimiert und die Anordnung der Komponenten auf Streptavidin-bedeckten Mikro-kugeln in Emulsionströpfchen gezeigt. Schließlich wurden die Plasmide in Bakterien transfomiert und die Expression der Komponenten in Bakterien wurde mit der Expression im zellfreien Expressionssystem verglichen.

In einem dritten Projekt wurde die Reversibilität der Bildung von Filamenten von *E. coli*-Bakterien mittels einer induzierbaren Version eines Systems auf Basis der CRISPR-Interferenz (CRISPRi) untersucht. In diesem Kontext wurde CRISPRi mit einer Antisense-RNA-Strategie kombiniert und

in einem zellfreien Expressionssystem betrachtet. Anschließend wurde durch CRISPRi die Expression des *ftsZ*-Gens in *E. coli* unterdrückt, was zu einem Verlust der Fähigkeit zur Zellteilung führte, nicht aber zum Abschalten des Zellwachstums. Wir konnten zeigen, dass sowohl statische als auch induzierbare Schwellwerte den Umschaltprozess beeinflussen können. Schließlich konnten wir durch die Kombination mehrerer Messmethoden die Effizienz des Umschaltprozesses von sich teilenden Bakterien hin zu filamentierenden Bakterien charakterisieren.

Chapter 1

Introduction

This section serves as an introduction into the field of synthetic biology. We briefly introduce different approaches and then take a look at the proclaimed goals and challenges of the field. Then the focus changes and we look at the development of synthetic biology in bacteria, mainly *E. coli*, and cell-free synthetic biology. Finally, we will give a brief overview of the content of the thesis.

1.1 Synthetic biology

"Synthetic biology is the science of designing biological systems" [1] and brings an engineering mindset to biology. It is an interdisciplinary field enabled mainly by advancements in molecular biology and biotechnology. Traditionally, research was limited to understanding and optimizing natural systems, whereas today one can create and study artificial biological systems. The research can range from engineering of biomolecules such as proteins to changing the genome of living organisms. One can identify two main approaches, which are called *top-down* and *bottom-up*. The top-down approach starts with existing biological organisms and reduces them until one can understand aspects of it. A classical example for a 'top-down' approach to synthetic biology is the minimized genome of the *Mycoplasma mycoides* JCVI-sny1.0, where the genome was reduced from about 1079 kilo base pairs (kbp) to 531 kbp which was the minimum for the organism to still grow and replicate [2]. Bottom-up synthetic biology, starts from non-living molecules and combines them to create molecular systems or even an artificial cell. An example for the 'bottom-up' approach is a microfluidic assembly line that carefully constructs lipid vesicles with trans-membrane proteins inside that can use protein gradients to catalyze the creation of ATP [3]. In order to gain

more clarity about the definition and direction of the research field, the next two subsections will focus on more general goals and challenges of synthetic biology.

1.2 Goals of synthetic biology

The MIT Synthetic Biology Center defines that the goal of "synthetic biology is to make the construction of novel biological systems into a practical and useful engineering discipline" [4]. However, as there are different sub-areas under the term synthetic biology, the immediate goals in these sub-areas will differ. The spectrum ranges from finding new ways to manufacture molecules within engineered organisms to the construction of molecular computers to solve computational problems within or even outside of cells [5] [6] [7].

One unifying aspect of all the existing sub-areas in synthetic biology is the creation of novel functional parts from bio-molecules. This means that the field is not only carried by fundamental insights into the molecular world of biology but also driven by the hypotheses of the designer of a novel functional parts. The formulation of an hypothesis allows for falsification which can result in the generation of relevant research questions and advance fundamental research in the biological sciences. Hence, from an academic standpoint the boundaries of the field are particularly interesting. When a construct does not work as planned a new research question is formulated. In this spirit, Richard Feynman said: "what I cannot create I do not understand". However, sometimes the creation of functional parts such as the genetic construct for the bio-production of a malaria drug precursor via engineered yeast [5] is more of a story of trial-and error than the establishment of engineering rules [8]. Hence one of the challenges of synthetic biology can be articulated in a similar fashion as the quote from Richard Feynmann: what I can create, I do not necessarily understand [9]. The next section will go into more detail about the challenges in synthetic biology and how this can serve the field to reach its goals.

1.3 Challenges in synthetic biology

As outlined by Rollié *et al.*, electrical or chemical engineering have established themselves as traditional engineering disciplines by successfully implementing four general methods to reduce the complexity of the physical world: abstraction, modularization, standardization and optimization [10]. The introduction of engineering concepts to biological systems [11] is discussed in

this section. How well synthetic biology was able to apply and implement these four conceptual steps of engineering in its research is more controversial [8].

Standardization

In an effort to standardize genetic parts and inspired by standards from mechanical engineering the idea of BioBricks was introduced by Thomas Knight in 2003 [12]. BioBricks parts are nucleic acid sequences that form a functional unit such as a transcription promoter and are combined to create BioBrick devices such as genes with all the components necessary for expression. Furthermore there is a BioBrick assembly standard to assemble parts and devices into systems that can perform a task [13] (Fig. 1.1A). The list of parts are accessible in an online repository called "The Registry of Standard Biological Parts" which is mainly maintained by undergraduate students that take part in the international Genetically Engineered Machines (iGEM) competition [14]. However, one of the problems with the registry is that there is no warranty for the performance and quality of the parts, making it difficult to reliably combine parts into functional modules [15]. However, the efforts for the development of tools and parts in order to precisely control gene expression in bacteria, yeast and mammalian cells are ongoing [16] [17] [18]. The creation of a reliable toolbox for genetic parts is an important requirement for their abstraction, so that one can create genetic circuits from well characterized parts without having to consider the molecular details of the parts themselves. The next section is about the concept of abstraction in synthetic biology.

Abstraction

The process of abstraction is a projection or reduction towards a simpler system. A systems can be abstracted if it is known which details one can neglect and which details one has to consider. Biological systems are challenging to abstract due to a large number of molecules and molecular species present. Collisions between nearly all molecules within a cell create a large amount of interactions. Many parameters such as reaction rate constants or concentrations are difficult to access. Furthermore there are hidden interactions between molecules due to shared interaction partners or biochemical noise due to the stochastic nature of some reactions. These facts can hinder the formulation of predictive models about the behavior of synthetic systems within cells. For example in 2000, Leibler and Elowitz constructed a genetic oscillator termed "repressilator" due to its topology, which consists of three

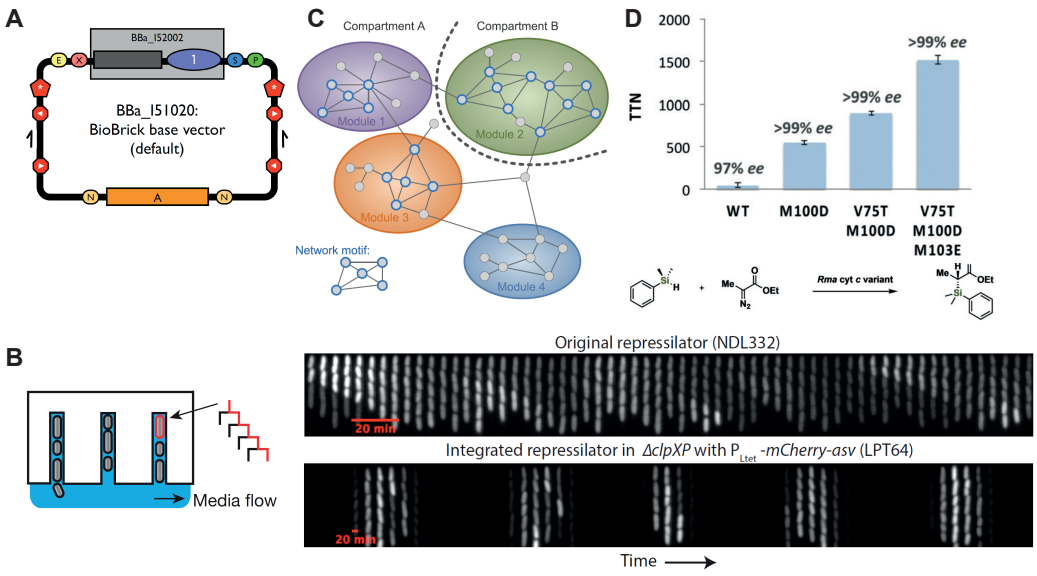

Figure 1.1: **A** Design of the BioBrick base vector. The pictographs on the vector each represent a different part. The vector is designed in such a manner that the construct which is framed in the gray box can be easily exchanged with the BioBrick assembly standard. The image is adapted from Ref. [13]. **B** Schematic representation of the microfluidic trap for bacteria to monitor single cell lines over several generations. Right upper image: kymograph from microscopy images of the fluorescence signal of bacteria growing in the microfluidic trap. The bacteria contain the original repressilator plasmid system from Ref [19]. Right lower image: kymograph from microscopy images of the fluorescence singal of bacteria with the revised 'repressilator' plasmid system from Ref [20]. The subfigure is adapted from Ref. [20]. Adapted by permission from Springer Nature, 2016. **C** Schematic representation of modularization by compartmentalization. Network graphs where each node represents a part and their interaction is represented by an edge, are separated by an compartment (dotted line). The subfigure is adapted from Ref. [10]. Copyright (2012), with permission from Elsevier. **D** An example of how directed evolution improved an enzyme (cytochrome *c* from *Rhodothermus marinus*) that catalyzes chiral Si-C bond formation. The bar graph shows the the total turnovers (ttn) of the enzymes in the wild type form, and with one, two and three mutations. The subfigure is adapted from Ref. [21].

transcription factors, repressing each other in a "rock-paper-scissors" fashion [22]. The oscillations of the transcription factors are observed via the co-expression of a fluorescent protein. Experiments then showed that out of approximately 100 analyzed bacteria about 60 % did not oscillate as was expected by the design of the genetic circuit. In 2016, Potvin-Trottier *et*

al. then published a revised version of the 16 year old genetic circuit, which decreased the coefficient of variation in period length of the oscillator from 0.35% to 0.14% [20]. Having the possibility of tracking single bacteria over generations, Potvin-Trottier *et al.* were able to identify and eliminate sources of noise in the oscillations. It could be experimentally shown that by considering new parameters it was possible to improve the performance of an genetic circuit (Fig. 1.1B). Some parameters they considered for the new design were the fluctuation of the plasmid copy number and the coupling through shared degradation pathways and the stochastic nature of gene expression. This genetic circuit can now be regarded as a synthetic clock module within *E. coli*. This example illustrates how synthetic biology struggles with the process of abstraction due to the complexity of biological systems and how it can overcome this problem by identifying the important and relevant mechanisms at play. It is possible to design a system in such a way that it can be reduced to its function, such as being a genetic clock.

Modularization

Modularization, meaning the interchangeability between different functional parts without affecting the functionality of the other parts or breaking the whole system, is a concept that is closely tied to all engineering disciplines. Due to the fact that most biological functions probably evolved together with the rest of the organism, modular architecture is not a necessity. Functional units can co-depend in intricate ways on other functional units or have visible or hidden redundancies within a cell or organism. This does not imply that functional systems in a cell are not modular, it can signify that the modules are hidden and intertwined. For example about 80% of the protein-coding genes are non-essential for the survival in most prokaryotes and eukaryotes, which means that their role can be compensated by other genes [23]. As more of the gene regulation networks in bacteria is uncovered, the more modules and functional units are identified [24]. Hence modular architecture is present, however it may be hidden or hard to detect. There are two main approaches to create modular systems. One is to create interaction between molecules which are orthogonal to the rest of the system. For example, Meyer *et al.* were able to create six variants of the RNA polymerase from phage T7 where each specifically recognizes a different promoter [25]. The other solution is to separate a set of components from the rest of the system by compartmentalization [11] (Fig. 1.1C). Compartments allow the separation of reactions via a membrane. A membrane can be designed to control the input and output signals of the compartment so that compartments are still able to communicate with each other however only via these designated

communication channels.

Compartments have evolved in most organisms such as protein-based micro-compart-ments in bacteria or organelles with a lipid membrane in eukaryotic cells [26]. Recent efforts have shown that prokaryotic micro-compartments can be expressed and assembled in different species of bacteria and even in yeast. The compartments can be used to encapsulate a proteins of interest or active enzyme and protect them against degradation [27] [28]. In a different study, Avalos *et al.* were able to compartmentalize the Ehrlich pathways into the mitochondria of yeast and thus greatly increase the isobutanol production [29]. In conclusion, although modules have evolved in living organisms, it can be challenging to detect them. Compartmentalization is one of the most promising strategies for achieving modularization as it establishes clear boundaries. Thereby synthetic biologist are able to build on solution that have already been created by nature.

Optimization

Optimization is the improvement of a functional system so it becomes more efficient. Often optimization means an increase in product yield when concering durgs or a or an increase in the robustness of a sensor system. However, optimization of metabolic pathways for the expression of certain molecules can be extremely challenging due to the fact that the trial-and-error was the most common approach to optimize synthetic genetic constructs [8].

Evolution optimized functional parts/elements in biological systems. The algorithm of mutation and selection over up to billions of years evolved all solutions that exist in biological systems today. Synthetic genetic constructs will also be subjected to spontaneous mutations once inserted into a cell. This can create a obstacle as the biological system will not be optimizing the function encoded in the synthetic construct. It will rather follow a fitness function that is tied to growth and survival of the organism. Often the two functions correlate negatively with each other. The production of any biomolecule will be costly in energy and resources. It forces engineers to opt for one of the two solutions.

First one can try to create artificial systems that do not have to survive/sustain on their own such as in cell-free synthetic biology. Here functional modules can be kept running due to engineered solutions like a mechanical pumping syste [30]. In addition, the creation and optimization of synthetic systems can be greatly facilitated by having an environment that is simpler than the cell. This can simplify prototyping and debugging of genetic constructs. Here cell-free expression systems provide a solution [31].

The other approach for engineers to work with is that of *directed evo-*

lution. Here both functions that of the synthetic construct and the fitness function of the organism are designed in such a way that they are tied together. This approach has proven useful in finding novel biochemistry and optimizing enzymatic pathways [21] (Fig. 1.1D). However, it does not solve the problem of genetic mutations inside of cells once a specific function has been evolved. Certain constraints, like spontaneous mutations of synthetic genetic constructs within cells have to be taken into account. The effect of mutation rates has to be approximated and mechanisms like DNA proofreading, which reduce them, have to be implemented. Alternatively there are work-around solutions outside of the cell.

This section provided an general overview of challenges within the field of synthetic biology. Furthermore it illuminated synthetic biology in regard to four different concepts of an engineering discipline: standardization, abstraction, modularization and optimization [10]. Despite the vast amount of challenges that accompany working with living organisms, synthetic biologist have demonstrated solutions within the constraints of knowledge and technologies. The idea of engineering biological systems is driven by the need for sustainable and economic solutions to protect the ecosystem and inspired by natural systems and human made machines. The next section provides an overview of important advancements of synthetic biology in bacteria within the last two decades.

1.4 Synthetic biology in bacteria

This section will first outline the development of tools that enabled synthetic biology in the twentieth century. This will be followed by a brief review of the first steps in creating artificial genetic systems in bacteria at the beginning of the twenty-first century. Finally, we mention some experiments that show current and potential developments in the area of synthetic biology in bacteria.

The foundation for the creation of most synthetic genetic constructs is in the understanding of the simplest biological organism such as prokaryotes. Especially the gram-negative bacteria species, *Escherichia coli*, serves as a model organism in molecular biology and now in synthetic biology. In addition, most tools for genetic modification and molecular cloning were discovered by studying bacteria in combination with bacteriophages [32]. In 1952, it was found that some bacteria can defend themselves against the infection of bacteriophages and in 1968, restriction enzymes were identified as the cause of this ability [33] [34]. In 1965 the observation was made that linear DNA of bacteriophages is circularized after the infection of the host,

and two years afterwards DNA ligase enzymes were successfully isolated independently by five different groups [35] [36] [37] [38] [39] [40]. Finally, in 1973 for the first time a recombinant DNA molecule was digested, ligated and transformed into *E. coli* [41]. Since then the techniques for molecular cloning have constantly improved, alongside with two very important technological innovations, the polymerase chain reaction (PCR) which allows amplification of recombinant DNA [42] and DNA sequencing, allowing verification of cloning[43].

In parallel, the unraveling of molecular mechanisms of the regulation mechanism of genes provided tools for the creation of gene switches. For example in the early 1960s, Jacob and Monod discovered the mechanism of regulator proteins within the lac operon [44]. Today, the so-called transcription factors are a widely used tool in synthetic biology [45].

This paragraph will present three early examples of synthetic biology in bacteria that have shown to be the basis for further developments in the field. In a review paper from 2009, Purnick and Weiss declare that the first wave of Synthetic Biology started around 2000 and entailed combining specific functional parts like transcription promoters or transcription factors to systems within cells [48]. Three following experiments have provided researchers with conceptual and technical ideas of how to pursue the construction of more complicated synthetic circuits. The first example is a bistable switch, called the 'toggle switch'. One can switch between the expression of two genes by the transient addition of a chemical inducer for either of the genes. As each of the genes represses the expression of the other gene it is ensured that only one of the two proteins is expressed at a time and that it remains in that state [49][1]. The genetic device consisted of about 10 different functional parts such as transcription promoters, ribosome binding sites (RBS), open reading frames for three proteins and transcription terminators (Fig. 1.2A). Another early example of a synthetic circuit that is created from functional parts is the aforementioned protein oscillator - termed *repressilator* in *E. coli* [19] (Fig. 1.2B). It consists of two plasmids, of three promoters two of those are derived from phages and of three corresponding transcription repressors. It was the first example of an autonomous synthetic circuits in bacteria as it creates protein oscillations in a regular manner in bacteria without further intervention by the researcher. The third and final example of a milestone experiment was a publication by Weiss and Knight in 2000, in which they demonstrate the creation of a sender and receiver module in *E. coli* (Fig. 1.2C). This allowed them to establish a chemical cell-to-cell communication which was naturally foreign to *E. coli* [51].

[1]The first 'toggle switch' for yeast was published in 2004 [50].

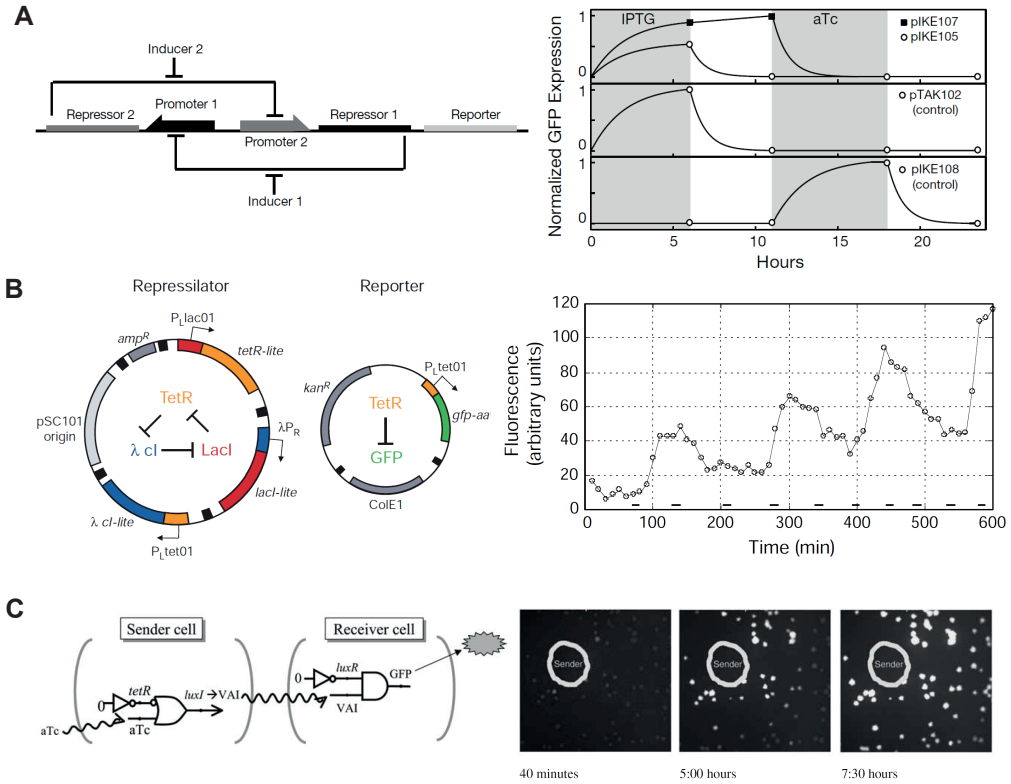

Figure 1.2: **A** Schematic representation of genetic construct that encodes the 'toggle' switch . Two inducible promoters express each a gene that represses the other gene. The switching behavior is visualized by the co-expression of GFP with the protein repressor 1 (induced by IPTG), where as the protein repressor 2 is induced by aTc. The subfigure is adapted from Ref. [46]. Adapted by permission from Springer Nature, 2000. **B** Scheme of the repressilator plasmid system. Three transcription factors that repress the expression of each other in a 'rock-papers-scissor' fashion that results in oscillations in the concentration of proteins in the bacteria. Left: GFP intensity of a single cell over time. The subfigure is adapted from Ref. [22]. Adapted by permission from Springer Nature, 2000. C Scheme of a sender and receiver system in bacteria that uses a chemical signal (here abbreviated VAI) to induce gene expression in the receiver cell. Left, sender cells in the circle produce the chemical signal and induce the receiver cells in the colonies around which is visualized by GFP. The subfigure is adapted from Ref. [47]. Adapted by permission from Springer Nature, 2003.

A crucial next step in genetic circuit design that build on the first wave of synthetic biology was published approximately a decade later, in 2016, when Nielsen *et al.* provided a software that automates the assembly of genetic circuits from previously characterized parts [52]. They demonstrated

its utility by designing 60 different genetic circuits in *E. coli* out of which approximately 45 worked as intended. However, new tools are constantly being developed. Two prominent recent examples of tool that are able to control the expression of genes are CRISPR interference (CRISPRi) and RNA toehold switches [53] [54]. CRISPRi is based on a mutated non-cleaving version of the Cas9 protein and can be used to halt transcription of a gene. The technology is described in more detail in Section 2.6. The RNA toehold switch is a post-transcriptional gene switch based on a hairpin structure around the RBS that inhibits ribosome assembly on the mRNA itself. Translation can be switched on by opening the hairpin structure *via* the binding of another RNA strands. Both tools have been applied to create logic circuits within bacteria [55] [56].

The efforts to create applications for the engineered bacteria are growing and especially medical application within mammalian guts are being explored [57]. In addition, strategies to fight cancer with engineered bacteria are being developed. Din *et al.* have recently demonstrated that certain bacteria that grow in and around cancer tumors, could be reprogrammed so that they lyse after a certain population density is reached. During lysis a genetically encoded cargo is being released which potentially can destroy cancer cells. After lysis, few bacteria survive so that the population regrows, creating pulses of the released cargo around cancer cells which results in repeated and coordinated attacks on the tumors.

Advancements in the engineering of bacteria in the microbiome serve as the last example of more recent developments in synthetic biology. Potential applications are promised in the field of diagnosis and personal medicine. For example, recently an interspecies communication between different engineered gut bacteria was established in mice. The communication was recorded in the bacteria using a genetic circuits that was used as a memory element [58].

In conclusion, functional systems within bacteria can be constructed from well defined parts. New tools are constantly being developed and enable the construction of new synthetic systems. In parallel, different applications of genetically engineered bacteria are starting to be realized despite the complexity of the biological systems. The next section addresses cell-free synthetic biology, which provides the possibility of synthetic genetic construct outside of cells.

1.5 Cell-free synthetic biology

This section provides an introduction into cell-free synthetic biology by first taking a glance at the history of cell-free biology. A brief overview of recent developments in the research area defined as cell-free synthetic biology is provided.

As the name cell-free biology suggests living organisms are not involved. Cell-free biology takes place outside the cell, but employs components and biomolecules used and synthesized by cells such as DNA, RNA and proteins. A large part within cell-free biology is based on cell lysates, which have been used for decades by biologists to investigate biological processes and mechanisms. They originate from different kinds of cells such as: rat liver cells, bacteria, human and rabbit reticulocytes, yeast, wheat germs and tobacco plants [59] [60]. Cell lysates provide the possibility of cell-free protein synthesis (CFPS) from supplemented DNA or mRNA. Researchers have used the fact that it is an open system to decipher molecular mechanisms. In 1961, Nierenberg and Matthaei were able to show that the triplet of UUU in RNA codes for the amino acid phenylalanine using purified RNA in a cell-free expression system. They thereby cracked the genetic code [61]. Since 1994, researches were able to unravel the complex process of translation initiation in eukaryotes using yeast cell extracts [62] [63]. Furthermore, Shin *et al.* have demonstrated the replication of the genetic material of the bacteriophage T7 and its self-assembly inside a cell-free expression system, providing a platform to further understand, engineer and produce phages [64].

At the interface of cell-free biology and synthetic biology there is a subdirection called cell-free synthetic biology, which combines elements from biochemistry, molecular biology and nanosciences [65] [66]. Recently there has been an increased interest in CFPS from researchers outside of biology due to the availability of simple protocols explaining how to create cell extract from *E. coli* [67] [68]. In addition there are commercial suppliers for cell-free expression systems based on crude cell extracts to purely reconstituted components, called the PURE[2] system [69]. With these open and relatively simple systems, researchers can create reductionist versions of a cell using only selected components to study molecular mechanisms under better control and accessibility. In 2005, Isalan *et al.* used artificial chambers filled with agarose, cell extract and patterned genes to emulate the pattern formation of the *Drosophila* embryo [70]. Another creative approach to study pattern formation was enabled by creating an artificial cell like compartment on a silicon chip that entails gene brushes and is continuously supplied with

[2]PURE stands for "protein synthesis using recombinant elements"

fresh cell-extract [71] (Fig. 1.3B). These compartments were then connected to study spatiotemporal patterns [72]. Recently, Dupin and Simmel used a controlled arrangement of droplets separated by a lipid bilayer [73] to study signaling and differentiation processes in artificial multicellular system [74].

Cell-free synthetic biology can be used to prototype genetic circuits. In 2003, Noireaux *et al.* have demonstrated the operation of synthetic gene circuits in a cell-free gene expression system [75] (Fig. 1.3A). Since then researcher have used cell-free expression system to prototype synthetic genetic constructs more rapidly than *in vivo* [76] [77]. Niederholtmeyer *et al.* have used a nano-liter microfluidic reactor that can maintain a cell-free expression reaction in a steady-state [78] to prototype genetic circuits that code for oscillatory protein expression before transforming them into bacteria [31] (Fig. 1.3C). In parallel, cell-free expression systems themselves are the focus of research. In order to understand their limitations and explore the possible applications, researchers are further characterizing and optimizing the constituent parts and reactions within cell-free expression systems [79] [80].

The chemical interface of cell-free expression systems with biological organisms has been used to explore applications in the field of diagnostics. In combination with paper-based microfluidics, RNA toehold switches and CRISPR-Cas9 technology, Pardee *et al.* have created a test that can detect a virus like Zika in purified blood samples and provide an optical read-out within hours [82]. Furthermore, cell-free systems with genetic circuits can sense chemicals in lungs samples from a *P. aeruginosa* infection which is a common danger for people with cystic fibrosis [83].

In recent years, cell-free expression systems have also been used to synthesize bioactive eukaryotic proteins that can be used for therapeutics [84] or personalized medicine [85]. In addition metabolic pathways are engineered within the cell-free expression system. For example, recently it was demonstrated that engineered cell-free expression system were able to transform glucose to valuable metabolites such as dihydroxyacetone phosphate (DHAP) or isobutanol, a next generation biofuel [86]. Finally, scaling-up of batch reactors of cell-free expression systems to about 100 liters has been demonstrated [30] making industrial application of cell-free metabolic engineering more feasible.

Furthermore, cell-free synthetic biology is connected to the field of bottom-up synthetic biology. The idea behind a large German research consortium called "MaxSynBio" is to re-create modules of living organisms from scratch and to combine them to create an artificial cell [87]. Exciting results from the module *replication and division* come from experiments where purified proteins that are naturally involved in cell-division in bacteria are encapsulated in vesicles. The proteins interact with the vesicle membrane and create

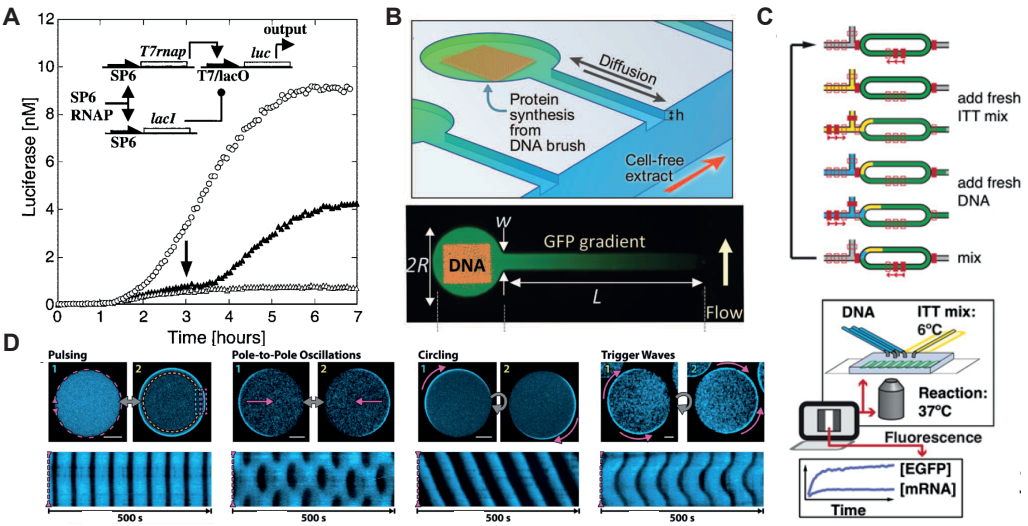

Figure 1.3: **A** Demonstration of a genetic circuit in a cell-free expression system. Luminescence of luciferase was used as a read-out. Open circles show the signal with IPTG (at t=0), close triangles with IPTG at t=3h and open triangles without IPTG. The subfigure is adapted from Ref. [75]. **B** Upper image: Sketch of an gene expression within an artificial mirco-compartment on a chip from gene brushes. The cell-free expression system is pumped and exchanges material with the the compartment over diffusion. Lower image: GFP expression from the DNA brush within the compartment. The subfigure is adapted from Ref. [71]. Reprinted with permission from AAAS. **C** Outline of a microfluidic system that operate a cell-free expression reaction in a steady-state with the use of valves. The reaction chamber can be imaged in real-time with an inverted microscope. The subfigure is adapted from Ref. [78]. **D** Oscillations and spatiotemporal patterns on vesicle walls by encapsulated proteins. The kymographs show pulsing, pole-to-pole oscillations, circling and trigger waves. The subfigure is adapted from Ref. [81].

oscillations and other spatiotemporal patterns [81] (Fig. 1.3D). Furthermore the protein oscillations can change the shape of the vesicle, which resulted in periodic budding and merging of vesicles. In the module *metabolism*, Schwander *et al.* were able to re-create the CETCH cycle, a carbon dioxide fixation pathway, *in vitro* with 17 enzymes from nine different organisms out of which three enzymes were engineered [88]. This opens up the possibilities for artificial photosynthesis.

Cell-free systems are furthermore at the heart of another topic of research called bionanosciences and molecular computing, which use mainly DNA to construct nanostructures or materials and perform calculations [89] [90] [91] [92]. DNA enzymes and DNA strands have been used to construct molecular circuits that create spatiotemporal patterns [93] [94]. The research in

these fields is closer to nanotechnology or computer science than to synthetic biology.

In conclusion, synthetic cell-free biology is a fairly new research field attracting researches from the outside of biology since it appears closer to an engineering discipline than synthetic biology in living organisms. Furthermore, it is closely tied to the field of bottom-up biology, that tries to create artificial cells from scratch. The full potential of cell-free synthetic biology lies in the fact that it is an open system. The biochemistry of living organisms hence can be more easily accessed and engineered than in cells. Having introduced bacterial and cell-free synthetic biology, the next section is about the structure and the content the thesis.

1.6 Overview of the thesis

The objective of the thesis is to explore the advantages of cell-free gene expression systems over living organisms and to elucidate to what extent one can use them to develop and prototype functional aspects of genetic constructs. Moreover, we have focused on synthetic RNA based regulators that can control the expression of genes or serve as a structural scaffold for proteins. Cell-free expression systems provide an open platform in which more parameters of gene expression are accessible. In addition, artificially structured environments like emulsion droplets or microfluidic chambers were introduced to facilitate the control of external parameters. The content of this thesis is nearly exclusively focused on *E. coli* and *E. coli* based cell-free expression systems.

The second chapter provides background information about gene expression, the control of gene expression and noise in gene expression. A brief introduction into CRISPR, CRISPR interference and RNA nanostructures is provided.

The third chapter looks at the establishment of chemical communication between *E. coli* and a cell-free expression system via autoinducer molecules. The content of this chapter has been published in Reference [95].

The forth chapter explores the expression and assembly of a multi-component protein-RNA nanostructure in a cell-free expression system, where the RNA serves as a scaffold for proteins. The content of this chapter has been published in Reference [96].

In the fifth chapter, we examine gene expression in a cell-free system and in *E. coli*, where CRISPR interference serves as repressor. Additionally we examine an anti-sense RNA approach to CRISPRi. The content of this chapter has been published in Reference [97].

Chapter 2

Background

2.1 Transcription and translation

In 1958, five years after the structure of DNA was resolved, Francis Crick introduced the central dogma of molecular biology [98]. It states that information for proteins is stored in DNA which is transcribed by enzymes into RNA that is then translated into proteins via an adapter molecule. The process of creating RNA from DNA is called transcription and is mediated by RNA polymerase (RNAP). The step from RNA to protein is mediated by the the ribosome and is called translation. In combination, the whole reaction from DNA to protein is termed gene expression or simply expression.

2.1.1 DNA and RNA

DNA and RNA both have their information encoded in four different nucleotides that differ in their bases[1]. The four bases for DNA are adenine (A), thymine (T), guanine (G) and cytosine (C). RNA shares the base of G,C and A but has an uracil (U) as the fourth base. A DNA or RNA strand is a sequence of bases connected by a phosphate-sugar backbone. RNA and DNA also differ in the form of the sugar present in the backbone. DNA has a deoxyribose whereas RNA has a ribose. The backbone has two different ends, a 5' and a 3' end, which are named after the number of the carbons in the sugar to which the first/last phosphate is attached. This provides the backbone with an orientation (direction). In living organisms DNA is normally double stranded, where two strands of opposite direction are intertwined forming a double helix. Within the helix the bases are paired according to rules called

[1]There also exit non-canonical bases, however in the scope of this section we limit the description.

Watson-Crick base-pairing and are as follows: C pairs with G and A with T or U for RNA. The information is encoded in the sequence of bases, and due to the base pairing rules symmetric and available in both strands. RNAP uses one of the DNA strands as a template and strings together ribose nucleotides according to base-pairing rules with the DNA template. The concatenation of such nucleotides is then the RNA strand. The next subsection describes this process in more detail by looking at transcription by the RNAP from the phage T7, which is widely used in the field of synthetic biology.

2.1.2 Transcription by T7 RNA polymerase

As described above, transcription is the polymerization of ribonucleotides (rNTPs) to form RNA oligomers. The reaction itself consists of three phases: *initiation, elongation* and *termination*. Here the process described will be limited to the *in vitro* transcription by the RNAP from the phage T7, which is a special case in the sense that the enzyme requires very few components for transcription - mainly, a DNA template, Mg^{2+} ions, and rNTPs. The initiation is proceeded by binding of RNAP to DNA. However to start transcription the RNAP has to denature the double helix. Only at a specific DNA sequence does the RNAP bind stably enough to start transcription and this is called the promoter region: the promoter region for the T7 RNAP is 23 base pairs (bp) long:

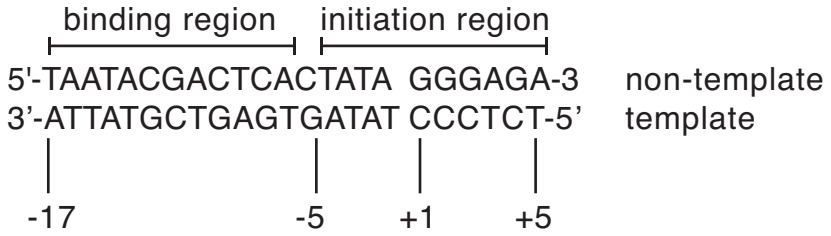

Figure 2.1: The T7 promoter sequence has 23 bases pairs (bp). It can be divided into two section, one for binding of the T7 RNA polymerase and one for the initiation of transcription. The first base that will be used in the RNA strand is at position +1.

Mutations in the first 12 bp affect the binding probability of the enzyme, whereas transcription initiation is sensitive to mutations in the remaining 11 bp. The first base to be transcribed and that begins the RNA sequence is at position +1 (Fig. 2.1). Once the RNAP has transitioned into elongation, it migrates along the DNA strand, transiently denaturing the double helix (transcription bubble) as it proceeds and catalyzing bond formation between

the nascent RNA strand and new ribonucleotides. The transcription bubble is about 8 bp long [99].The effective speed of transcription of the T7 RNAP was measured to be about 97 nucleotides per second [100]. The sequence of the RNA strand is complementary to the template strand of the DNA.

At the end of non-circular DNA strands the RNAP just falls off. However in plasmids and long DNA strands the transcription is terminated by transcription terminator sequences. A typical transcription terminator is made up of a hairpin structure followed by an A-U rich sequence. The hairpin structure folds on itself after transcription and pulls on the A-U rich DNA:RNA hybrid[2]. Transcription is terminated when the RNA:DNA helix dissociates due to the pull of the hairpin formation [101].

2.1.3 Transcription by *E. coli* RNA polymerase

The native RNA polymerase from *E. coli* consists of multiple components. For initiation of transcription a subunit called σ factor binds to the core enzyme which then mediates the binding of the complex, also called RNA polymerase *holoenzyme*, to the promoter sequence [102]. Once the complex has found and bound the promoter region, structural changes in the enzyme and DNA unwind the double helix and rNTPs can diffuse and position themselves on the template DNA strand. The enzyme catalyzes the formation of a phosphodiester bond between between adjacent nucleotides. The polymerase remains in this state until a strand of about 15 nucleotides has been synthesized. Then the core enzyme will undergo another conformational change and start transciptional elongation. The σ factor dissociates again and the polymerase proceeds at a speed of about 50 nucleotides per second [102] along the DNA strand. The transcription is terminated in the same manner as with T7 RNAP.

2.1.4 Translation

The ribosome, originally proposed by Francis Crick as an abstract adapter molecule between RNA and protein, was structurally resolved in 2000 [103]. Interestingly the ribosome itself is a RNA-protein structure (65% RNA and 35 % protein) [104]. The *E. coli* ribosome is a complex of three different ribosomal RNA strands (rRNA), two of which form the large subunit (50S) with 31 associated proteins. While the other strand forms the small subunit (30S) with 21 associated proteins [103]. The rRNA holds the mRNA in place and even catalyzes the peptide bond formation. The main role of the proteins

[2] A:T and A:U base pairs are energetically less stable than G:C base pairs.

seems to be stabilizing the RNA core [102]. So called transfer RNA (tRNA) molecules are adapters, which are charged with an amino acid and have a three base pair sequence (called the anticodon) that pairs with the three complementary bases on the mRNA (codon). For different amino acids there are different tRNA molecules with different anticodons. Hence the mRNA sequence codes for a specific chain of amino acids.

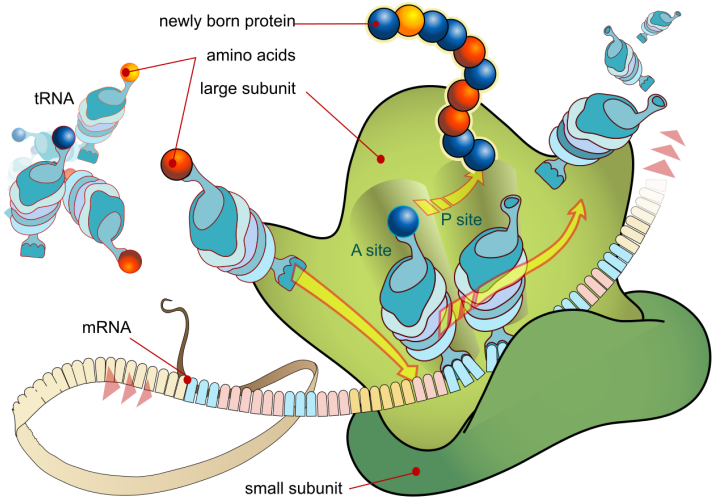

Figure 2.2: Schematic representation of translation. The small and large subunit of the ribosome assemble on the mRNA. tRNA molecules charged with amino acids are guided to the A site of the ribosome and bind to the corresponding codon on the mRNA. The amino acid on the tRNA at the P site is linked to the rest of the chain of amino acids. The uncharged tRNA molecule diffuses away, so that the process can repeat. The image is adapted from [105].

Each amino acid is encoded in three nucleotides, therefore exact initiation of translation is crucial as a shift by one nucleotide changes all codons. The sequence with the correct start and end is called open reading frame (ORF) and is flanked by a start codon (AUG) that codes for methionine, meaning all proteins start with this amino acid and stop codons (UAA, UAG or UGA), which do not code for an amino acid, but initiate the release of the peptide chain *via* the recruitment of *release factors* [102].

The assembly of the two ribosomal subunits occurs on the mRNA, and is promoted by ribosome binding sites (sometimes called Shine-Dalgarno sequence) on the mRNA which forms base-pairs with the small subunit of the ribosome. Once formation of the whole complex is initiated and the first tRNA has bound to the codon the ribosome starts to read the mRNA from the direction of the 5' end to the 3' end until it reaches a stop codon which

results in dissociation of the whole complex.

2.2 Controlling gene expression

Control over gene expression levels is one of the most important strategies for bacteria to adapt to changing environments. As bacteria are thought to be growing and dividing for about 3 billion years, this strategy has proved to be robust. The biochemical repertoire of an *E. coli* mostly comes from the few million proteins that are encoded by about 4000 different genes on its genome of about 4.6 billion base pairs [106]. As a response to changing external conditions, an *E. coli* can regulate the expression levels of the 4000 genes, and for that reason it has about 300 different gene regulators, also called transcription factors [106].

Regulation of protein expression can occur on the DNA, transcriptional and translational level. For example, there are small RNAs (sRNA) which are non-coding RNA molecules that bind to the region of the RBS on the mRNA thus blocking the assembly of the ribosome [107]. Furthermore, there are so called toxin-antitoxin protein systems, that selectively target the degradation of mRNA of certain genes [108]. In so called *riboswitches*, the mRNA creates a binding pocket for a small molecule which then stabilizes a certain conformation in the mRNA secondary structure that influences the amount of protein created from that mRNA [109].

For the scope of this thesis, we focus on a mathematical description of gene regulation by transcription factors. If one gene, gene a activates a second gene b, then the rate of production of gene b is a function of the concentration of gene a. By considering a simplified scheme based on mass-action kinetics, one can formulate the general case between the activator protein A that forms an n-mer and then binds to promoter sequence p (Fig. 2.3A and B).

$$\mathrm{nA} + \mathrm{p} \underset{\mathrm{k^{\cdot}-}}{\overset{\mathrm{k^{\cdot}+}}{\rightleftharpoons}} \mathrm{nA{:}p} \tag{2.1}$$

Then:

$$\frac{d[nA:p]}{dt} = k_+[A]^n[p] - k_-[nA:p] \tag{2.2}$$

where $[p]_0 = [p] + [nA:p]$ is the total concentration of promoters. Assuming a steady-state between the binding partners:

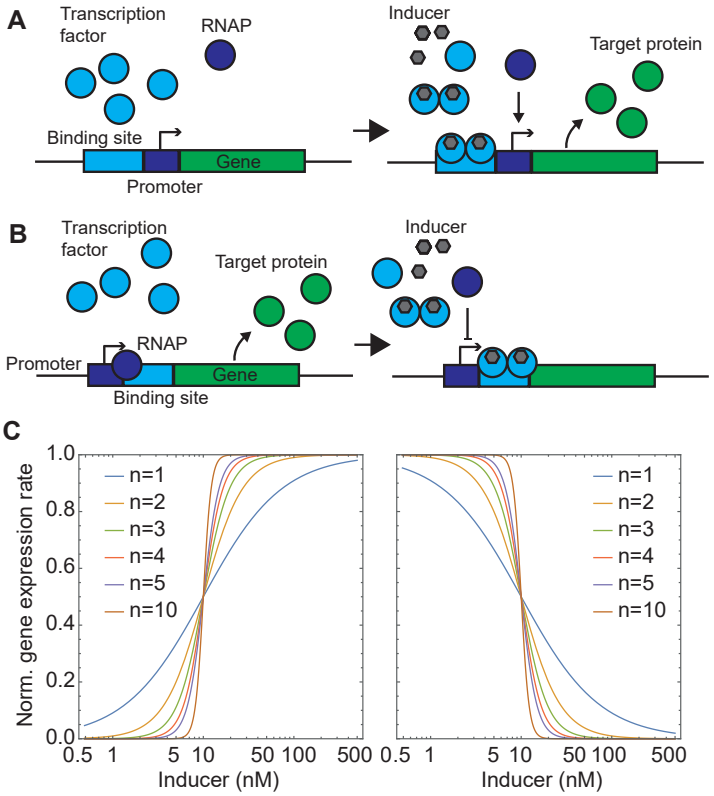

Figure 2.3: **A** Schematic representation of gene activation by transcription factors. In this example, the inducer molecule binds to the transcription factor and induces dimerization. The protein-inducer complex binds to the binding site next to the promoter and activates transcription of the target gene by RNA polymerase (RNAP). **B** Schematic representation of gene repression by transcription factors. Here the inducer binds to the transcription factor and induces dimerization of the transcription factor. The protein-inducer complex binds to the binding site which inhibits transcription of the target gene. **C** Hill curves with different Hill-coefficients. Left: the inducer activates gene expression (Eq. 2.11). Right: the inducer represses gene expression (Eq. 2.12. For both plots: $\alpha = 1$, $\alpha_0 = 0$ and $K_m = 100$ nM.).

$$\begin{aligned} \frac{d[nA:p]}{dt} &= 0 \\ \Rightarrow [nA:p] &= \frac{k_+}{k_-}[A]^n[p] \\ &= \frac{1}{K^n}[A]^n[p] \end{aligned} \tag{2.3}$$

with

$$K^n = \frac{k_-}{k_+}. \tag{2.4}$$

Then the fraction of occupied to total promoters $[nA : p]/[p]_0$ can be calculated from equation 2.3 by substituting $[p]_0 = [p] - [nA : p]$:

$$\begin{aligned} [nA : p] &= \frac{1}{K^n}[A]^n \left([p]_0 - [nA : p]\right) && (2.5) \\ \Rightarrow \frac{[nA : p]}{[p]_0} &= \frac{[A]^n}{K^n + [A]^n} && (2.6) \end{aligned}$$

The form of equation 2.6 is called a Hill equation. The concentration of protein B should be approximately proportional to the ratio of activated promoters [106] so that:

$$\begin{aligned} \frac{d[B]}{dt} &\approx \alpha \frac{[nA : p]}{[p]_0} && (2.7) \\ \Rightarrow \frac{d[B]}{dt} &= \frac{\alpha[A]^n}{K^n + [A]^n} + \alpha_0 && (2.8) \end{aligned}$$

where α is the maximal expression rate that can be reached in units concentration per time and α_0 captures the expression from non-activated promoters. n is the Hill coefficient that determines the steepness of the sigmoid function and quantifies the non-linearity in the relationship between concentration of gene a and b. K is the activation coefficient and marks the concentration of $A = K$ at which $\frac{d[B]}{dt}$ is at half-maximum.

In case the protein C from gene c represses the expression of protein D from gene d the Hill equation can be derived in analogous manner to equation 2.8.

$$\begin{aligned} \frac{d[D]}{dt} &\approx \alpha \frac{[nC : p]}{[p]_0} && (2.9) \\ \Rightarrow \frac{d[D]}{dt} &= \frac{\alpha K^n}{K^n + [C]^n} + \alpha_0 && (2.10) \end{aligned}$$

There are quite some simplification steps from equation 2.6 to 2.8 or 2.10 as one infers the expression rate of the protein from the fraction of activated promoters by using a linear relationship. In addition, the Hill equation in this form is often used to describe the response of a gene to a certain concentration of a small inducer molecule m which activates/inhibits a transcription factor

$[m] + [A] \rightarrow [A^*]$. In this case the transcription factor is constitutively expressed but remains inactive until bound by the inducer, and its exact concentration is unknown. In contrast the concentration of the small inducer molecule $[m]$ is known and if one assumes that $[A^*]$ is proportional to $[m]$ then for activation one can describe the change in concentration of B:

$$\frac{d[B]}{dt} = \frac{\alpha[m]^n}{K^n + [m]^n} + \alpha_0 \tag{2.11}$$

and for repression:

$$\Rightarrow \frac{d[D]}{dt} = \frac{\alpha K^n}{K^n + [m]^n} + \alpha_0 \tag{2.12}$$

where K and n are now the induction threshold and Hill coefficient for the small molecule. Equations 2.11 and 2.12 can be used to describe the response in gene expression rate due to the concentration of a small genetic inducer (Fig. 2.3C). In synthetic biology, genetic switches are often controlled by small molecules as they can be added externally and permeate the membranes of bacteria. These equations serve to describe the input-output behavior of such a switch.

2.3 Noise in gene expression

In the previous section gene regulation based on mass-reaction kinetics was discussed. However in systems where reactants are below are certain number, the reactions that are based on two molecules colliding become stochastic. That means the randomness of when two reactants meet becomes noticeable as fluctuations in the speed at which chemical reaction progress. Single cell analysis in the study of the repressilator circuit in *E. coli* showed the performance of the oscillator was affected by the stochastic nature of some reactions inside bacteria. Since then the origins and the quantification of noise in gene expression is a research focus in synthetic biology [22].

In 2002, Swain and Elowitz started to distinguished two sources of noise in gene expression: the first source is the "intrinsic" noise from the stochastic nature of chemical reactions within a cell (Fig. 2.4). This will be uncorrelated between cells. The second source of noise is called "extrinsic" noise from fluctuations and differences in the outside conditions of the cell. These fluctuations are shared by all cells since they are exposed to the same conditions [110].

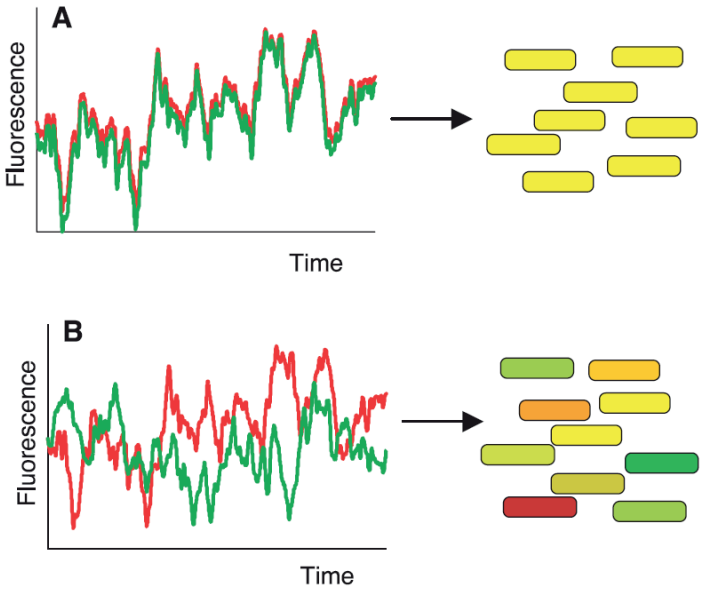

Figure 2.4: **A** The concept of extrinsic noise for gene expression within a bacterial population. As all bacteria are subjected to the similar conditions , the fluctuations in gene expression of two different genes (red and green) with the same promoter correlate. **B** An illustration of the concept of intrinsic noise. The fluctuations in gene expression of two different genes (red and green) due to intrinsic noise does not correlate between individual cells. The figure is adapted from Ref. [111]. Reprinted with permission from AAAS.

By introducing stochasticity the transcription and translation model was able to reproduce the cell-to-cell variations in a population of bacteria. Transcription is described to occur only rarely and mRNA molecules are degraded quickly, so that very few mRNA molecules exist at a time. However, translation comes in bursts, where about 50 proteins are created from one mRNA molecule. This model produces a gamma-distribution for the protein concentration across cells. This was confirmed by *in vivo* experiments [112].

However for abundant proteins (above about 10/cell), stochastic bursting can not explain the cell-to-cell variation [113]. Then usually extrinsic noise is used as an explanation without further explanation of what that actually entails. Recently, there are more arguments that the gamma distributions for the protein concentration across population of cells is a special case that occurs when the transcription rate for that protein is extremely low [114]. In the absence of extreme condition such as the very irregular transcription rate, protein concentrations across a bacterial population are expected to be

log-normally distributed. This might become more important for synthetic biologists, as they are able to engineer parameters such as transcription rates in their genetic constructs.

2.4 Using chemical reaction networks to model a cell

This section looks at quantitative descriptions of gene expression and their accuracy when stochasticity is taken into account. Stochasticity can become important for the prediction and design of genetic switches and circuits in synthetic biology. It can also offer a more accurate picture of the cell-to-cell variations of mRNA and protein concentrations in an isogenic population of bacteria [115]. The closer synthetic biology and engineered bacteria come to application the more important it is to consider not only the genetic construct as such but also the fact that bacteria mostly occur in populations with some underlying heterogeneity between individual cells. In addition any protein species of interest will be only one species in a thousand. In order to simplify this pictures and to predict the effect of this heterogenity on genetic circuits performance, one can use computational models that capture the fundamental mechanisms.

The following paragraph provides an overview of an algorithm published by Furusawa and Kaneko [116] [117], in which a bacterial cell is reduced to chemical reaction networks. The goal of the model system is to derive experimental observations such as the cell-to-cell variations in a protein concentration within a homogeneous population of cells from abstract models that require a minimal amount of biological details. This way the model may remain more general and help in the development of engineering principles for synthetic biology.

The internal state of a cell is described by a set of m different chemical species within a cell volume v. The number of molecules of each species is represented by n_i with i ranging from $i = 1$ to m, with a total amount of molecules in the cell $N = \sum_i^m n_i$. Then a randomly chosen catalytic network between the species is created, where some chemical k catalyzes the reaction of chemical i to j ($i + k \rightarrow j + k$). All reaction coefficients have the same values and the probability of two chemicals to be connected (eg. i and j) is given a connection rate ρ (the connection paths are chosen randomly). Furthermore, there are some species that can diffuse between the environment and the cells with a diffusion coefficient D, among which are the nutrients that have no catalytic activity and are constantly supplied

by the environment. The permeable nutrients are catalyzed to impenetrable chemicals so that the number of chemicals in a cell N increases until a certain threshold N_{max} is reached and the cell is divided (the cell's molecules are randomly split between the two daughter cells). The growth and division process is repeated. Any two molecules taking part in a chemical reaction are chosen randomly, same for any molecule in the diffusion process outside or inside the cell. There is a critical value for $D = D_c$ above which all the nutrients flow so fast into the cell that the chemical reactions cannot keep up with the influx. Continuous growth and division of cells is only possible for $D \leq D_c$.

Power-law distribution of abundance of chemical species

The simulation allows to calculate the abundance of each chemical species. The result of the rank-ordered distribution, which ranks each chemical species according to their abundance[3] follows a power-law that converges to the exponent of -1 at the critical point $D = D_c$ [116]. Interestingly, there are several examples of rank-ordered frequency plots with a similar shape found in experiments. For example, the mRNA species in human liver, *C. elegans* and yeast cells follow a power law with an exponent close to -1 [116]. This occurrence of rank-ordered frequency distribution with a power law exponent of or close to -1 in various research areas is called Zipf's law, after the linguist George K. Zipf, who studied the frequency of words in a language [118]. The result tells us that we can expect that relative changes in the concentration of one component will result in proportional changes in the concentration of another component, independent of their initial concentrations.

Log-normal distribution of a chemical species within a population of cell

The simulation by Furusawa *et al.* can also be applied to the question of cell-to-cell variations by looking at the distribution of each chemical's abundance over many cells for a large number of divisions [117].

Their results fall into two different cases, depending on whether the total sum of chemicals N_{max} is larger or smaller than the amount of chemical species m. When $N_{max} > k$, so there is more than k one molecule of each species, the population ratio, defined as $\frac{n_i}{N_{max}}$ is stable over division events so that daughter cells have a similar chemical composition to their mother cell. Then the abundances of all chemicals that are reproduced in the cell follow

[3]For example, checmical species i has rank 2 if n_i is the second largest value in the set of $n_1, ..., n_m$.

a log-normal distribution. When $k > N_{max}$, the population ratios are not stable and change each generation. Now, the resulting distribution for each chemical follows a Poisson-like distribution. However, for most cases where the chemical species do not change from generation to generation the former case with $N_{max} > k$ is more applicable. The main reason for the log-normal distribution of the chemicals comes from a small group of dominant reaction pathways that self-organized in a cascade-like fashion so that fluctuations in the concentration of one species propagate in a multiplicative way to the concentration of the next species. Experimental data from a bacterial culture expressing a fluorescent protein matched the result of the calculation. The the concentration of the protein across a bacterial populations was log-normal distributed [117].

Gene expression: chemical complexity leads to a log-normal distribution

There are a number of studies dealing with the log-normal like distribution of gene expression across bacterial populations. The following paragraph will outline an argument by Jacob Beal, that provides more of an heuristic explanation to why the log-normal distribution is found for many genes and why it is an emergent property of a complex reaction network [114].

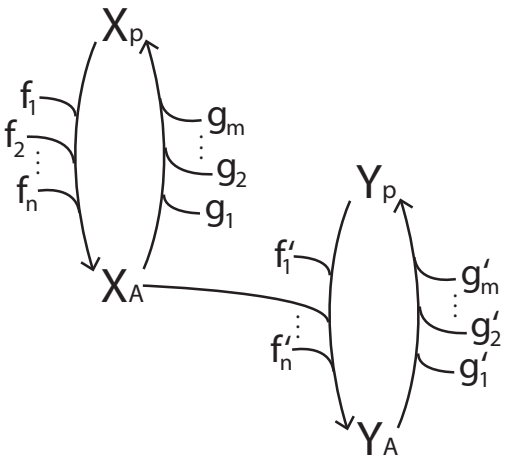

Figure 2.5: A schematic visualization of the chemical complexity in gene expression. Some molecular species X involved in gene expression has a passive state X_P and an active state X_A. Several other molecular species f enhance state X_A whereas species g enhance state X_P. Molecule X_A itself may be one of many enhancing factors that stabilizes another molecular component in gene expression to become active Y_A. The image is adapted from Ref. [114].

Let us pick one molecule X that is included in the process of gene expression and has two states: an active state (X_A) and a passive state (X_P). There are plenty of other molecules in the cell that serve as catalytic factors $f_1, f_2, ..., f_n$ that change the state of X from passive to active. Then again there is another set of chemical factors $g_1, g_2, ..., g_m$ that catalyze the active molecules (X_A) back to a passive state (X_P). For example, molecule X could be a RNA polymerase, that is transcribing a gene (active state) after it was recruited by a sigma factor such as protein σ^{70} (catalytic factor f). At the same time, for example, an anti-sigma factor like Rsd (catalytic factor g) can inhibit the association of RNA polymerase to a sigma factor so that the RNA polymerase stays in the cytosol (passive) (Fig. 2.5).

The rate equation for the activated molecule X_A can be formulated on the grounds of standard chemical reaction laws:

$$\frac{dX_A}{dt} = \left(k_x \cdot [X_P] \cdot \prod_{i=1}^{n} [f_i] \right) - \left(\lambda_x \cdot [X_A] \cdot \prod_{i=1}^{m} [g_i] \right) \tag{2.13}$$

where k_x and λ_x are reaction rate constants. Assuming that the expression of the gene under consideration does not dominate cellular resources drastically, the state of the system is quasi-static and the expected equilibrium concentration of X_A is given by:

$$0 = \left(k_x \cdot [X_P] \cdot \prod_{i=1}^{n} [f_i] \right) - \left(\lambda_x \cdot [X_A] \cdot \prod_{i=1}^{m} [g_i] \right) \tag{2.14}$$

$$[X_A] = \frac{k_x}{\lambda_x} \cdot [X_P] \cdot \prod_{i=1}^{n} [f_i] \cdot \prod_{i=1}^{m} [g_i]^{-1}. \tag{2.15}$$

The concentration of X_A is given by the product of input concentration ($f_1, f_2, ..., f_n, 1/g_1, 1/g_2, ...1/g_m$). Consequently the distribution of X_A is proportional to the product of many distributions. This argument can be extended such that X_A now serves as the catalytic factor for another molecule Y, promoting its active state Y_A. For example Y could be nucleotides (passive) which are catalyzed via the RNA polymerase X_A to mRNA (active). Then the equilibrium concentration of Y_A follows from equation (2.15) and is given by:

$$[Y_A] = \frac{k_y}{\lambda_y} \cdot [Y_P] \cdot \prod_{i=1}^{p} [h_i] \cdot \prod_{i=1}^{q} [j_i]^{-1} \tag{2.16}$$

$$[Y_A] = \frac{k_y k_x}{\lambda_y \lambda_x} [Y_P][X_P] \prod_{i=1}^{n} [f_i] \prod_{i=1}^{m} [g_i]^{-1} \prod_{i=1}^{p} [h_i] \prod_{i=1}^{q} [j_i]^{-1} \tag{2.17}$$

where k_y and λ_y are reaction rate constants, and h_i and j_i catalytic factors for Y analog to f_i and g_i for X. Now the distribution of Y_A is a product of the distribution of X_A and so on. The line of argument can be continued at least several times. The central limit theorem states that the sum of independent random variables converges to being normal distributed. Additionally, since the logarithm of a product is equal to the sum of the logarithms, the product of the independent random variables converges to a log-normal distribution. This means that the distribution which results from a product of many independent random variables is a log-normal distribution. Hence, due to the large amount of factors involved in expression of a protein, the expected distribution in a homogeneous bacterial population approaches a log-normal distribution. In practice, already a few factors suffice to create a distribution that is very close to a log-normal distribution.

Geometric statistics

The result from the previous section is that concentration of most proteins within a population of bacteria will follow a log-normal distribution. Since this is a skewed distribution, the statistics we use to evaluate and present the data should be adjusted. As for most skewed distribution geometric instead of arithmetic statistics can be more accurate as argued in the following paragraph.

A random variable X is log-normally distributed when $Y = \log(X)$ is normally distributed. Hence to characterize the log-normal distribution one can use the statistical tools which are applied to normal distributions to the log-transformed data. For example, for log-normal distributed data one can take the logarithm of the values and work with the log-transformed data (i.e. $Y \sim \mathcal{N}(\mu, \sigma^2)$)[4]. The arithmetic mean μ and the standard deviation σ of the log-transformed data serve as a measure of location and dispersion with the interval $\mu \pm \sigma$ containing 68.3% of the data.

However, there might be advantages to working directly with the non-transformed data. If so, one has to consider the fact that the log-normal

[4]In this chapter, the symbols μ and σ always describe the arithmetic mean and standard deviation of a normal distribution.

distribution is asymmetrical around the mean with a longer tail towards values that lie above the mode of X. This implies that the arithmetic mean μ generally used for the description of a normal distribution would overestimate high-valued outliers and thus strongly deviate from the median.

The expected value $E(X)$ (also known as the 1^{st} moment) or mean μ_{LN} of the log-normal distribution is given by:

$$\mu_{LN} = e^{\mu + \frac{\sigma^2}{2}}. \tag{2.18}$$

However, due to the fact that the log-normal distribution is positively skewed, the median might be more interesting as a measure of location. Since the mean μ is also the median of the log-transformed data, e^{μ} is the median of the non-transformed data. Then the median of the log-normal distribution can be determined by calculating the geometric mean μ_g of the original data:

$$\mu_g(x_1, ..., x_n) = (x_1 \cdot ... \cdot x_n)^{1/n} = e^{\frac{log(x_1)+...+log(x_n)}{n}} = e^{\mu} \tag{2.19}$$

where $x_1, ..., x_n$ is the measured data.

Furthermore the arithmetic standard deviation σ is also not an ideal representation for the dispersion of the data since the symmetric ranges around a value such as the mean do not have the same probability [119]. Here an underlying difference between a normal distribution, which is realized by the additive sum of random variables and the log-normal distribution, which is realized by the multiplicative product of many random variables surfaces. As for the case of gene expression, fluctuations in the concentrations of the factors (see eq. (2.17)) involved in the process propagate multiplicatively, so that multiplication/division of a factor should be preferred to represent dispersion around a value rather than addition/subtraction (e.g. $\pm\sigma$). The arithmetic standard deviation σ_{LN} of the log-normal distribution[5] is given by:

$$\sigma_{LN} = e^{\mu + \frac{\sigma^2}{2}} \sqrt{e^{\sigma^2} - 1}. \tag{2.20}$$

The back-transformed standard deviation, e^{σ}, is termed the geometric standard deviation[6] σ_g and can be estimated by:

$$\sigma_g(x_1, ..., x_n) = e^{\sqrt{\frac{1}{n-1} \cdot \sum_i^n (log\, x_i - log\, \mu_g)^2}}. \tag{2.21}$$

[5]The standard deviation (SD) is given by $SD[X] = \sqrt{Var[X]} = \sqrt{E[X^2] - (E[X])^2}$.

[6]Also called multiplicative standard deviation.

where $x_1, ..., x_n$ is the measured data. Now it is possible to re-define the interval $[\mu - \sigma, \mu + \sigma]$ associated with the normal distribution for the log-normal distribution as $[\mu_g/\sigma_g, \mu_g \cdot \sigma_g]$, which also covers 68.3% of the probability [120]. Hence, it is advised to express gene expression data using the geometric mean (= median) with error bars representing the geometric mean times the geometric standard deviation for the upper bound and divided by the geometric standard deviation for the lower bound (e.g. $\mu_g \,^{\times}\!/\, \sigma_g$) [114].

In case that the arithmetic mean of a sample $\bar{x}$ and the sample standard deviation s of some gene expression data is known, one can estimate the sample geometric mean [120]:

$$\bar{x}_g = \frac{\mu}{\sqrt{1 + \left(\frac{s}{\bar{x}}\right)^2}} \tag{2.22}$$

and the geometric standard deviation:

$$\bar{x}_g = e^{\sqrt{\log\left[1+\left(\frac{s}{\bar{x}}\right)^2\right]}}. \tag{2.23}$$

It should be mentioned again, that gene expression only follows the log-normal distribution if there are no other extreme conditions forcing it into another regime.

2.5 CRISPR-Cas9

This section provides a short introduction into the origins, mechanisms and applications of CRISPR-Cas9. Bacteria and archaea have an adaptive immune memory called CRISPR (clustrered regularly inter-spaced short palindromic repeats)-Cas(CRISPR-associated genes). In short, it works by inserting snippets (around 20 bp) of foreign genetic material from viruses or plasmids as spacers in CRISPR-arrays (Lamarckian-type inheritance) [121]. RNA molecules encoding the spacer sequences are used as recognition elements to find matching foreign genetic material which is then destroyed by Cas proteins [122].

Individual Cas proteins have become the subject of intensive research, one of the most prominent examples being the Cas9 endonuclease from *Streptococcus pyogenes*. In contrast to many other CRISPR systems, the CRISPR-Cas9 system involves only one protein in the effector complex cleaving the foreign DNA.

First the CRISPR locus is transcribed forming a long precursor CRISPR RNA (pre-crRNA) containing multiple spacer sequences and direct repeats [123]. A second RNA molecule, the trans-activating crRNA (tracrRNA)

binds to the direct repeats on the pre-crRNA. This long RNA complex is further dissected and processed by RNase III and some unknown nuclease so that one complex contains one tracrRNA, one direct repeat and one spacer sequence of about 20 nucleotides. In this configuration the RNA complex associates with Cas9 which is then guided towards the target sequence on the foreign DNA given that there is a required protospacer adjacent motif (PAM) present next to the target site, which is NGG[7] for Cas9 from *Streptococcus pyogenes*. Once bound to the DNA it cleaves the bonds in the sugar phosphate backbone of both DNA strands.

In 2012, the CRISPR-Cas9 system was engineered to improve its function as a gene editing tool [124]. Within the next year, it was found to be applicable for genomic editing in eukaryotic cells including yeast [125], plants [126] [127] [128], mouse [129], zebrafish [130] and even human cells [131] [132] [133].

In brief, the genome editing process entails identifying the target sequence in the genome with a PAM then designing and transcribing the corresponding RNA to guide the Cas9 protein. The resulting double strand break (DSB) activates the DNA repair mechanisms of the cell. In the presence of large quantities of donor template DNA containing the desired insert which are flanked by homologous regions, the homology-directed repair (HDR) pathway integrates the insert from the template DNA into the chromosome. Alternatively, in the absence of donor template DNA, the DSBs can be repaired by non-homologous end joining (NHEJ), which results in non-precise deletions or insertions of a few nucleotides.

One of the major developments of the CRISPR-Cas9 system was the merging of the two processed RNA molecules, crRNA and tracrRNA, to one single RNA molecule named single-guide (or synthetic-guide) RNA (sgRNA) [124]. In eukaryotic cells the removal of a putative Pol-III terminator and a stem-loop extension improved the sgRNA stability and enhanced the assembly with a Cas9 mutant [134].

Furthermore, different versions of the Cas9 protein were engineered by mutations and fusions. For example, the mutants Cas9D10A or Cas9H840A each cut only one of the DNA strands [124], whereas the version known as dCas9 (with both mutations D10A and H841A) does not cleave DNA at all [53]. The method of suppressing transcription of genes with dCas9 by either blocking the binding of the RNAP to its promoter or the elongation during transcription is named CRISPR interefence (CRISPRi). On the other hand, CRISPR activation (CRISPRa) enhances the transcriptional activity of a gene. For example dCas9 fused to the RNAP omega subunit (ω) that

[7]N stands for any arbitrary nucleotide.

stabilizes the binding of RNAP to its promoter increases the expression of a gene when placed upstream of its promoter [135]. This technique to position specific proteins on the genome via dCas9:sgRNA has been used in different ways.

2.6 CRISPRi

This section looks at CRISPRi as a means to regulate gene expression within bacteria. As described above, the method uses the mutated dCas9 protein which can be guided to a specific DNA site using the appropriate sgRNA sequence. The only requirement is the presence of a protospacer adjacent motif (PAM) next to the target sequence on the DNA.

The sgRNA consists of a 20-nt target specific complementary region, a Cas9-binding RNA structure and a transcription terminator derived from *S. pyogenes*. The sgRNA binds to the dCas9 and then finds the target sequence on the DNA.

It has been shown that the bound dCas9-sgRNA complex downstream from a promoter can stop the RNAP from transcription elongation [53] (Fig. 2.6A). Elongation of the RNAP is stopped more efficiently when the non-template DNA strand for transcription is targeted by the sgRNA than the template strand (Fig. 2.6B). In this case the PAM sequence is CCN at the 5' end of the non-template DNA strand. There are still differences in the repression efficiency depending on the position within the gene of interest that should be repressed. Results can vary but in general the shorter the distance of the target sequence to the start codon of the gene that should be repressed the stronger the repression [53]. Double or multi repression of a gene is stronger as long as the repression sequences do not overlap [53] [138]. In addition, the number of base pairs in the guide RNA that are complementary to the DNA influence the repression strength of the gene expression [137]. Finally the repression efficiency is strongly reduced above temperatures of 37°C and is lower for stronger transcription promoters. This has allowed the formulation of a "kick-out" out model where the RNAP that approaches the dCas9-sgRNA complex is either stopped with a certain probability $P(stop)$ or kicks the dCas9-sgRNA complex off the DNA and proceeds with $1 - P(stop)$ [137] (Fig. 2.6C). According to this model the expression level of the target gene α is given by:

$$\alpha = \alpha_0 \left(1 - P(stop) \cdot P(bound)\right) \tag{2.24}$$

where α_0 is the native transcription rate and $P(bound)$ is the probability

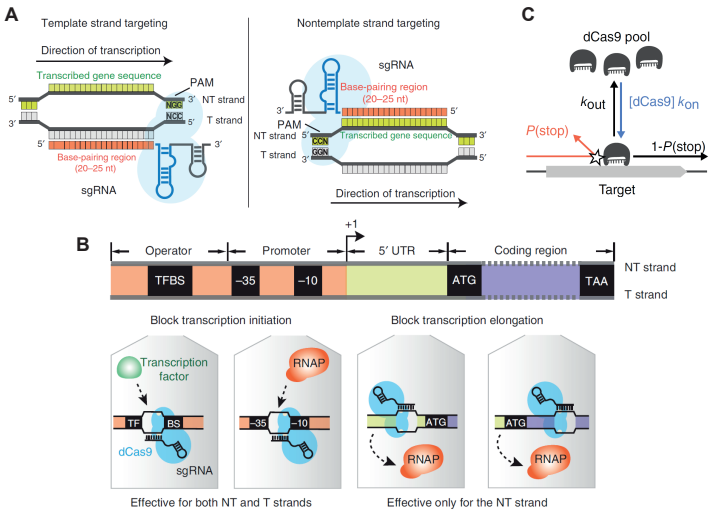

Figure 2.6: **A** CRISPR interference. Illustration of the sgRNA-Cas9 complex bound to the template DNA strand (left) and the non-template DNA strand (right) relative to the direction of transcription. The subfigure is adopted from Ref. [136] by permission from Springer Nature, 2013. **B** Depending on the target position on the DNA, CRISPRi was different operational mechanisms. Transcription initiation is blocked if the operator or promoter sequence are bound. Here both the template and non-template strand can be targeted. In case the target sequence is downstream of the transcription start, the sgRNA-dCas9 complex stops transcriptional elongation. Here only the non-template strand should be targeted. The subfigure is adopted from Ref. [136] by permission from Springer Nature, 2013. **C** Scheme of the "kick-out" model. P(stop) is the probability that transcription elongation is stopped when the RNAP runs into the bound sgRNA-dCas9 complex. The 'occupancy' of sgRNA-dCas9 is determined by the concentration of sgRNA-dCas9 and the binding and unbinding rates (k_{on} and k_{out}). However, the fact that incoming RNAP can kick out bound sgRNA-dCas9 complex is incorporated into the unbinding rate k_{out}. The subfigure is adapted from Ref. [137] by permission from Springer Nature, 2018.

that dCas9-sgRNA is bound to the target. This probability is given by:

$$P(bound) = \frac{k_{on}[dCas9]}{k_{on}[dCas9] + k_{out}} \tag{2.25}$$

where $k_{on}[dCas9]$ is the rate of binding and k_{out} is the rate of dCas9-sgRNA ejection. The ejection of the complex can be spontaneous unbinding or displacement by RNAP or possibly occur during DNA replication [137] [139].

Besides targeting the non-template strand in the codon region of a gene,

one can also target the template strand and the promoter region. It has been shown that targeting the template strand in the codon region is less efficient in repressing gene than the non-template strand [53]. Targeting the promoter region has yielded different results. Apparently here the dCas9-sgRNA complex blocks the RNAP from either binding to the promoter sequence or transitioning from transcription initiation to elongation. Hence it is important that the target sequence is within the region where the σ factor binds to the RNAP core enzyme. Here, in contrast to the blocking of transcription elongation, the difference between the template and non-template strand is not visible [53].

It has also been shown that the CRISPRi induced gene knock-down is reversible by dilution in bacteria by stopping to express dCas9-sgRNA, although full recovery of gene expression can take 6 - 15 hours [53] [140]. Furthermore it was demonstrated that an anti-sense RNA to the sgRNA was able to recover gene expression in bacteria even in the presence of dCas9-sgRNA expression [141].

In summary, CRISPRi provides a method to silence genes, which is programmable by changing the target sequence on the sgRNA. This potential is being harnessed to create multi-layered logic or change gene expression from the chromosome of the bacteria from a plasmid [55]. However, recently it has been also reported, that high levels of dCas9 without sgRNA in *E. coli* can cause growth defects such as stopping cell division and that specific target sequences can cause severe growth defects or even kill bacteria [142] [143].

2.7 RNA nanostructures

One of the applications of synthetic biology is the creation of biomolecular structures of particular interest. The following section describes the origins, developments and techniques of nucleic acid nanostructures with a focus on RNA nanostructures. In 1982, Nadrian Seeman proposed the idea of using biomolecules to build artificial structures by using artificial DNA strands to construct tiles and lattices [144]. This opened up a new field called DNA nanostructure and nanomaterials [145]. Since then DNA has been used to construct lattices, two and three-dimensional objects of nearly arbitrary shapes. Especially, a technique called 'DNA origami' allowed the creation of more complex structures [90], where a long strand of phage DNA is folded to a desired shape by small oligo-nucleotides called staple strands. Using this technique researchers were able to create DNA envelopes for enzymes that can be opened by RNA or DNA strands, create DNA nanopores in lipid membranes or create gigadalton sized assemblies of several DNA origamis

[146] [147] [148]. Recently, Praetorius and Dietz have used this principle to fold double stranded DNA into designed shapes with custom designed staple proteins [149].

In parallel, researchers have started to use RNA for the construction of nanostructures. In comparison to DNA, RNA is chemically less stable, but can be encoded genetically and produced enzymatically. Furthermore, living organisms have developed structural RNA motif that can be integrated in artificial structures. In 2000, Jaeger and Leontis used so called RNA tectonics to creates supramolecular RNA filaments by self-assembly due to interactions between hairpin tetraloops [150] (Fig. 2.7A). This technique was further developed to create programmable 2D and 3D structures [151] [152]. Furthermore, the DNA origami method was extended to the use of RNA [153]. As there is a library of structural RNA modules found in nature, scientists can also create different shapes by combining different modules [154]. Recently a new set of design rules for RNA tiles made from a single RNA strand has been proposed [155]. The size of the tiles ranges from 156 nt to 660 nt, can be produced enzymatically, fold co-transcriptional and form different lattice assemblies on surfaces.

Interestingly RNA structural motifs are not limited to RNA-RNA interactions. RNA structural motifs termed RNA aptamers can bind a ligand such as small molecules, proteins or even lipids. New RNA aptamers can be evolved, selected and enriched from a random sequence library in cycles in a well defined method called SELEX (Systematic Evolution of Ligands by EXponential enrichment) [159]. For example, aptamers have been developed to bind molecules and enhance their fluorescence [160]. RNA aptamers were also evolved in nature. For example, phage RNA often has a hairpin structure at the start codon of the viral replicase which can bind the coat proteins of that phage in order to repress the replication and facilitate packaging [161]. RNA nanostructure can be functionalized by the integration of RNA aptamers. For example, Delebecque *et al.* were able to combine aptamers that positioned two enzymes within a larger RNA nanostructure in proximity to each other [162]. As both enzymes were involved in the production of hydrogen, it was demonstrated that the scaffolding of the enzymes *via* the RNA increased the hydrogen production in bacteria.

Single stranded RNA is actively degraded in bacteria and only lives for a few minutes [106]. However, as there are stable RNA species in cells and phages, there must exist strategies to avoid degradation. One remarkable example is a three-way juction motif from an RNA strand in the phage $\phi 29$ that scaffolds a protein pore which is used for the translocation of the genome through the cell membrane [163]. This three-way junction was shown to be extremely resistant against degradation and denaturation [164] [165]. Using

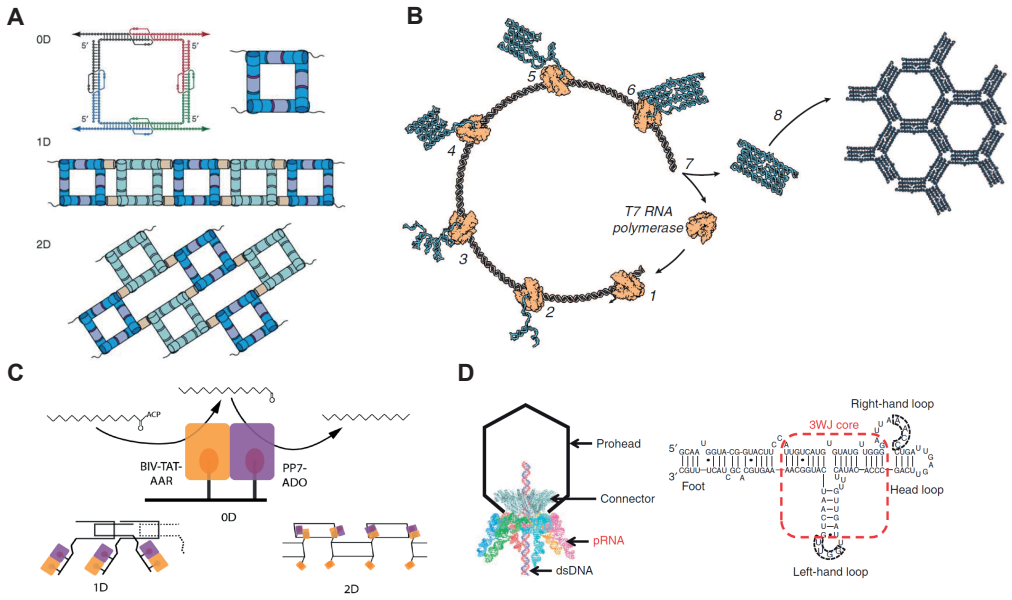

Figure 2.7: **A** An illustration of the RNA tectonic mechanism. Natural RNA structures can be decomposed into structureal motifs. These can be combined again in new ways from artifical and larger structures. The subfigure is adopted from Ref. [156]. Copyright (2006), with permission from Elsevier. **B** Schematic representation of the co-transcriptional folding of an RNA tile which assembles into a supramolecular lattice on a mica surface. The subfigure is adopted from Ref. [155]. Reprinted with permission from AAAS. **C** Illustration of an RNA scaffold for two enzymes. Due to their proximity a reaction pathway involving the the two enzymes is accelerated. The RNA scaffold can be incorporated into larger RNA nanostructures. The subfigure is adopted from Ref. [157] by permission of Oxford University Press, 2014. **D** The phage $\phi 29$ uses an RNA scaffold for the complex that trans-locates the DNA into its prohead. A three-way junction motif in these pRNA molecules has shown to be exceptionally stable against denaturation. This subfigure is adopted from Ref. [158]. Reprinted by permission from Springer Nature, 2013.

structural motifs like this can enhance the stability of RNA nanostructures [166].

In conclusion, RNA offers the possibility to genetically encode and enzymatically produce artificially designed nanostructures. The interactions of RNA with small molecules, DNA and proteins and its own catalytic activity provide the possibility to combine several functional aspects within RNA nanostructures. In short, RNA nanostructures have probably more potential applications than DNA nanostructures within living organisms and are easier

to design than proteins.

Chapter 3

Chemical communication between a cell-free expression system and bacteria

3.1 Biological background and motivation

The contents of this chapter have been published in the article: "Chemical communication between bacteria and cell-free gene expression systems within linear chains of emulsion droplets" by Matthaeus Schwarz-Schilling, Lukas Aufinger, Andrea Mückl and Friedrich C. Simmel in *Integrative Biology*, 2016,8, 564-570.

Multicellular organisms often develop from one or a few cells that divide. The progeny then differentiates into different types of cells that take part in fulfilling different functions [167]. In this sense multicellular organisms can achieve functions unachievable for single cellular organisms by *division of labor* using functional modules such as an eye, a muscle or a brain. The development and maintenance of such an architecture requires coordination and communication between cells. The field of embryogenesis studies the development of organisms and has identified the importance of chemical gradients across one cell or several cells as a means to provide orientation and positional information from which further development can occur [168] [169]. For example, the anterior-posterior axis in the embryo of *Drosophila melanogaster* is determined by pre-positioned mRNA molecules that create a gradient across the early oocyte [170].

Most of the time the chemical signals, also called morphogens, encode information in their concentration as a function of position and time. The natural mechanism of transport of chemicals is diffusion. There are different

ways of signal transmission between or within a cell: by simple diffusion across membranes, through protein channels or pores, within vesicles or by active transport via molecular motors along pre-determined pathways [102].

As artificial cellular systems are easier to control and engineer than living organisms, but most of the medical and biotechnological applications come from cells, chemical communication and coordination between artificial systems and living organisms provide the possibility for an interface between engineered systems and the biological world.

3.1.1 Artificial biomimetic systems

An advantage of artificial biomimetic systems over natural environments is that they are often more controllable. This allows for a suitable environment to both prototype modules, similar to aeronautical wind tunnels in aviation, and to screen specific parameters while fixing boundary conditions. Over the past decade, the investigation of bacterial growth, gene expression or population dynamics in artificially structured micro-environments has become increasingly popular [171], as they allow researchers to follow the dynamics of individual cells within a population over time, and also to precisely control their spatial, temporal, and chemical boundary conditions. In this context, the combination of fluorescence microscopy methods and microfluidic techniques has been shown to be particularly versatile [172]. Micro-fabricated bacterial traps have been applied to monitor gene expression dynamics over extended periods of time [173], and study, for example, stochastic effects in gene expression [174] [112] or bacterial oscillators [175] [176].

Compartmentalized bacterial consortia are of particular interest for applications in synthetic biology. For instance, compartmentalization could be used to create systems in which several bacterial species interact and cooperate, although they might be incompatible in co-culture (due to different growth conditions or predation of one species on the other). Furthermore, hybrid systems could be created in which some droplets contain cell-free gene expression systems, while others contain bacteria, or simply nutrients or other chemicals.

3.1.2 The Lux-system

Bacterial cells communicate with each other via the secretion of signaling molecules, direct exchange of genetic material. Here we focus on communication via signaling molecules. When detected, the signaling molecules activate or change the expression of specific target genes in the receiving cells. Due to the fact that the signaling molecules often induce their own

production, they are also referred to as *autoinducers* [177]. As a result of the positive feedback mechanism, an increase in the density of cells within a certain volume can lead to a rapid increase in the concentration of the autoinducers. This allows bacterial communities to perform *quorum sensing*, where certain genes are over-expressed once the population density surpasses a threshold value. For example, one of the first examples of quorum sensing was discovered in *Aliivibrio fischeri* [178]. The bacterium has a symbiotic relationship with a range of marine fishes and squids. The bacteria populate specific places in the large organism where they grow to high densities and produce light through quorum sensing induced expression of bio-luminescent proteins. In response, the the marine organism provides the bacteria with nutrients. Outside of the marine organism in the seawater, nutrients are scarce. In this ouside environment, the quorum sensing molecule concentration remains low due to dilution and low growth, which means that the cells do not spend resources on the expression of bio-luminescent proteins [179]. Besides bioluminescence, quorum sensing can coordinate the formation of biofilms [180], the production of antibiotics [181] or virulence factors [182], or regulate competence [183].

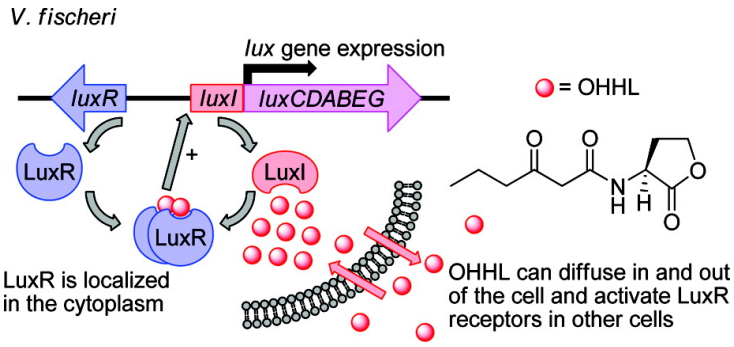

Figure 3.1: Illustration of the quorum sensing system in *A. fischeri* with the transcripational activator LuxR, the AHL-synthase LuxI and the quorum sensing molecule AHL (here abbreviated with OHHL). Reprinted with permission from Warren *et al.* [184]. Copyright (2011) American Chemical Society.

Generally, autoinducers are species specific. For example, the autoinducers for most gram-negative bacteria are acyl-homoserine lactones (AHL). AHL-based systems are relatively simple and consist of two main components: an AHL-synthase and a transcriptional activator protein with the corresponding promoter (Fig. 3.1). In the context of synthetic biology, Weiss and Knight, separeted the two components of the quorum sensing system from *A. fischeri* and created an AHL-sending and a AHL-sensing *E. coli* strain [51]. The AHL-synthase from *A. fischeri* is called LuxI and produces

3-oxo-C6-AHL for communication, which naturally does not occur in *E. coli* cells [184]. The AHL molecule binds to the transcriptional activator protein, LuxR, and promotes the formation of homodimers of the protein. The transcriptional activation from the corresponding promoter, also called *pLux*, is induced by the LuxR dimer binding to the *lux* box, which is a 20 nucleotide sequence residing about 40 nucleotides upstream of the transcriptional starting point. The bound LuxR then stabilizes the RNAP complex by binding to the σ^{70} factor and the α-subunit carboxy-terminal domain (αCTD) of RNAP [185].

3.2 The AHL sender-receiver plasmids

In previous work, Weitz *et al.* created an "AHL sender" and an "AHL receiver" plasmid based on the Lux-system and the transcriptional repressor protein LacI [186] (Fig. 3.2).The "AHL sender" plasmid contains *lacI* and a T7 promoter with a *lacO* site upstream of *luxI*. The "AHL receiver" plasmid has LuxR under the control of an endogenous *E. coli* RNAP promoter with a *lac0* site for *luxR*. In addition, GFP is under the control of a *pLux* promoter that is active when an AHL-LuxR complex has bound (Fig. 3.2). Hence the LuxI-LuxR system is activated by IPTG and expresses GFP as a read-out.

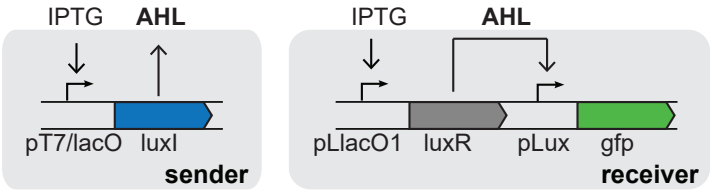

Figure 3.2: Two plasmid sender-receiver system. The sender plasmid encodes the AHL-Synthase LuxI, which is under the control of an T7 RNAP promoter that is inducible with IPTG. The genetic AHL receiver module responds to the inducers IPTG and AHL: IPTG induces expression of the transcriptional activator LuxR, which is under the control of a lac promoter. Only in the presence of AHL, LuxR activates the expression of GFP, which is controlled by a *lux* promoter.

3.3 Water-in-oil droplets

It was previously found that a fluorinated emulsion droplet system that is used to encapsulate bacteria can support communication via indcuer molecules IPTG and AHL, as both of them are able to partition in to the continuous oil

phase and diffuse between different water-in-oil droplets [186]. The water-in-oil droplets are stabilized against coalescence by a surface active agent ('surfactant'), which is in this case a triblock-copolymer that assembles at the droplets surface. It consists of a poly-ethylene glycol (PEG) head group that faces towards the droplets inner aqueous phase. The head-group is flanked by 10 to 60 perfluoropolyether (PFPE) groups that form the two tail groups, which assemble on the outside towards the oil phase [187]. The continuous oil phase is made from a fluorinated carbon oil, such as FC-40 (see Sec. 3.12). It was then shown that bacteria with the AHL receiver plasmid that were induced with IPTG and encapsulated in droplets were able to express GFP in the vicinity of droplets containing either bacteria with the AHL sending plasmid or directly purified AHL [186].

3.4 Outline

Here we studied spatially distributed gene expression in strictly linear arrangements of micro-droplet compartments. A quasi one-dimensional geometry allows for a better control of boundary conditions, and also facilitates a straightforward analysis of the experiments. One of the most important aspects for our present study is the stronger mutual coupling of neighboring compartments in a 1D geometry.

3.5 TXTL: AHL receiver plasmid

We transferred the AHL sender-receiver plasmids in an *E. coli* based cell lysate which is supplemented with an energy solution so that transcription (TX) and translation (TL) still work [68]. The cell-free expression system is thus abbreviated TXTL.

3.5.1 Hill curve

The chromosomal DNA of the bacteria is degraded during the preperation of the TXTL so that the TXTL solution can be programed by the addition of external DNA. Hence the addition of the AHL receiver-plasmid to the TXTL creates a solution that expresses GFP as a function of the concentration of AHL. In plate reader experiments, the maximal rate of GFP expression increases with the concentration of AHL and follows roughly a Hill curve

$$r(C) = \alpha \frac{C^n}{K^n + C^n} + \alpha_0 \tag{3.1}$$

with a coefficient of $n = 1.5$ and a threshold value of $K \approx 12.8$ nM (Fig. 3.3A), where C is the concentration of the inducer (in this case AHL), α the theoretical maximal expression rate at full induction, α_0 is the basal expression rate and r is observed expression rate of GFP as a function of the inducer. Then, ideally the normalized maximal expression rate $\frac{r-\alpha_0}{\alpha}$ is proportional to the fraction of active to inactive promoters.

3.5.2 Comparison to bacteria

A similar response to AHL is registered for the receiver plasmid in *E. coli* (Fig. 3.3B[1]) [186]. Here the fit with a Hill function results in a Hill coefficient of $n = 2.6$ and a threshold value of $K \approx 1.2$ nM. However, Weitz *et al.* also measured the response function of a plasmid without the LacI control of LuxR but otherwise identical and they found a Hill coefficient of $n = 1.6$ and a threshold value of $K = 15$ nM, which are closer to the values we measured in TXTL [186]. This is a remarkably sensitive system as a threshold value around $K \approx 10$ nM for an *E. coli* corresponds to about 8 molecules per cell [188][2]. Most importantly, the AHL receiver plasmids maintains that sensitivity in the TXTL as the threshold value is in a similar range as for bacteria. It seems from Fig. 3.3 that the response of the receiver plasmid in TXTL does not follow the hill curve as smoothly as for bacteria ($\frac{RMSE_{TXTL}}{RMSE_{Bac}} \approx 13$)[3]. The reasons for this have not been systematically investigated in the scope of this thesis.

However, there are differences in the experimental conditions between the samples which make a direct comparison difficult. For example, the sample for bacteria in the plate reader has a volume of $V = 300\ \mu l$ at $T = 37°C$ instead of $V = 15\ \mu l$ at $T = 29°C$ for TXTL. The overall protein concentration in the TXTL is about 30x more diluted than in bacteria [189]. Furthermore, during sample preparation, the components for the TXTL reactions such as DNA and the cell lysate are under the direct influence of pipetting errors between samples whereas bacteria in the exponential growth phase have a control system for the plasmid copy number and the concentration of the components of the expression machinery that depend on their growth rate [190] [191].

[1]The data with bacteria was measured by Ronja Berg in her Bachelor thesis under the supervision of Maximilian Weitz and Friedrich C. Simmel.

[2]molecules per cell = $(12.8\text{ nM}) \cdot (1\ \mu\text{m}^3) \cdot (N_A) = 8$.

[3]The root mean squared error (RMSE) is defined as $\frac{\sqrt{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}}{T}$, with T being the number of data points and y_t and $\hat{y}_t$ being the data point and the predicted data point at t.

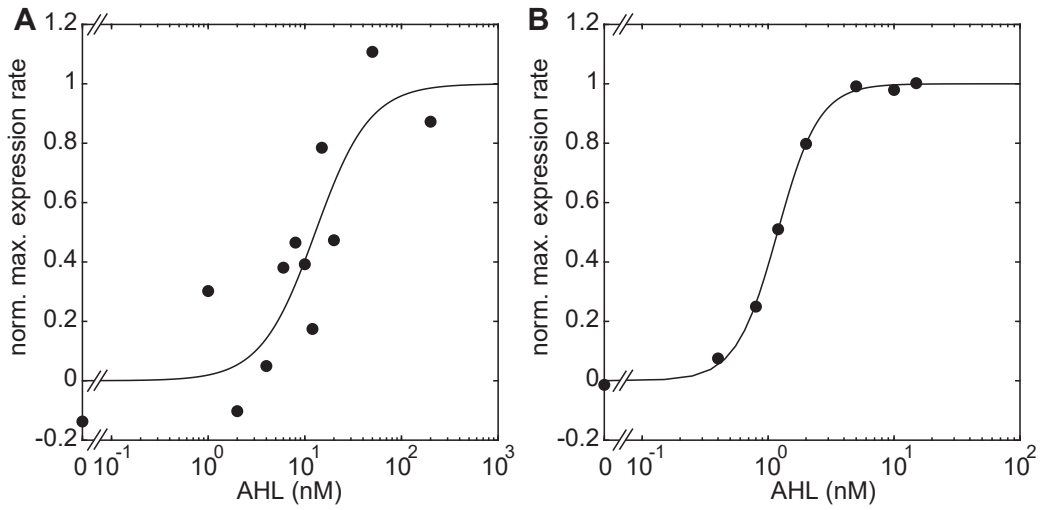

Figure 3.3: **A** Normalized maximum GFP expression rate of the AHL receiver plasmid (7.5 nM) in TXTL for varying concentration of AHL and $c(IPTG) = 10$ mM, measured in a plate reader. A fit with a Hill curve (continuous line) results in a Hill exponent of $n = 1.5$ and an induction threshold at $K = 12.8$ nM. **B** Normalized maximum GFP expression rate of the AHL receiver plasmid in *E. coli* in the exponential growth phase (within 5 hours) at $c(IPTG) = 2$ mM for different AHL concentrations. A fit with a Hill curve (continuous line) results in a Hill exponent of $n = 2.6$ and an induction threshold at $K = 1.2$ nM. Bacteria data in (B) replotted from [186].

3.5.3 Related studies

It has previouly been shown that quroum sensing systems from bacteria can function in cell lysates [192] [193]. The focus in these studies is to show that cell-free assays are rapid, robust and cost-effective for the detection of autoinducers. Using purified LuxR and DNA on gel shift assays, Urbanowski *et al.* determined that the binding of AHL to LuxR is non-cooperative with a hill coefficient of $n = 0.9$. However, the AHL concentration at which 50% of the LuxR was unbound, which corresponds to the threshold value, was determined to be about $K = 100$ nM [194].

In another study, where the response of a very similar plasmid was carefully studied in *E. coli* the Hill coefficient and threshold constant found are $n = 1.6$ and $K = 1.5$ nM, confirming the values found by Weitz *et al.* [195] [186].

In 2018, Halleran and Murray compared the cross-talk of several quorum sensing system in bacteria and TXTL that was made with the same protocol as in our experiments [196]. They found that the cell-free reactions accurately

predicted cross-talk of the quorum sensing system, however it underestimated the intensity. The difference in the ratio of transcription factor proteins to free promoters between the TXTL and bacteria was identified as one reason for the discrepancy in the intensity of crosstalk.

3.5.4 LacI control of the receiver plasmid

LacI is present in the *E. coli* stain from which the TXTL is made, so that there should be endogenous protein in the reaction. However, the additional IPTG control of the GFP expression only worked in the TXTL after supplementation of purified LacI proteins (about 200 nM) at a DNA concentration of 7.5 nM (Fig. 3.4A). Hence the response of the maximal production rate of GFP as a function of IPTG did not follow a Hill curve (Fig. 3.4B) for the TXTL that was not supplemented with additional LacI. However, also the response of the plasmid to IPTG titration in bacteria (with 15 nM AHL) did not follow a smooth hill curve (Fig. 3.4C[4])

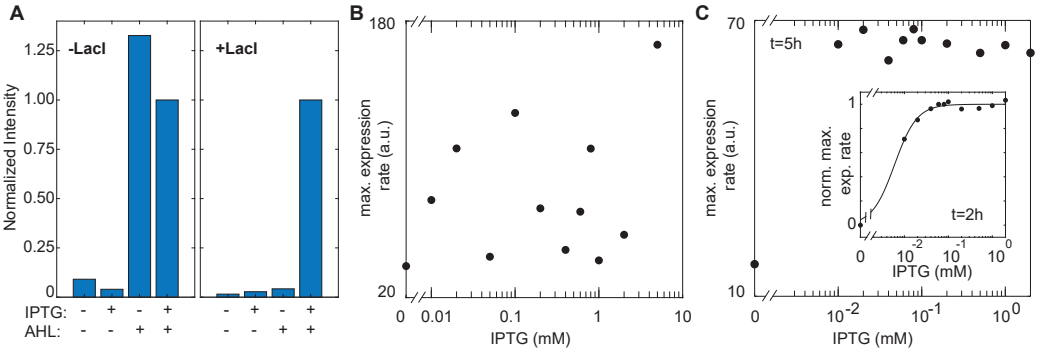

Figure 3.4: **A** GFP fluorescene intensity of TXTL reactions with the AHL receiver plasmid (7.5 nM) at t=5h. IPTG concentration $c = 5$ mM, AHL concentration $c = 200$ nM. Left: without the addition of purified LacI. Right: with the addition of about 200 nM His-tagged purified LacI. **B** Maximum GFP expression rate of the AHL receiver plasmid (7.5 nM) in TXTL for varying concentration of IPTG at $c(AHL) = 200$ nM, measured in a plate reader, without purified LacI. **C** Maximum GFP expression rate of the AHL receiver plasmid in *E. coli* for varying concentration of IPTG at $c(AHL) = 15$ nM within 5 hours, measured in a plate reader. Inset: GFP expression rate at $t = 2$ hours. Bacteria data in (C) replotted from [186]

For this reasons we focused on the characterization of the AHL receiver plasmid in respect to AHL, while saturating the reactions with IPTG.

[4]The data was measured by Ronja Berg in her Bachelor thesis under the supervision of Maximilian Weitz and Friedrich C. Simmel.

3.6 Encapsulated TXTL or bacteria with an AHL receiver plasmid in a spatial AHL concentration gradient

3.6.1 Experimental set-up

To introduce a spatial dimension to the gene expression, either the bacteria or the TXTL with the receiver plasmid were encapsulated in picoliter-sized emulsion droplets, which corresponds to a droplet diameter of about $40 - 50$ μm, using a microfluidic droplet generation system. The droplets were then loaded into a squared glass capillary with inner lengths of 50 μm so that the diffusion between droplets is effectively reduced to two neighboring droplets. To study the diffusion mechanism of the inducers between the droplets in the capillary, we introduced a spatial inducer concentration gradient at the opening of the capillary. This was realized by placing the capillary in a reservoir with IPTG and AHL. As the aqueous phase in the droplets was not saturated with AHL, AHL would start to diffuse into the capillary at its open ends once placed in the reservoir. The logarithm of the partitioning constant for octanol/water for OC6-HSL is $\log K \approx 1 - 2$ which means that the concentration of AHL will be 10-100x higher in the continuous phase for octanol [197] [198]. As we use fluorinated emulsion the situation might be quite different. However, Gruner *et al.* showed that there is molecular transport between droplets in fluorinated emulsions through surfactant assemblies such as micelles [199].

3.6.2 Considerations: inducer gradients

The concentration of AHL along the long axis of the capillary can be approximated by solving the 1D diffusion equation

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2} \tag{3.2}$$

where $C(x,t)$ is the AHL concentration, D its diffusion coefficient. The inducer concentration at the opening of the capillary $x = 0$ was assumed to remain constant ($C = C_0$) since the volume of the reservoir is about 1000 times larger than the interior volume of the capillary. Therefore any decrease in AHL concentration due to molecules leaving the reservoir and entering the capillary is quickly buffered by the rest of the reservoir.

Since the length of the capillary is $l \geq 1$ cm and the run time of the encapsulated TXTL is about $t \approx 5 - 6$ hours, we assumed that the inducer

concentration is sufficiently small at the center of the capillary so that we can neglect it [200]. These assumption allows us to solve Eq. 3.2 using the boundary conditions of an semi-infinite medium

$$C = C_0 \text{ for } x = 0 \text{ and } t > 0 \text{ and } C = 0 \text{ for} x \to \infty \tag{3.3}$$

and intial conditions

$$C = 0 \text{ for } x > 0 \text{ and } t = 0. \tag{3.4}$$

The solution can be found with the help of Laplace transformations [201]:

$$C(x,t) = C_0 \operatorname{erfc}\left(\frac{x}{\sqrt{4Dt}}\right). \tag{3.5}$$

where $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ is the complementary error function. With time, the point $x_K(t)$ at which the inducer concentration has reached the threshold value for GFP expression ($C(x = x_K, t) = K$) is given by

$$x_K(t) = \sqrt{4Dt}\operatorname{erfc}^{-1}\left(\frac{K}{C_0}\right), \tag{3.6}$$

where erfc^{-1} is the inverse complementary error function. We tested the solutions of Eq. 3.5 and 3.6 for our set-up using a rough approximation for the diffusion constant of AHL in the capillary, based on the reported value for aqueous solutions ($D \approx 100 - 1000\ \frac{\mu m^2}{s}$ [202] [203]). In addition we took the effect of higher viscosity of the TXTL and the fluorinated oil ($\approx$ factor 4 compared to water) into account which is given by the Einstein-Smoluchowski relation. Hence the solutions for $D = 25 - 250\ \frac{\mu m^2}{s}$ are plotted in Fig. 3.5 and show that the concentration at the center of the capillary ($x = 5$ mm) would only stay below the threshold concentration if the diffusion coefficient is at the lower bound.

It is important to note here that D is an apparent diffusion coefficient, which simplifies the description of the diffusion process. The next section takes a more detailed view on the diffusion process in a medium that has more than one phase.

Permeability

For the transport mechanism of the inducer between emulsion droplets, one can use descriptions that where developed for lipid membranes such as the

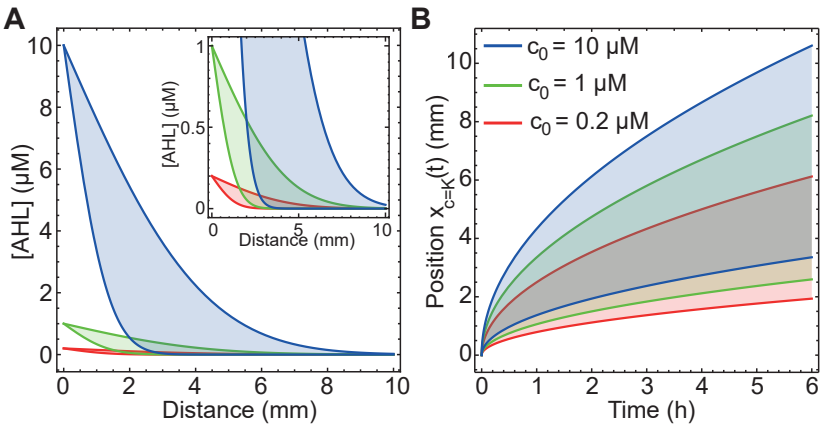

Figure 3.5: **A** The solution of Eq. 3.5 for three different AHL reservoir concentrations $C_0 = 200$ nM (red), $C_0 = 1$ μM (green) and $C_0 = 10$ μM (blue) for diffusion coefficients $D = 25 - 250$ $\frac{\mu m^2}{s}$. Inset: Zoom-in on the concentration. **B** The solution of Eq. 3.6 for the three AHL resevoir concentrations for the range of diffusion coefficients. It can be seen that x_K reaches the center of the capillary (≈ 5 mm) within the life time of the TXTL reaction (≈ 5 h) for diffusion coefficients at the upper bound.

Meyer-Overton Rule [204]. It states that the trans-membrane flux of a permeating molecule can be described in terms of the partitioning coefficient:

$$K_{oil} = \frac{C_{oil}}{C_{aq}} \tag{3.7}$$

where C_{aq} is the aqueous phase, and C_{oil} the continous phase. The surfactant mostly increases the solubility within the continuous phase, increasing the value of partitioning coefficient for the emulsion system. Since it was found that an interface between the continuous and dispersed phase is not a significant energy barrier for the molecules passing [205], Fick's first law for the flux between two neighboring droplets (see Fig. 3.6B) can be written as:

$$J = -D_{oil}\frac{\Delta C_{oil}}{\Delta x} = -D_{oil}\frac{(C_{oil}(1) - C_{oil}(2))}{\delta} \tag{3.8}$$

$$= -D_{oil}K_{oil}\frac{(C_{aq}(1) - C_{aq}(2))}{\delta} = -P_{oil}(C_{aq}(1) - C_{aq}(2)) \tag{3.9}$$

$$P_{oil} := \frac{KD_{oil}}{\delta} \tag{3.10}$$

where P is called the permeability, D_{oil} is the diffusivity within the continuous phase and C_{oil} and C_{aq} are the concentrations at the interfaces in

the continuous phase and aqueous phase at Position 1 and 2 (see Fig. 3.6B). Using the Derjaguin approximation, the droplets can be approximated by cylinders with radius d and the continuous phase with a radius δ (see Fig. 3.6A) [206]. Hence δ is the width of the oil gap between the droplets. As previously mentioned, molecular transport in the continuous phase can be mediated by surfactant molecules and assemblies such as reverse micelles and vesicles [199]. This means that there are several ways of the inducer to diffuse through the continuous phase between two droplets. This is equivalent to the description of several membranes in parallel each with permeability P_i. The total permeabilities then is given by:

$$P = \sum_i P_i. \tag{3.11}$$

When considering several droplets in a linear array, the oil gaps between the droplets can be thought of as several membranes with P_i in series. Then, the overall permeability P is given by:

$$\frac{1}{P} = \sum_i \frac{1}{P_i}. \tag{3.12}$$

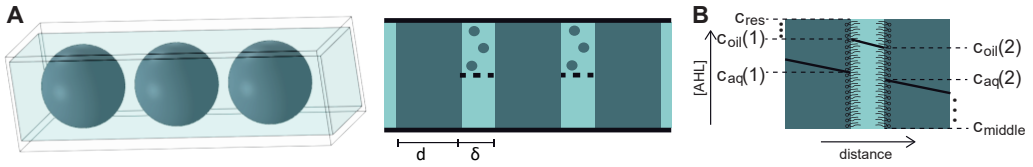

Figure 3.6: **A** Sketch of the emulsion droplets in the glass capillary and an abstraction of the linear array of droplets using the Derjaguin approximation. The approximation reduces the droplet geometry to a quasi 1D array of rectangles of width d that are separated by thin oil films of thickness δ. The oil film is separated by a dashed line to illustrate parallel diffusion paths through the oil phase. The upper half shows transport between droplets via inducer surfactant assemblies. **C** An illustration of an AHL concentration profile accors two water reservoirs separated by an thin oil film that is stabilized by surfactant molecules with an partition coefficient $K > 1$. c_{res}, $c_{aq}(1)$, $c_{aq}(2)$, $c_{oil}(1)$, $c_{oil}(2)$ and c_{middle} are the AHL concentration at different positions.

Hence, the apparent diffusion coefficient D for inducers in the linear arrangement of emulsion droplets can be approximated by [207]:

$$\frac{1}{P} = \frac{1}{P_{aq}} + \frac{1}{P_{oil}} \tag{3.13}$$

$$\frac{d+\delta}{D} = \frac{d}{D_{aq}} + \frac{1}{P_{oil}} \tag{3.14}$$

$$P_{oil} = \frac{K_{oil} D_{oil}}{\delta} \tag{3.15}$$

where P_{aq}, P_{oil}, P are the permeabilities and D_{aq}, D_{oil} and D are the diffusion constants for the aqueous phase, the oil phase and the whole system and K_{oil} as defined by Eq. 3.7. If we assume that $d >> \delta$ and that $D_{aq} >> P_{oil}d$ we can simplify Eq. 3.14 to:

$$\frac{d+\delta}{D} \approx \frac{d}{D} = \frac{d}{D_{aq}} + \frac{1}{P_{oil}} \tag{3.16}$$

$$\frac{P_{oil}d}{D} = \frac{P_{oil}d}{D_{aq}} + 1 \approx 1 \tag{3.17}$$

$$D = P_{oil}d \tag{3.18}$$

This results underlines the importance of the permeability on the apparent diffusion coefficient. For the case that the inducer molecules can diffuse through the continuous phase in different ways such as in surfactant assemblies of different sizes or as a small molecules without surfactant molecules the permeability is given by:

$$P_{oil} = \sum_i P_i = \sum_i \frac{K_i D_i}{\delta} \tag{3.19}$$

with P_i, D_i and K_i are the permeabilities, diffusion and the partition coefficients for the different phases in the continuous phase. As different partition coefficients can depend on each other a clear description of each permeability can be experimentally challenging.

3.6.3 Reaction-diffusion system

As stated previously, we infer the diffusion dynamics of the inducer from the kinetics of the protein expression. Thus, in an attempt to increase the accuracy, we can include the protein expression, protein maturation and the inducers life time in our model. This step adds a term to Eq. 3.2 and

introduces two new equations:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - \phi C(x,t) \tag{3.20}$$

$$\frac{\partial G}{\partial t} = \alpha_{exp}(t)\frac{C^n}{K^n + C^n} - \alpha_{mat} G \tag{3.21}$$

$$\frac{\partial G_{mat}}{\partial t} = \alpha_{mat} G \tag{3.22}$$

where C, G and G_{mat} are the concentrations of the inducer, nascent GFP and maturated GFP, ϕ is the degradation rate of the inducer and α_{mat} is the GFP maturation rate. Eq. 3.21 describes the protein expression with a Hill curve activation for the inducer like in Eq. 3.1 and a degradation term. For the TXTL, as shown by Karzbrun *et al.*, the the GFP expression rate is assumed to exponentially apporach the expression rate of full induction α_{max} so that $\alpha_{exp}(t) = \alpha_{max}(1 - \exp[-(t - t_1)/t_2]$, with the delay time t_1 in protein expression at the beginning and the lifetime of mRNA t_2 [189]. Finally in Eq. 3.22 the protein maturation is taken into account, since only maturated GFP is detected by fluorescence.

3.6.4 TXTL droplets in a spatial AHL concentration gradient

The receiver plasmid in TXTL supplemented with IPTG was encapsulated and the emulsion droplets loaded into a capillary. The capillary was then placed in a reservoir with M9 medium, $c(IPTG) = 10\ \mu$M and either $c(AHL) =$ 200 nM, 1 μM or 10 μM. It is possible to track the gene expression within droplets in a time series of epifluorescence microscopy images (Fig. 3.7). In order to make sure that the gene expression is only affected by the AHL inducer gradient from one end of the capillary we restricted our analysis to the first 1 mm of the capillary from one end.

Due to the establishment of an AHL gradient inside the capillary, gene expression is activated faster and stronger in the droplets closer to the reservoir. As the TXTL system has a certain life-time of about $\tau_{TXTL} \approx 5 - 6$ hours after which it stops expressing protein [200] and the GFP is not degraded, we see a stable spatial gradient in GFP concentration (Fig. 3.8A). Droplets at a certain distance from the reservoir were not able to express any detectable amount of GFP above background levels, because the sufficient AHL concentration at these droplets was not reached within the time window of τ_{TXTL}.

Similarly, by following the dynamics of the GFP expression of each droplet, one can see that the maximal rate of GFP expression also decreases with the

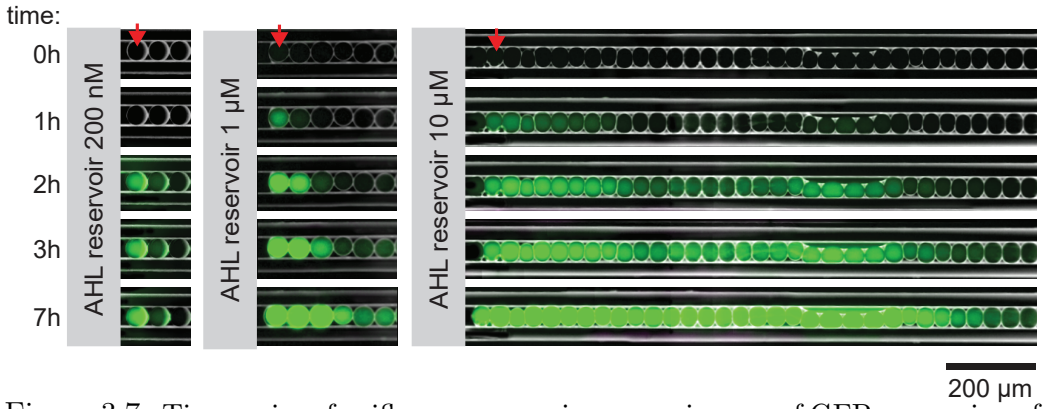

Figure 3.7: Time series of epifluorescence microscopy images of GFP expression of TXTL droplets containing the AHL receiver plasmid. The left end of the capillary is connected to an AHL reservoir of different concentration: 200 nM (left), 1 μM (center), 10 μM (right). As AHL diffuses into the capillary it activates the GFP expression within the droplets. Time=0h corresponds to the beginning of the acquisition.

distance from the reservoir (Fig. 3.8B). Although larger distances from the reservoir will result in slower kinetics due to lower AHL concentrations, this is not a one-shot AHL titration experiment which can be predicted by the AHL dependence of the gene expression rate established in Fig. 3.3. The dynamics of GFP expression are overlaid by the depreciating performance of the TXTL, which has not been explicitly measured as a function of time and by the AHL concentration flux. One simple way to change the AHL concentration flux is to change the concentration of the reservoir. With increasing AHL concentration in the reservoir, the GFP expression in more droplets was activated, for the cost of having less spatial differentiation in GFP concentration between neighboring activated droplets. Interestingly, there is a discrepancy between the expectations of the number of droplets (n) with GFP above the background from back-of-the envelope-calculations (see Sec. 3.6.2) and the results for the AHL reservoir concentrations of $C_{res} = 200$nM ($n \approx 3$) and $C_{res} = 1$ μM ($n \approx 5$). In the case of the highest amount of concentration flux ($C_{res} = 10$ μM), is there a clear relationship between the time of occurence of the max. expression rate and the distance of the droplet from the reservoir (Fig. 3.8C).

3.6.5 Apparent diffusion coefficient

In order to gain a more quantitative understanding of the transport mechanism we fitted two hours of the expression dynamics with the set of reaction-

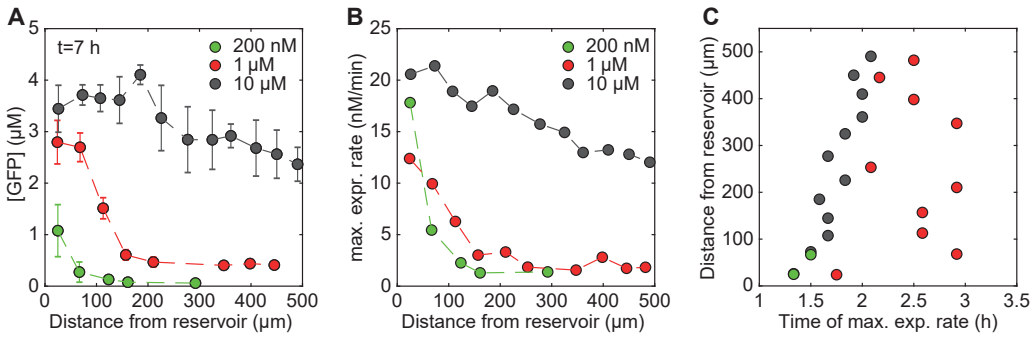

Figure 3.8: **A** The GFP concentration of droplets for different distances from the AHL reservoir for three different concentrations after $t = 7$h of acquisition. The mean concentration with s.e.m. of 3 samples for res. conc. of 200 nM and 10 μM and of 2 samples for 1 μM of AHL. **B** Average maximal expression rate of GFP per droplet for different distances from the three different AHL reservoirs. S.e.m. is not shown, due the large size of the error (avg. $CV = 0.15$ for $C_{res} = 200$nM, avg. $CV = 0.11$ for $C_{res} = 1$ μM and avg. $CV = 0.21$ for $C_{res} = 10$ μM). **C** The time of occurrence of the avg. maximal expression rate per droplet shown against the distance from the reservoir. The legend is $C_{res} = 200$nM (green), $C_{res} = 1$ μM (red) and $C_{res} = 10$ μM (black).

diffusion equations 3.20-3.22 (Fig. 3.9). In the experiments with an AHL reservoir concentration of $c = 1$ and $c = 10\,\mu$M we considered the dead time of about 30 minutes before the acquisition due to experimental setting-up. The equations were solved numerically and fitted with MATLAB using the *pdepe* solver and the *fminsearch* function which was used to minimize the least squared error between the solution and the experimental data. The only variable parameters were the apparant diffusion constant D and the maximal expression rate α_{max}. The best solution for D and α_{max} and the values of the remaining constants are listed in Tab. 3.1.

The best solution for both D and α_{max} vary between the different AHL reservoir concentrations. The different values for α_{max} most probably come from batch-to-batch variations in the overall performance of the TXTL, as variations in α_{max} mostly scale the amount of GFP produced in all droplets. However, only by changing D is it possible to account for the time delay of the GFP expression in different droplets as a function of the distance from the reservoir.

By considering the GFP expression of the first two hours of acquisition, the value for the diffusion coefficients were dependent on the AHL reservoir concentration (Table 3.1). Whereas the apparent diffusion constant had a value of around $D \approx 0.1\ \mu m^2 s^{-1}$ for reservoirs with $c_{AHL} = 200\ nM$ and $c_{AHL} = 1\ \mu M$ ('low' reservoir concentration), it increased by factor of about

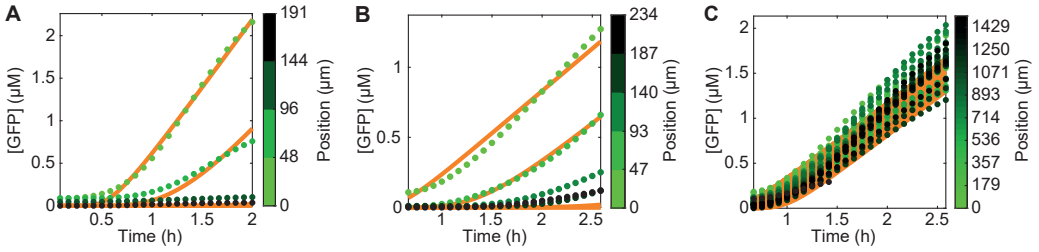

Figure 3.9: The GFP expression kinetics for the first two hours of acquisition for the three AHL concentrations of the reservoir **A**: 200 nM, **B**: 1 μM and **C**: 10 μM are fitted with the reaction diffusion equations in order to estimate the apparent diffusion coefficient of AHL.

$c_{AHL}(x=0)\ (\mu M)$	$D_a\ (\mu m^2 s^{-1})$	$\alpha_{max}(pMs^{-1})$
0.2	0.15	458
1	0.09	155
10	29.89	224

Table 3.1: Apparent diffusion constants

300 to $D \approx 30\ \mu m^2 s^{-1}$ for the reservoir with $c_{AHL} = 10\ \mu M$ ('high' reservoir concentration).

3.6.6 Depreciation of TXTL

There are several potential reasons for the concentration dependence of the apparent diffusion coefficient. As previously mentioned, the performance of the TXTL system depreciates with time. For example, the pH of the TXTL changes from about 7.1 to about 6.2 in 12 hours (Fig. 3.10). Encapsulation using the surfactant and oil, might even speed up the depletion of resources since it has been shown that substantial amounts of ions that are key to transcription and translation such as Mg^{2+} where > 90 % are transferred to the continuous phase via surfactant molecules within the time frame of about an hour [208].

Smaller diffusive flux due to lower AHL reservoir concentrations means that droplets at the same position are induced at later times. Hence the slower GFP expression kinetics for the second and consecutive droplets from the capillary opening for the reservoir concentration of 200 nM and 1 μM might also be related to the fact the TXTL has depreciated.

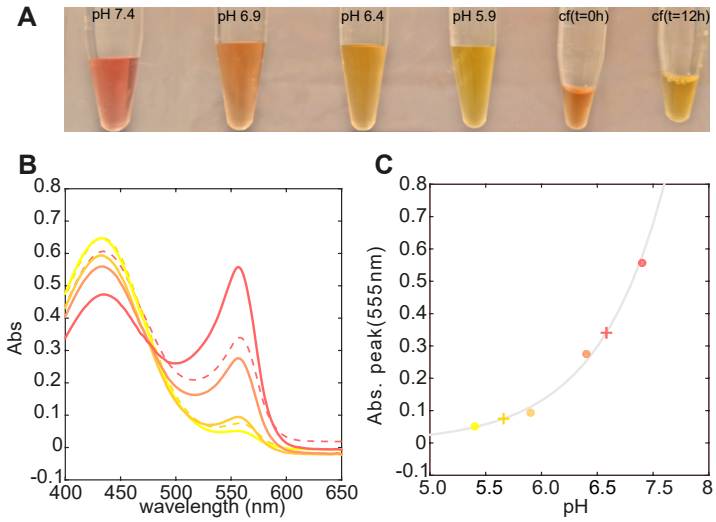

Figure 3.10: pH Test with phenol red and the cell free system: **A** Phenol red calibration solutions (100 μM in PBS) between pH 7.4 and pH 5.9 and cell free reactions supplemented with 100 μM phenol red. **B** Absorption spectra of calibration solutions and the cell free samples (dashed lines) at t=0h and t=12h after start of experiment. **C** Absorption maxima at 555 nm plotted against the corresponding pH values. Data points were fitted with an exponential curve for estimating pH values of the cell free samples.

3.6.7 Surfactant micelle formation

As described in Sec.3.6.2 the transport mechanism within the capillary can be described by a series of water reservoirs separated by thin oil films with permeability P. So the AHL partitions into the continuous phase, diffuses to the droplet interface, partitions into the aquous phase, diffueses to the other side of the droplet, diffuses through the surfcatant layer and partitions into the continuous phase again. The surfactant used in this experiments is not expected to form adhesive bilayers and it was found that the droplet boundary is not thought to be rate limiting.

Studies on the transport mechanisms between droplets using fluorophores that are less soluble in the continous phase showed that the rate limiting step is the diffusion through the continuous phase [208]. Since the surfactant interface is not a barrier, the key parameters for the permeability are the spacing between the droplets and the partitioning constant. The partitioning constant and thus the permeability can be increased with increasing surfactant concentration [199]. It has been shown that the transport of molecules between the aqueous phase through the continuous phase can be mediated by assemblies of surfactant molecules such as vesicles, micelles or reverse mi-

celles with a diameter of about $100 - 1000$ nM [199]. It could also be the case that AHL itself has a low solubility in fluorinated oil.

The different ways to diffuse through the continuous phase might also explain the discrepancy in the diffusion coefficient. For AHL reservoir concentrations of $c_{AHL} = 200$ nM and $c_{AHL} = 1$ μM, the concentrations of AHL molecules that diffuse over a faster pathway are below the AHL induction threshold (see Fig. 3.3), so that only the slower diffusing surfactant assemblies push the concentration above the threshold. In the case of the higher AHL reservoir concentration of $c_{AHL} = 10$ μM, the amount of AHL molecules diffusing without surfactant molecules could be high enough to induce the GFP expression, so that the apparent diffusion coefficient reported by the GFP expression is dominated by this diffusion pathway.

3.6.8 Binding to LuxR

Additionally, the depletion of AHL due to binding of LuxR in upstream droplets could potentially explain the dependency of the apparent diffusion coefficient on the reservoir concentration. However, simulation with LuxR did not show a significant difference (Fig. 3.11). An AHL binding term was added to the reaction-diffusion equation using the model by Basu *et al.* [209]. In addition the LuxR expression was simulated with the parameters determined by Karzbrun *et al.* [189].

$$\frac{\partial A}{\partial t} = D\frac{\partial^2 A}{\partial x^2} - \phi A - k_p \cdot L^2 \cdot A^2 \tag{3.23}$$

$$\frac{d(A:L)}{dt} = k_p \cdot L^2 \cdot A^2 \tag{3.24}$$

with $A = [AHL]$, $L = [LuxR]$ and $A:L = [AHL:LuxR]$.

3.6.9 1D geometry

Given that the apparent diffusion coefficient is relatively small for the low reservoir concentrations, underlines the importance of restricting the diffusion along one dimension in order to see an effect in more than one droplet. In the linear arrangement, we have only two neighboring droplets, whereas in a 2D and 3D arrangement of closely packed droplets there are 6 and and twelve neighboring droplets, decreasing the coupling between droplets due to dilution of inducer molecules.

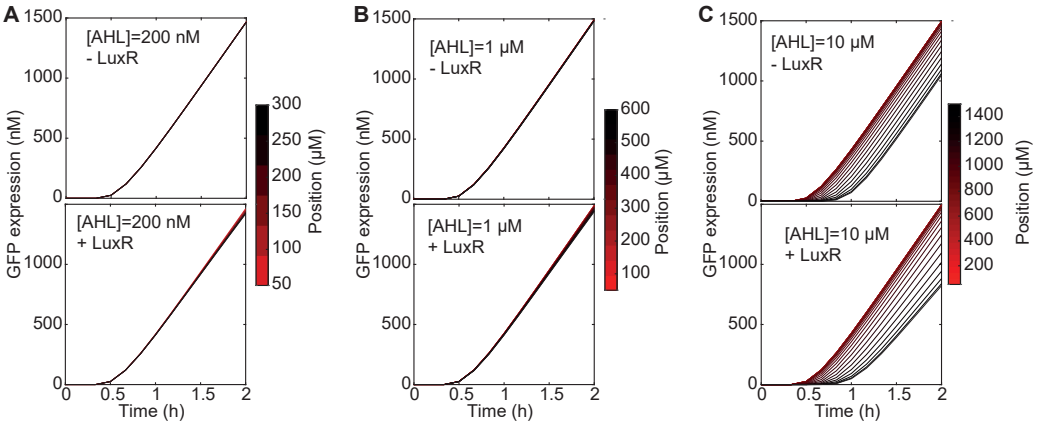

Figure 3.11: Simulation of GFP expression without (upper) and with (lower) LuxR expression for an AHL reservoir concentration of **A**: 200 nM, **B**: 1 μM and **C**: 10 μM. The GFP expression are slightly delayed in the presence of LuxR, however this effect cannot alone explain the large difference between the apparent diffusion coefficients between the low and high resevoir concentrations.

3.7 TXTL droplets vs. bacteria droplets with AHL reservoir

We also encapsulated *E. coli* bacteria with the AHL receiver plasmid in the emulsion droplets to see their response within an AHL concentration gradient using the experimental set-up explained in the previous section. The bacteria were grown in M9 minimal medium supplemented with IPTG to an optical density of $\approx 0.1-0.2$ so that there are only a few bacteria per droplet at the beginning and loaded into an capillary which was then placed in an AHL reservoir (Fig. 3.12A).

Constant	Value	Reference
ϕ_{AHL}	$1 \cdot 10^{-5}\ s^{-1}$	[210]
α_{mat}	$3.33 \cdot 10^{-3}\ s^{-1}$	[211]
t_1	20 min	[189]
t_2	12 min	[189]
n	1.5	
K	$12.8\ nM$	
k_p	$10^{-1}\ nM^{-3}\ s^{-1}$	[209]

Table 3.2: Parameter for the reaction-diffusion equations.

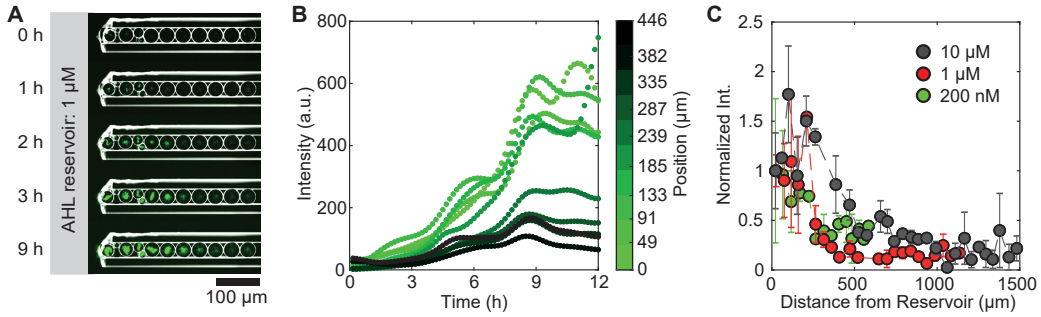

Figure 3.12: **A** Time series of epifluorescence images of emulsion droplets inside a glass capillary, containing bacteria with the AHL receiver plasmid. The capillary is placed in a reservoir with $[AHL] = 1\mu$M. As AHL diffuses into the capillary it induces GFP expression inside the bacteria that are growing in the droplets. **B** Tracked GFP fluorescence intensity of the different droplets in the capillary for $[AHL] = 1$ μM. **C** The GFP fluorescence intensity level of droplets for different distances from the AHL reservoir for three different concentrations after $t = 9$h of acquisition. The mean concentration with s.e.m. of 3 samples for res. conc. of 200 nM and of 2 samples for 1 μM and 10 μM of AHL.

In contrast to the TXTL, the bacteria are still growing within the droplets during the subsequent induction, so that the signal of the induced GFP fluorescence is overlaid by bacterial growth. This has an effect on the fluorescence time traces of the droplets (Fig. 3.12B) in that it is harder to quantify the GFP concentration per droplet, as the bacteria change their position within the droplet which can cause a change in the detected fluorescence intensity. In addition, the GFP fluorescence level is dependent on the number of bacteria in the droplets and not only on how strong they are induced. For that reason, we did not convert the fluorescence intensity into a GFP concentration as for the TXTL.

By comparing the end level (t=12h) of the bacteria for different AHL reservoir concentrations, we were able to see that in all cases the GFP signal dropped to a background level for $x > 500$ μm from the opening of the capillary (Fig. 3.12C). As oxygen is required both for bacterial growth and GFP maturation, this signal drop might not be caused by an absence of inducers molecules but rather a lack of oxygen availability within the capillary. This effect was not observed for the TXTL droplets. For these reasons, bacterial droplets are inferior to TXTL droplets for the characterization of the diffusion process within the droplets in the capillary.

3.8 IPTG concentration gradient

We looked at IPTG diffusion within the capillary, by saturating all droplets with the receiver plasmid in either TXTL or bacteria and the reservoir with $c_{AHL} = 200$ nM, but only the reservoir with $c_{IPTG} = 10$ mM. As the IPTG diffuses through the capillary the droplets start the expression of GFP as shown in Fig. 3.13.

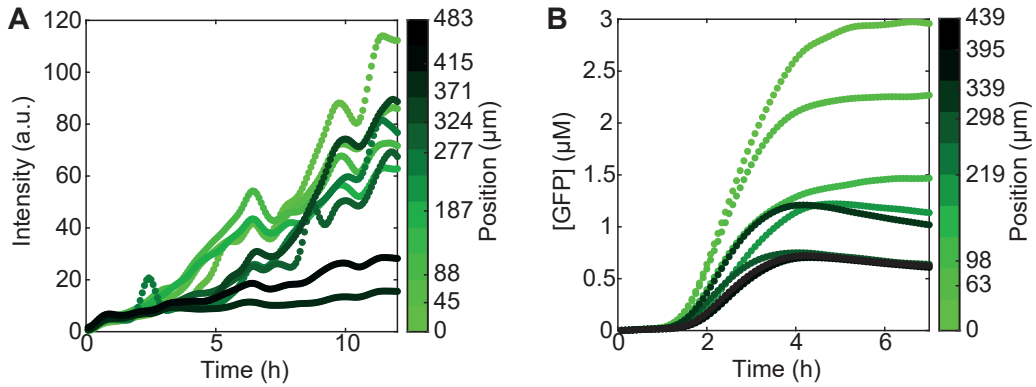

Figure 3.13: Tracked GFP fluorescence intensity of droplets with the receiver plasmid in **A** bacteria or **B** a cell-free expression system in the capillary with different distant to the reservoir $[IPTG] = 10$ mM.

The signal for bacteria is again overlaid by the growth dynamics and the variations in intensity due to the movement of the bacteria within the droplet. The droplets within $l \leq 200$ μm express GFP stronger within the first $t = 6$ h then the the rest of the droplets, whereas all the droplets up to $x \leq 400$ μm reach about the same intensity after $t = 12$ h, the droplets beyond this distance seem to express significantly less amount of GFP. For the TXTL, only the first two droplets can be differentiated from the rest, whereas for the bacteria the first four droplets have a significantly more GFP at about $t = 6$ h than the rest, making the bacteria filled droplets more sensitive to the inducer concentration gradient. However, it was surprising to find the GFP expression to be dependent on the distance at all, as the bulk experiments in Fig. 3.4A did not suggest that gene expression should be dependent on IPTG unless LacI protein is supplemented (which was not the case in the experiments here).

3.9 Reservoir inducer droplets

Inspired by previous work we encapsulated inducer in droplets (termed 'reservoir droplets') and mixed them with droplets contaning bacteria/TXTL with the receiver plasmid, which did not contain the encapsulated inducer [186]. With this we can restrict the gene expression to droplets in the vicinity of the reservoir droplets (Fig. 3.14). We either encapsulated $c_{AHL} = 200$ nM and/or $c_{IPTG} = 10$ mM, so that we can approximate the reach using the macroscopic reservoir experiments from the previous sections.

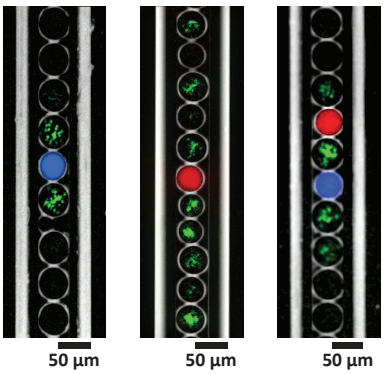

Figure 3.14: Fluorescence microscopy images of gene induction from reservoir inducer droplets. **Left** capillary: [IPTG]=10 mM global, [AHL]=200 nM in reservoir droplet (blue). Bacterial receivers in close proximity to the blue AHL reservoir droplet express GFP (green). **Centre** capillary: [AHL]=200 nM global, [IPTG]=10 mM in reservoir droplet (red). GFP expression in bacterial receivers is induced in all droplets between IPTG reservoir droplets. **Right** capillary: [AHL]=200 nM in reservoir droplet (blue) and [IPTG]=10 mM in reservoir droplet (red). GFP expression is only activated in the vicinity of both reservoir droplets.

Hence for the experiment, with IPTG distributed globally and AHL only locally, we found, as expected from the experiment from Sec. 3.7, that GFP expression is activated for droplets that are either direct neighbours to the reservoir droplets or one droplet further away. However, in the case of IPTG reservoir droplets with AHL distributed globally, GFP expression was not restricted to only neighboring droplets. We ruled out that all the GFP expression comes from a leakiness of the receiver plasmid to IPTG, by looking at the fluorescence level in droplets in the absence of IPTG reservoir droplets. Since the GFP expression for bacteria with the receiver plasmid is found to be already maximally induced at 10 μM (see Fig 3.4C) which is 0.1% of the concentration inside the reservoir droplets, we assume that the IPTG con-

centration is already high enough to induce GFP expression. However, the encapsulated TXTL for these experiments was not supplemented with extra LacI protein, so that there should not be an IPTG dependence (see Fig. 3.4A, B). Using a combination of AHL and IPTG reservoir droplets, it was then again possible to spatially confine gene expression only to bacteria/TXTL close to the AHL reservoir droplets.

3.10 From AHL senders to AHL receivers using TXTL and bacteria droplets

The receiver-sender plasmid system (Sec. 3.2) in principle allows to produce AHL *in situ*. Hence we tested the production of AHL in TXTL in bulk by the addition of both the receiver and the sender plasmid. The GFP signal was strongly reduced without the addition of AHL at the beginning, but it was possible to increase the GFP signal slightly by supplementing the TXTL with S-adenosyl methionine (SAM), a precursor for the enzymatic production of AHL with LuxI (Fig. 3.15). Additionally, the supplementation of T7 RNA polymerase leads to an earlier GFP signal, suggesting that the AHL production is sped up.

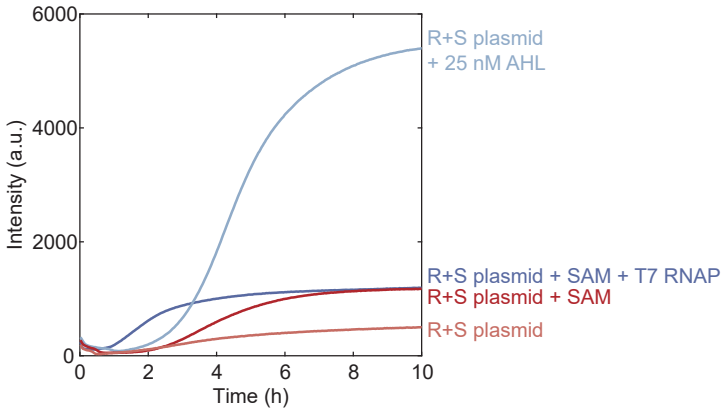

Figure 3.15: GFP expression from the receiver plasmid ($[R] = 7.5$ nM) in TXTL in a plate reader. The GFP expression is strongly reduced if the AHL is not directly added at the beginning of the reaction (e.g. 25 nM of AHL), but instead the sender plasmid ($[S] = 2$ nM) is added so that AHL is enzymatically produced. The addition of S-adenosyl methionine (SAM) which is a substrate for the production of AHL via LuxI can increase the response of the receiver plasmid to the sender plasmid in TXTL. Addtionally purified T7 RNA polymerase speeds up the GFP response due to faster expression of LuxI.

We then proceeded to encapsulate the sender and receiver plasmid in either bacteria or TXTL in order to create, to our knowledge for the first time, chemical communication between either bacterial senders and cell-free receiver droplets or *vice versa*. As shown in Fig. 3.16A, the bacterial receivers express GFP as a result of the AHL production in TXTL senders. In addition, we saw that the GFP intensities of the droplets with bacterial receivers is strongly dependent on the position within the capillary. After about one hour the GFP signal for the droplets at the center of the capillary plateaued, whereas the singal of the receiver droplets closer to the ends of the capillary continued to rise. The droplets closest to the opening of the capillary had the strongest GFP signal after about 6 hours of increase (Fig. 3.16b). We assume that oxygen availability limits the GFP signal within the capillary, since both bacterial growth and GFP maturation require oxygen. Given that the oxygen has a high solubility in fluorinated oil (about $\approx$ 77 ml of oxygen gas per 1 l (at 1 bar) and can be transported relatively fast through the continous phase with a diffusion coefficient of about $D \approx 8300\ \frac{\mu m^2}{s}$ [212], we assume that the fresh oxygen diffusing in at the end of the capillary gets consumed by the bacterial droplets at the opening of the capillary before it can reach the center of the capillary, establishing an oxygen gradient that is reflected in the GFP signal.

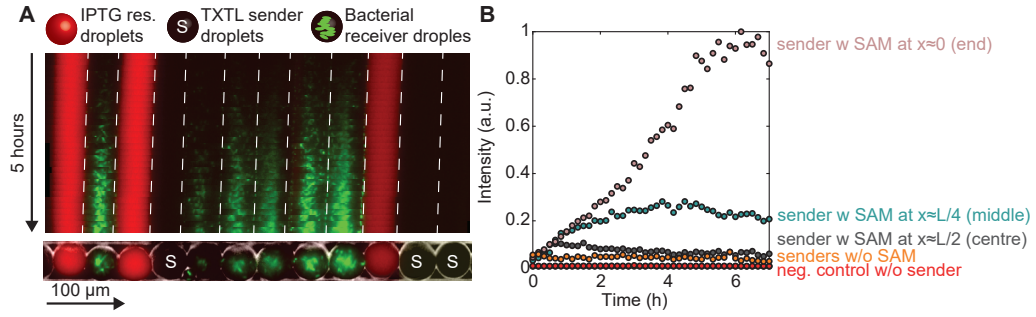

Figure 3.16: **A** A kymograph from a fluorescent microscopy image showing communication between TXTL sender droplets (with SAM) and bacterial receiver droplets in the presence of IPTG reservoir droplets (50 mM). The AHL is produced in the TXTL sender droplets and diffuses the the bacterial senders droplets, in which the bacteria grow and start to express GFP (green). **B** The average GFP fluorecnce intensity time trace of about 20 bacterial receiver droplets without any TXTL sender droplets, with TXTL sender droplets without SAM, and with TXTL sender droplets with SAM for three different positions in the capillary of length L. At the opening of the capillary $x = 0$, half-way to the center of the capillary $x = L/4$ and at the center of the capillary $x = L/2$.

In order to make the TXTL receiver droplets sense the AHL produced

by bacteria, we changed the plasmid to the sender plasmid in *E. coli* and the receiver plasmid in TXTL. Then the direction of communication was reversed so that GFP expression was induced in the TXTL droplets by the AHL from the sender bacteria (Fig. 3.17A). The GFP signal had the same intensity throughout the whole capillary, in contrast to the bacterial receiver droplets. Furthermore, we were able to differentiate TXTL receiver droplets from their GFP expression kinetics as being a direct neighbor to a bacterial sender droplet or not (Fig 3.17B).

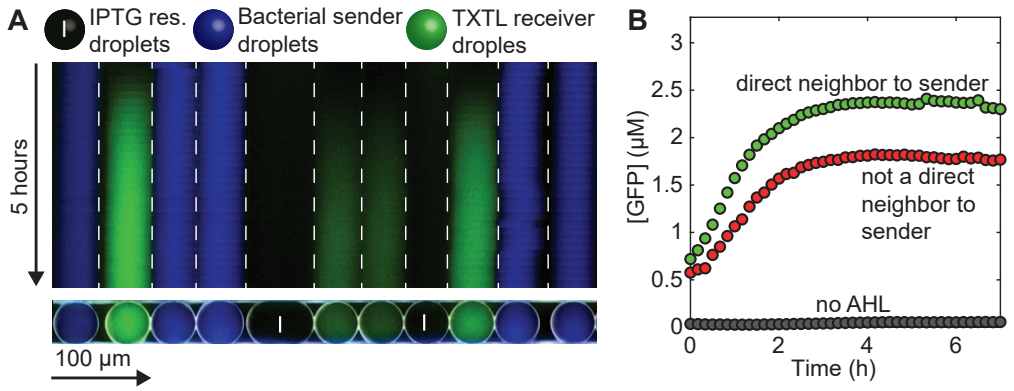

Figure 3.17: **A** A kymograph from a fluorescent microscopy image showing communication between bacterial sender droplets and TXTL receiver droplets in the presence of IPTG reservoir droplets (50 mM). The AHL produced in the bacteria diffuses to the TXTL droplets and induces GFP expression (green). **B** Time trace of the GFP expression kinetics in the TXTL receiver droplets without and with bacterial sender droplets. Furthermore one can differentiate if a bacterial sender is a direct neighboring droplet to a TXTL receiver droplet.

3.11 Conclusion

We have demonstrated spatially differentiated gene expression in linear chains of emulsion droplets containing either bacteria or cell-free gene expression systems. Spatial effects can be either achieved by establishing concentration gradients of genetic inducers such as AHL and IPTG via macroscopic reservoirs or local sender droplets. For low inducer concentrations, the apparent diffusion coefficient of AHL is reduced by several orders of magnitude, consistent with micelle-mediated transport between neighboring droplets. Confinement to an effective 1D geometry leads to a stronger chemical coupling of neighboring droplets and an extended spatial range of the inducer molecules. This also enabled us to establish communication between cell-free sender

droplets and neighboring bacterial receiver droplets and, conversely, between sender bacteria and cell-free receivers. Our experiments represent a step towards artificial multicellular hybrid systems, in which the "cells" are constituted either by encapsulated bacterial consortia, artificial cells containing cell-free gene expression systems, or supplementary cells with chemical supplies, nutrients, etc. The 1D geometry utilized in this work proved particularly useful to achieve spatial differentiation in such systems and to generate interesting spatiotemporal effects, thus emulating simple developmental processes in an artificial context.

3.12 Materials and Methods

3.12.1 Preparation of the cell-free gene expression system

Crude S30 cell extract was obtained by beat beating a BL21-Rosetta2(DE3) mid-log phase culture with 0.1 mm glass beads in a Minilys device (Peqlab, Germany) as described in Ref. [68]. Instead of 3-phosphoglyceric acid (3-PGA), phospho-enolpyruvate (PEP) was utilized as an energy source [213] in the composite buffer (50 mM Hepes pH 8, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/ml tRNA, 26 μM coenzyme A, 0.33 mM NAD, 0.75 mM cAMP, 68 μM folinic acid, 1 mM spermidine, 30 mM PEP, 1 mM DTT, 2 % PEG-8000). All components were stored at -80°C before thawing on ice. The composition of a single cell-free reaction was: 33 % (v/v) S30 cell extract mixed with 42 % (v/v) buffer and 25 % (v/v) DNA plus additives. Each final reaction mix was supplemented with 13.3 mM maltose for ATP regeneration [200]. 300 μM of S-(5'-Adenosyl)-L-methionine chloride dihydrochloride (Sigma-Aldrich, A7007) and 1 U/μl of T7 polymerase (Epicentre, TM910K) was added to the cell free sender droplets. LacI was His-tagged and purified by gravity-flow chromatography with Ni-NTA Agarose Beads (Qiagen).

3.12.2 Bacterial strains and culture media

The receiver plasmid was cloned into Escherichia coli DH5αZi (ExpressSys) and the sender plasmid into *E. coli* BL21(DE3)pLysS (Promega) as described previously [186]. As culture media, Luria-Bertani (LB) medium (Carl Roth, X968.1) and M9 minimal supplemented with 20 mM glucose and 300 μg/ml thiamine hydrochloride (Sigma-Aldrich, T1270) were used. For IPTG reservoir experiments with reiceiver bacteria, M9 minimal medium was supple-

mented with 0.4 % (w/v) glycerol instead of glucose. Cells from glycerol stock were grown overnight in Falcon tubes with 5 ml LB medium containing 100 μg/ml carbenicillin (Carl Roth) and 30 μg/ml chloramphenicol (Carl Roth) at 37°C shaken at 250 rpm. The overnight cultures were diluted to an initial optical density (OD 600nm) of 0.01 in Falcon tubes containing fresh LB or M9 minimal medium (Sigma-Aldrich, M6030) for sender or receiver cells, respectively (LB medium is autofluorescent in the GFP channel) with 50 % reduced antibiotic concentrations. Before encapsulation of bacteria into microemulsion droplets, the cultures were incubated for 2 - 3 h until an OD 600nm of 0.1 – 0.2 was reached. Sender bacteria droplets were grown in LB medium supplemented with 1 mM IPTG, two hours before encapsulation.

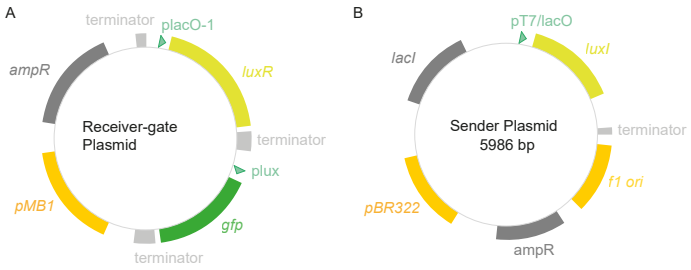

Figure 3.18: Plasmid Maps of constructs used in this study: the AHL receiver plasmid (a) is based on the BioBrick part pSB1A3 with insert T9002 completed with lacO-1 promoter. (b) pETDuet-1 vector was used for the sender plasmid with the BioBrick insert C0061 (luxI). Contsruction details can be extracted from Ref. [186].

3.12.3 Microfluidics

For the generation of droplets, we utilized a microfluidic flow-focusing geometry, which was defined using soft lithographic techniques and the elastomer poldimethysiloxane (PDMS). PDMS devices were micromolded from silicon masters containing the channel structures, which were defined using the negative photoresist Epocore 20 (micro resist technology, Germany). The cured microfluidic device was bonded onto a glass cover slide after oxygen plasma treatment and baked at 200°C for 3 hours [214]. To avoid cross contamination, each droplet species was produced in a separate device. Fluid flows were generated with a pressure controller OB-1K (Elveflow, France) and appropriate PTFE-tubings (inner diameter: 0.8 mm). Pressures were set between 150-300 mbar. The droplet device was loaded with the carrier oil Fluorinert FC-40 (Sigma-Aldrich, Germany) blended with 2% (w/w) surfactant (EA, RainDance Technologies, USA) and focused with the appropriate aqueous

phase. The critical micelle concentration of the surfactant is ≈ 0.02-0.04 % [215]. The reservoirs and the collection tubes for the cell free system were placed on ice during droplet production. Different species of droplets were collected separately and then mixed in a new test tube in defined volume ratios with a pipette. For capillary measurements with inducer reservoirs, the capillary ends were placed in a chamber containing the reservoir solution. The chamber was built with two stripes of melted parafilm on a glass slide, immediately covered with another glass slide. The chambers were sealed with silicone vacuum grease and nail polisher.

3.12.4 Bulk characterization

Cell-free expression of GFP was characterized via plate reader measurements (BMG FLUOstar Optima) using 15 μl reaction volumes in 384-well plates.

3.12.5 Microscopy

Video microscopy of droplets containing only bacteria was performed with an inverted epifluorescence microscope IX81 (Olympus, Japan) using a 10x magnification objective, an EMCCD camera (iXon3, Andor, UK), and a mercury fluorescence excitation light source (X-Cite 120Q, Excelitas Technologies, USA). The microscope was equipped with an incubator box (Okolab, Italy) to maintain a temperature of T=37°C. All time-lapse microscopy measurements containing the cell-free system, also in combination with bacteria, were conducted on an IX71 microscope (Olympus, Japan) using a 10x objective, a CCD camera (LucaR, Andor, UK) and a 4-wavelength fluorescence LED source (Thorlabs, USA). The samples were thermostatted at 30°C using a heating plate (Tokai Hit Co. Ltd., Japan).

3.12.6 Data analysis

Image analysis is performed using a customized automated droplet tracking software programmed in MATLAB [216]. Time traces of cell-free droplets were generated by taking the mean fluorescence intensities over the whole droplet area (determined from bright field images). Since bacteria accumulate in the centre of the droplet, in this case the mean was taken only from bright pixels (defined by automated thresholding). Intensity traces were then normalized by subtracting the minimum of all traces of one capillary. Based on the distance of the droplets the time traces of multiple (2-3) capillaries were averaged. Subsequently, time traces were normalized to the maximum intensity for bacterial receivers. For cell-free receivers, a calibration with

droplets containing purified GFP was performed. Finally, most of time traces were smoothed in MATLAB with the function smoothn (S=10-100) to reduce the noise and account for missing data points. For data spanning more than one observation area, i.e. 1 μM and 10 μM AHL reservoirs, images A and B were recorded with an overlap so that a reference droplet was present in each image. Accordingly, extracted data was stitched by transforming the positions of droplets in B to the coordinate system of A. It was ensured that droplets present in both images were not taken twice. Positions of droplets were calculated as the distance to the first droplet plus the radius of the first droplet. Pixel sizes of the cameras are 8 μm and 13 μm for IX71 and IX81, respectively. For data of hybrids, droplets were sorted manually dependent on their number of sender neighbours/their position inside the capillary and averaged.

3.12.7 Filling of the capillaries

Squared borosilicate glass capillaries with inner dimensions 50 μm x 50 μm x 5 cm (VitroCom, USA) were fixed by placing one end of the capillary between a freshly activated PDMS block and a glass slide (Fig. 3.19). The end of the capillary was positioned within a pre-punched hole in the PDMS block (diameter 1.25 mm), into which the droplet solution was pipetted. The capillaries were filled by manually applying pressure on the PDMS block. Filled capillaries were sealed with vacuum grease and nail polisher. For the experiments with several droplet species, we scanned for and then focused on appropriate droplet configurations.

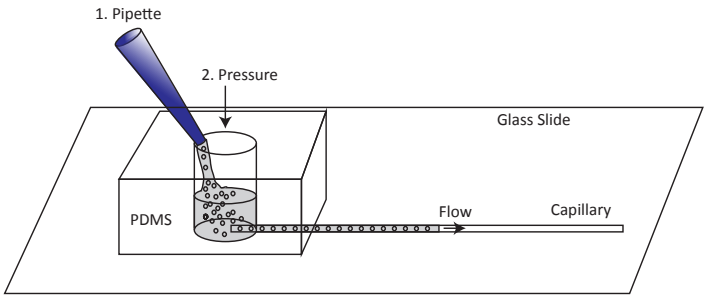

Figure 3.19: Filling of capillaries.

3.13 Contributions

The experiments were conceptualized by Matthaeus Schwarz-Schilling (M.S.) and Friedrich C. Simmel (F.S.). The cloning of the sender-receiver plasmids was done by Andrea Mückl (A.M.). The idea of using glass capillaries came up in a discussion with Michael Heymann. Experimental work was done by M.S. and Lukas Aufinger (L.A.). Characterization of the TXTL in plate-reader was done by M.S and of bacteria by Ronja Berg and Maximlian Weitz. Preparation of cell-free expression system was done by M.S. and A.M. Tracking of fluorescence intensity time traces in droplets was performed by L.A. Loading of the capillaries was done by M.S. Chemical communication experiments were conducted by M.S. Data analysis and modeling were done by M.S., L.A., and F.S.

Chapter 4

RNA-protein nanostructure in a cell-free expression system

4.1 Bionanotechnology

Living organisms are full of intricate structures on the nanoscale made from different molecules. Since the twentieth century, researchers are able to resolve the complex structure of biomolecules within cells, from the discovery of the double helix of DNA to the structure of the ribosome. Structural determination is often coupled to understanding the mechanisms behind the molecules' functions. There are libraries of structural motifs from determined protein structures (rcsb.org) and from nucleic acid structures (ndb-server.rutgers.edu).

With the understanding of structural and functional aspects of bio-molecules comes the will to design and create artificial structures. First, artificial nanostructures from biomolecules were proposed in 1982 by Nadrian Seeman [144]. The field of DNA nanotechnology has progressed since then and has also inspired the use of RNA to design artificial nanostructures. Some examples of RNA nanostructures have been introduced in section 2.7.

4.2 Summary

The contents of this chapter have been published in the artice: "Optimized assembly of a multi-functional RNA-protein nanostructure in a cell-free gene expression system" by Matthaeus Schwarz-Schilling, Aurore Dupin, Fabio Chizzolini, Swati Krishnan, Sheref S. Mansy and Friedrich C. Simmel in *Nano Letters*, 2018, 18 (4), 2650-2657.

The expression and assembly of an RNA-protein nanostructure is demonstrated and optimized within a cell-free expression system. The RNA structure is composed of different structural motifs. In total four different aptamers and one three-way junction derived from the phage ϕ29 RNA (pRNA) have been used. We used a variety of assays to verify that each aptamer works as intended.

The fluorescence enhancing effect of the malachite green (MG) binding aptamer was tested using a fluorescence spectrometer. The function of the aptamer is to report the concentration of the RNA in an expression experiment. Although there are other fluorescence enhancing RNA aptamers, like Spinach [217] with higher brightness , the advantage of MG is that its fluorescence emission is in the red spectrum with a peak at about 650 nm. In this way the signal does not interfere with other fluorescent probes that are also used in this study, and absorb and emit in the green and yellow parts of the light spectrum.

The main function of the RNA structure is to serve as a scaffold for two proteins. For this purpose we used a previously optimized RNA scaffold unit for two proteins from Sachdeva *et al.* [157]. The scaffold consists of two aptamers. The first is from the phage PP7 that binds a domain of the PP7 coat-protein (PCP) [161]. The second aptamer is from a trans-activating response RNA element in the bovine immunodeficiency virus (BIV) that binds the Tat peptide, a protein which promotes the enhancement of transcriptional elongation of the viral genes [218]. The two RNA binding domains of the proteins are fused to fluorescent proteins that form a FRET-pair. The Tat peptide is fused to the FRET donor mTurquoise2 which is from the CFP family. The PCP peptide is fused to the FRET acceptor protein YPet which is from the YFP family. The PCP was chosen for the FRET acceptor since it forms non-aggregating dimers and in this way there are one donor and two acceptors per RNA molecule. The scaffold unit was designed in such a way that proteins are positioned within the FRET radius of about 5.7 nm [219]. In total, the FRET system is fully genetically encoded with each component being fluorescent on its own or with the help of MG in case of the RNA nanostructure. This allows for real-time observations of a) the expression of each component and b) the assembly of the components into the RNA-protein complex.

The fourth aptamer is a streptavidin binding aptamer. It was tested using a binding assay with streptavidin coated magnetic beads and subsequent elution by biotin. The RNA-nanostructure and the RNA-bound proteins can be localized on the microbeads after expression in droplets *via* the aptamer. Co-localization of each component on the bead is another method to verify assembly of the RNA-protein nanostructure..

The four aptamers are connected via the three-way junction (3WJ) domain of the pRNA molecules of the phage ϕ29. The three-way junction was already proven to guide the folding of aptamers attached to its three arms [220]. Additionally it provides stability against denaturation by temperature or urea. Even more important for the expression within a cell-free expression system and cells, it protects the aptamers within the structure against degradation by RNase. We followed the fluorescence intensity of the MG aptamer for various versions of the RNA structure in the *E. coli* derived cell-extract. The closer the aptamers were to the 3' end of the RNA the faster they were degraded. The half-life of the middle arm where the protein FRET pair is assembled was measured to be about $\tau = 9.55 \pm 0.75$ hours.

Then the assembly efficiency of the RNA-protein complex was assessed by titration of purified proteins against the purified RNA with a fluorescence spectrometer. The donor quenching due to FRET of the scaffolded proteins, in combination with known concentrations of each component and literature values for the equilibrium binding constants, allowed for a quantification of the assembly process. The efficiency of complex assembly was quantified by considering the fraction of fully assembled complex to the sum of monomers. This metric captures the fact that the production of monomers that are not used in the assembled complex is simply a waste of resources. We believe that this aspect is worth to be taken into account in the process of designing the production of molecular complexes in closed compartments or even in cells with finite amount of building blocks.

We chose a cell-free expression system for the expression and assembly of the RNA-protein nanostructure, as we can easily screen parameters such as DNA template concentration for different components. Furthermore, it provides more control over the expression process as individual components and reaction reagents can be added in a step-by-step manner. After the expression and assembly of the RNA-protein was optimized for the cell-free expression system, we encapsulated the genetic constructs encoding the RNA-protein nanostructure in water-in-oil emulsion droplets together with streptavidin coated beads. Under a microscope we were able to follow the expression of the proteins and the RNA nanostructure. In addition we saw a co-localization of proteins on the streptavidin coated beads, suggesting that the RNA is the link that assembles the three proteins, the FRET-pair and streptavidin.

Finally we transformed the plasmids encoding the RNA-protein structure in *E. coli*. Using a fluorescence spectrometer we were able to verify the expression of the FRET-pair and the RNA nanostructure in cells. However, as the expression was not optimized in bacteria, the ratio of the components was not ideal for assembly. The constraints for the expression of the complex in bacteria are stronger. We noticed that the protein expression slows down

the growth rate, and that the relative expression levels of the two proteins compared to each other can fluctuate between experiments. This had a significant effect on the reproducibility of the expression levels and reduced the FRET signal of the assembled complex.

4.3 Introduction

Over the past years there has been tremendous progress in the assembly of complex and large nanostructures from biomolecules. Structural DNA nanotechnology, and the DNA origami technique [90] in particular, has enabled the creation of large molecular lattices and crystals [145] with tunable properties, as well as almost arbitrarily shaped discrete molecular objects [221] [222] [148]. RNA has also been successfully utilized for the creation of nanostructures [223] [156] [224]. In contrast to DNA self-assembly, which is most often based on thermal annealing of the constituting DNA strands, RNA nanostructures can be created enzymatically via transcription from a DNA template by RNA polymerase [155] [152]. RNA folding occurs intramolecularly during transcription, which is under kinetic rather than purely thermodynamic control. Recently, protein nanotechnology has taken major strides, and a variety of artificial protein nanostructures [225] as well as de novo-designed small proteins and protein assemblies have been demonstrated [226] [227].

Most likely future biomolecular nanostructures will consist of combinations of several types of biomolecules, i.e. RNA or DNA combined with peptides or proteins [228] [229], possibly involving lipids [230] and small molecules ("co-factors") [160] [231] RNA-protein nanostructures are of particular interest in this context, as both RNA and protein components can be produced enzymatically via transcription and translation of the corresponding gene sequences. In fact, whereas biological systems barely utilize DNA in a structural context, RNA nanostructures do occur in nature, where they are most often found interacting with proteins. Prominent examples of such structures are the ribosome [232], the spliceosome [233], RNase P [234], or the portal motor of bacteriophage ϕ29 [235].

Interactions between RNA and other biomolecules such as proteins are mediated by small aptamer motifs, and these have already been frequently utilized for tagging of RNA molecules with fluorescent proteins or enzymes [162]. Functional aptamers as well as structural RNA motifs can be modularly combined to create RNA nanostructures,[154], which can be utilized as enzyme scaffolds that increase the flux through metabolic pathways [157] or perform regulatory actions [236].

The constituents of RNA-protein nanostructures potentially can be coexpressed and assembled within a single reaction compartment or inside of cells [237]. However, efficient production of multicomponent structures in this way faces a number of challenges which have not been systematically addressed so far. For the assembly of multiple proteins along an RNA scaffold one has to consider the potentially different expression levels of the components as well as their variable binding strengths. Accordingly, uncoordinated production of the components under resource-limited conditions decreases the yield of correctly assembled structures.

In order to address this issue, we here quantitatively study the expression and assembly of an RNA-protein hybrid structure using a cell-free gene expression system. The RNA nanostructure investigated here (Fig. 4.1) is based on the pRNA three-way junction (3WJ) motif [164] and is functionalized with four different aptamers, three of which interact with proteins (streptavidin and two fluorescent proteins), while one binds the small molecule dye malachite green (MG). Utilizing the fluorescence of the MG aptamer [238] and Förster resonance energy transfer (FRET) between the fluorescent proteins bound to the RNA nanostructure scaffold, we can assess critical assembly parameters such as the efficiency of protein co-assembly and the stability of the hybrid nanostructure. Using a variety of genetic constructs encoding the RNA and protein components of the structure, we screen for conditions resulting in the most efficient assembly inside of a cell-free expression system. We then demonstrate the production of the multifunctional RNA nanostructures and the fluorescent proteins as well as their co-assembly in a single cell-free expression reaction, and also their aptamer-mediated localization onto streptavidin-coated substrates. We finally show that our in vitro-prototyped hybrid nanostructure can also be assembled in vivo.

In a related study by Jepsen *et al.* [239], an RNA nanostructure modified with two fluorescent RNA aptamers (requiring chemically synthesized chromophores) was utilized as a novel type of in vivo FRET sensor. By contrast, the FRET-generating components of the RNA-protein nanostructure developed here are fully genetically encoded. While in the present work, FRET is used merely as an analytical tool for the optimization of protein assembly along the RNA scaffold, an application as in vivo sensor is also conceivable.

4.4 RNA nanostructure design

We chose the 40 nt-sized pRNA three-way junction (3WJ) as the central structural motif of our RNA nanoscaffold (Fig. 4.1A) to ensure a compact tertiary structure and to improve co-transcriptional folding of the four ap-

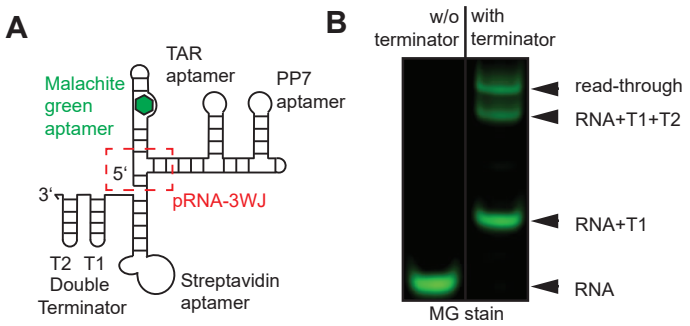

Figure 4.1: Design and characterization of the RNA scaffold. **A** Overview scheme of the RNA nanostructure. The RNA is functionalized with four aptamers: a malachite green (MG) aptamer at the 5' end, TAR and PP7 aptamers, which are positioned one helical turn away from each other, and a streptavidin aptamer at the 3' end. The core of the structure is formed by a three-way junction (3WJ) motif from bacteriophage phi29 (boxed in red). Two transcriptional terminators are placed behind the structure. **B** Native gradient PAGE (4-20%) stained with malachite green, of the RNA nanostructure after in vitro transcription with T7 RNAP from a linear template with and without terminators.

tamers. The pRNA 3WJ motif has been previously shown to be an unusually robust structure that dominates the tertiary structure of RNA strands up to four times its length [220]. The first RNA aptamer attached to the scaffold is the fluorescent aptamer for Malachite Green (MG) [240], which serves as a fluorescence reporter for the total amount of RNA nanostructure generated. Instead of brighter fluorescent RNA aptamers such as Spinach [217], we chose the MG aptamer for this study since its fluorescence spectrum does not overlap with the spectra of the protein-based FRET-pair. The second and third RNA aptamers are used to recruit proteins to the RNA scaffold. In their biological contexts, the 28 nt long TAR aptamer from bovine immunodeficiency virus (BIV) binds the 2 kDa transcriptional activator protein Tat [241], while the 25 nt long PP7 aptamer forms a complex with a 122 amino acid (a.a.) peptide subsequence of the phage PP7 coat-protein (PCP) [242]. The two aptamers are placed one helical turn away from each other in order to minimize the distance between the recruited proteins. Finally, a fourth, streptavidin-binding RNA aptamer [243] is used to facilitate localization of the RNA-protein nanostructures onto streptavidin-coated substrates.

For an initial structural characterization, we in vitro transcribed a linear gene template coding for the RNA nanostructure and purified the RNA via phenol-chloroform extraction, elution from a spin column, and ethanol precipitation (see Sec. 4.11). Using transmission electron microscopy, the ter-

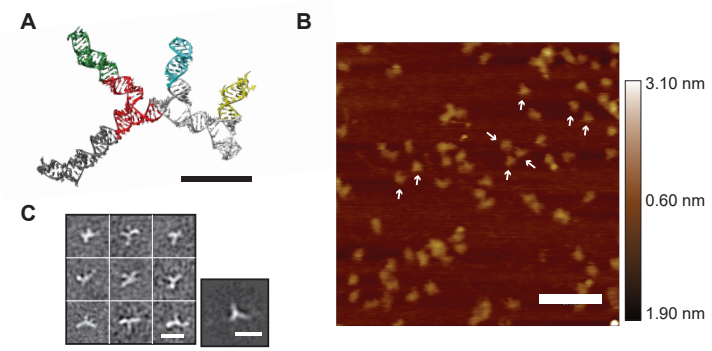

Figure 4.2: **A** An illustration of the predicted three-dimensional structure of the RNA without the terminators (image created with Vfold3D [246], scale bar: 5 nm). **B** AFM image of purified RNA structures on mica. The images in air confirmed the presence of a compact tertiary structure (white arrows). Scale bar: 100 nm. **C** Negative stain TEM micrographs of purified RNA nanostructures without terminators and a EM class average image (right). Scale bar: 20 nm.

tiary structure of the RNA with its three arms and appended aptamers could be visualized (Fig. 4.2C). Imaging using atomic force microscopy (AFM) in air confirmed the presence of a compact tertiary structure (Fig. 4.2B). As we used the strong T7 RNA polymerase (RNAP) [244] for *in vitro* transcription of the RNA nanostructure, we had to extend the RNA sequence by two transcriptional terminators [245], resulting in $\approx$ 70% termination overall (Fig. 4.1B).

4.5 Protein scaffold optimization using FRET

In order to assess the efficiency of co-assembly of two proteins *via* the TAR and PP7 aptamers, we tagged two fluorescent proteins with the corresponding peptides. Specifically, we were interested in the fraction of correctly assembled complexes, which we inferred from the efficiency of FRET between the fluorescent proteins.

Our FRET pair (with a Förster radius of about 5.7 nm [219]) consisted of the CFP variant mTurquoise2 [247] as a donor fused to the Tat peptide and the YFP variant YPet [248] as an acceptor fused to the PP7 coat-protein (for the following quantitative treatment of assembly efficiency, RNA scaffold, CFP and YFP proteins are abbreviated with r, c, and y, respectively).

Binding of the proteins to the RNA was first confirmed using an electrophoretic mobility shift assay (EMSA) and RNA-protein complexes were imaged using the transmission electron microscopy (Fig. 4.3B and D). The

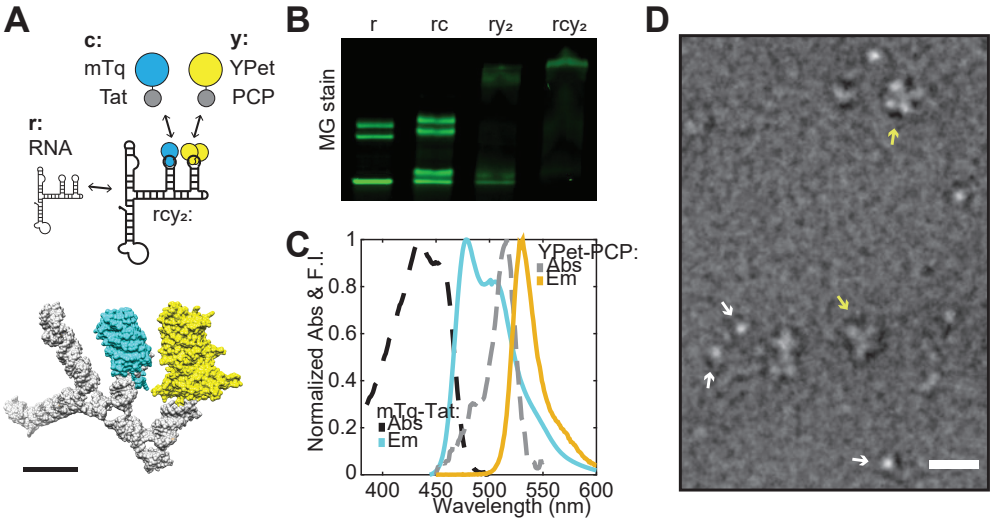

Figure 4.3: Characterization of fluorescent protein binding. **A** Top: Scheme showing mTurquoise2 fused to the Tat peptide (abbreviated c) and YPet fused to the PP7 coat protein (PCP) (abbreviated y) binding to the corresponding aptamers on the RNA structure (abbreviated r), resulting in the four-component complex rcy2. PCP preferentially binds to the PP7 aptamer as a dimer. Bottom: illustration of the assembled complex rcy_2 based on the RNA structure and the crystal structures of the proteins and peptides (created with UCSF Chimera, scale bar: 5 nm). **B** EMSA (native gradient PAGE) stained with MG showing the binding of the His-tag purified proteins c and y to r. **C** Normalized absorption and fluorescence emission spectra of the purified protein FRET-pair, mTurquoise2-Tat (mTq-Tat) and YPet-PCP. **D** Negative stain TEM micrographs of purified RNA nanostructures without terminators and purified proteins (white arrows) and assembled RNA-protein nanostructures (yellow arrows). Scale bar: 20 nm.

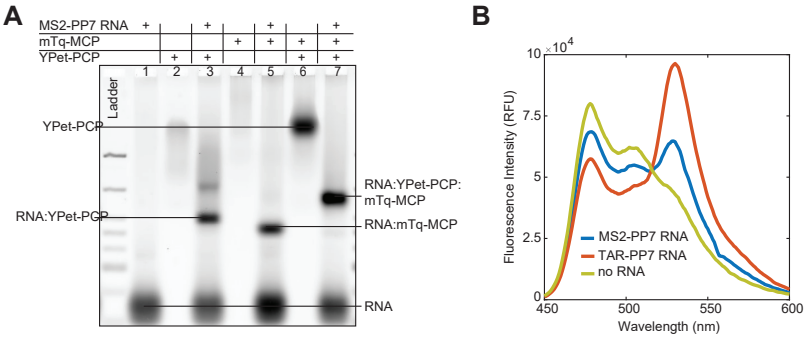

Figure 4.4: Characterization of fluorescent protein binding. **A** EMSA using agarose gel electrophoresis, 4% (w/v), stained with an intercalating dye showing the MS2-PP7 RNA and the appearance of a new band while binding to mTurquoise2-MCP and YPet-PCP (auto-fluorescence is also partially visible). The concentration of YPet-PCP and mTurquoise2-Tat in lanes 6 and 7 is two times as high as in lane 2-5. **B** Fluorescence emission spectra of purified mTurquoise2-Tat (250 nM) and YPet-PCP (500nM) without any RNA, with RNA combining the TAR and PP7 aptamers (250 nM). As a comparison, the fluorescence emission spectra of an RNA with the MS2 and PP7 aptamers (250 nM) plus mTurquoise2-MS2 (250 nM) + YPet-PCP (500 nM). Samples were excited at 440 nm. The donor signal at 480 nm is lower for the TAR-PP7 sample and the acceptor signal at 530 nm is higher, although the same concentrations of the components were used in both samples, indicating more energy transfer for the RNA with the TAR and PP7 aptamer.

fluorescence emission spectrum of the FRET donor protein overlaps with the excitation spectrum of the FRET acceptor protein (Fig. 4.3C).

We also investigated the use of the commonly used peptide binding aptamer from phage MS2 instead of TAR for binding of mTurquoise2 to the RNA nanoscaffold. Interestingly, even though MS2 has a lower K_d than TAR for binding to its cognate peptide, we found that in terms of FRET efficiency the combination of TAR with the PP7 aptamer was superior (Fig. 4.4), which potentially is due to the different spacing and orientation between the scaffolded proteins in this case.

4.5.1 Determining the complex concentration with FRET

The yield of correctly assembled complexes was estimated from the decrease of donor fluorescence intensity due to FRET upon the addition of the RNA scaffold (Sec. 4.5.1 and 4.5.2). Using literature values for the equilibrium

constants of the TAR-Tat and PP7-PCP interactions as well as considering the dimerization of PCP, we calculated the concentration of fully assembled complexes $[r \cdot c \cdot y_2]$ as a function of the total concentrations of its components (c_0, y_0 and r_0) (Table 4.1 and Fig. 4.6C and D).

The FRET efficiency having 1 donor and 2 acceptors [249] as a function of the distance d between the proteins on the RNA is given by:

$$E_2(d) = \frac{2E_1}{1+E_1} = \frac{2R_0^6}{d^6 + 2R_0^6} \tag{4.1}$$

The total fluorescence intensity of mTurquoise2-Tat + YPet-PCP + RNA nanostructure at $\lambda = 474$ nm is given by

$$F(474) = m_{474}([c] + [cr] + [rcy_2](1 - E_2)) \tag{4.2}$$

Where m_{474} is defined by Eq. 4.17. Therefore, the difference in fluorescence intensity ΔF before and after addition of the RNA scaffold is:

$$\frac{\Delta F}{m_{474}} = [rcy_2] \cdot E_2 \tag{4.3}$$

$$\rightarrow E_2 = \frac{\Delta F}{m_{474} \cdot [rcy_2]} \tag{4.4}$$

In order to calculate $[rcy_2]$ we assume that the binding of Tat to the RNA is independent to that of PCP-dimer to the RNA. Then the concentration of the fully assembled complex is given by the product of the fraction of RNA scaffold bound to Tat or PCP and the concentration of the RNA.

$$[rcy_2] = f_{cr} \cdot f_{ry} \cdot [r]_0 = \frac{[cr]}{[r]_0} \cdot \frac{[ry_2]}{[r]_0} \cdot [r]_0 \tag{4.5}$$

Here

$$[cr] = \frac{1}{2}\left([c]_0 + K_c + [r]_0 - \sqrt{([c]_0 + K_c + [r]_0)^2 - 4[c]_0[r]_0}\right) \tag{4.6}$$

$$[ry_2] = \frac{1}{2}\left([y_2] + K_y + [r]_0 - \sqrt{([y_2] + K_y + [r]_0)^2 - 4[y_2][r]_0}\right) \tag{4.7}$$

and taking the dimerization of PCP into consideration:

$$[y_2] = \frac{1}{2}\left(2[y]_0 + K_{2y} - \sqrt{(2[y]_0 + K_{2y})^2 - 4[y]_0^2}\right) \tag{4.8}$$

Plugging Eq. 4.8 into 4.7 and then 4.6 and 4.7 into Eq. 4.5 allows us to calculate $[rcy_2]$ as a function of $[c_0], [y_0], [r_0]$:

$$\begin{aligned}
[rcy_2] &= \frac{1}{2[r]_0}\left(\frac{[c]_0+[r]_0+K_c-\sqrt{([c]_0+[r]_0+K_c)^2-4[c]_0[r]_0}}{2}\right) \\
&\cdot \left\{\frac{1}{2}\left(2[y]_0+K_{2y}-\sqrt{(2[y]_0+K_{2y})^2-4[y]_0^2}\right)+[r]_0+K_y\right. \\
&- \left[\left([y]_0+K_{2y}-\sqrt{(2[y]_0+K_{2y})^2-4[y]_0^2}+\frac{[r]_0+K_y}{2}\right)^2\right. \\
&- \left.\left.2\left(2[y]_0+K_{2y}-\sqrt{(2[y]_0+K_{2y})^2-4[y]_0^2}\right)[r]_0\right]^{\frac{1}{2}}\right\} \qquad (4.9)
\end{aligned}$$

Combining Eq.4.4 and 4.1 we can solve for the distance of the proteins with the values from Table 4.1 and $[rcy_2] = 320$ nM ($[c]_0 = 428$ nM, $[y]_0 = 1072$ nM, $[r]_0 = 500$ nM) :

$$d = \sqrt[6]{2R_0^6\left(\frac{m_{474}[rcy_2]}{\Delta F}-1\right)} = 5 \text{ nm} \qquad (4.10)$$

In our experiments, the concentration of YFP-PCP is well above the K_d of the PP7 aptamer ($y >> K_y$), and therefore $[r \cdot c \cdot y_2]$ is simply proportional to $[y]_0$ until it saturates roughly when the concentration of the RNA scaffold is reached (Fig. 4.5). Dimerization of PCP effectively reduces the amount of active binding partners, which shifts the point of saturation towards higher concentrations of YPet-PCP monomers. Conversely, due to its higher dissociation constant the binding curve of CFP-Tat to the RNA is more sigmoidal in our experimental regime and therefore saturates substantially above the RNA scaffold concentration (Fig. 4.5).

4.5.2 The fraction $f_{complex}$

While it is possible to generate any desired concentration of complex by just producing enough of its components, the fraction of correctly assembled complexes that can be achieved is limited. We here use $f_{complex} = [r \cdot c \cdot y_2]/([r]_0 + [c]_0 + [y]_0)$ as a figure of merit that captures the trade-off between large complex concentrations and efficient use of components. For simplicity, we assume here that production of proteins and nanostructures cause comparable "costs" (we will argue below that this is justified). Given

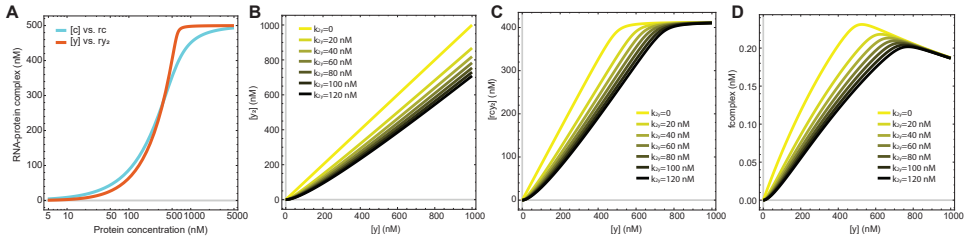

Figure 4.5: **A** Binary binding curves for mTurquoise2-Tat (c) and YPet-PCP (y) forming a complex with the RNA (either rc or ry_2) for $[r]_0 = 500$ nM. Calculated with Eq. 4.6 and 4.7 and the dissociation constants from Table 4.1. **B** The concentration of dimers as a function of monomers for YPet-PCP for different dimerization strengths. **C** The effect of the dimerization strength on the concentration of $[rcy_2]$ as a function of the monomers for $[r]_0 = 500$ nM and $[c]_0 = 700$ nM. The point of saturation of complex formation is pushed to higher monomer concentrations for weaker dimerization. Calculated using Eq. 4.9. **D** The weaker the dimerization the lower the maximum of $f_{complex}$ and the more monomers of YPet-PCP are needed to reach the maximum (calculated for $[r]_0 = 500$ nM and $[c]_0 = 700$ nM with Eq. 4.11).

a limited amount of resources, maximization of $f_{complex}$ therefore gives the largest achievable fraction of assembled complexes.

Using the value for $d = 5$ nm from above we can determine $E_2 = 0.81$. This allows us to calculate the value of $f_{complex}$:

$$f_{complex} = \frac{[rcy_2]}{[r]_0 + [c]_0 + [y]_0} \tag{4.11}$$

$$= \frac{1}{[r]_0 + [c]_0 + [y]_0} \frac{\Delta F}{E_2 m_\lambda} \tag{4.12}$$

Eq. 4.11 is only dependent on the dissociation constants (Table 4.1) and the initial concentrations of the components, which allows calculation by simply inserting into Eq. 4.9. This was used to create the heat map in Fig. 4.6C and D. On the other hand, Eq. 4.12 is based on fluorescence spectra shifts like in Fig. 4.6B. The results are compared in Fig. 4.6C.

We used the function *NArgMax* in Wolfram Mathematica v11.2 to numerically maximize $f_{complex}$ (Eq. 4.9) for $[r]_0$ from 1 to 1000 nM (in 1 nM steps) and find the corresponding concentrations of $[c]_{max}$ and $[y]_{max}$ at each point. The numerical results for the maximal points for $[rcy_2]$ are plotted in Fig. 4.6E and F .

The concentration of assembled complexes $[rcy_2]$ will always be smaller

Constants	Value	Reference
K_c	60 nM	[241]
K_y	1.6 nM	[242]
K_2y	20 nM	[251]
R_0	5.67 nm	[219]
$\Delta F/m_{474}$	230 nM	Fig. 4.6B

Table 4.1: Binding Constants and Förster radius for mTurquoise2-Tat and YPet-PCP

(or equal in the case of perfect binding) to the concentration of one of its components (the limiting species):

$$[rcy_2] \leq \min([r],[c],[y]) \rightarrow f_{complex} \leq \frac{1}{3} \tag{4.13}$$

Calculations for a fixed RNA concentration of $[r]_0 = 500$ nM indicate that $f_{complex}$ is bounded from above at $\approx 20\%$, which was experimentally confirmed in fluorescence spectroscopy measurements using purified components (Fig. 4.6B). With rising RNA concentrations $f_{complex}$ can maximally increase up to 1/3 (Fig. 4.6C and Eq. 4.13). If one allows the RNA concentration to exceed the protein concentrations, the amount of fully assembled complexes starts to decrease again due to the accumulation of scaffolds with only one protein bound (cf. the so-called "prozone" effect [250], Fig. 4.6G).

Maximization of $f_{complex}$ under constraints can be used to define the optimum expression levels of each component. For a total RNA concentration of $[r]_0 \geq 400$ nM, the concentrations of mTurquoise2-Tat and YPet-PP7 required for a maximum $f_{complex}$ are found to be approximately equal, since in this concentration regime, both proteins are well above the K_d values of their respective aptamers. For RNA concentrations $[r]_0 < 400$ nM, more mTurquoise2 is needed than YPet (i.e., $[c] > [y]$) (Fig. 4.6F), since then the Kd and the concentrations of the TAR-Tat components are of similar size. In order to increase $f_{complex}$ above 1/3, cooperative interactions between the scaffolded proteins themselves would have to be established [252] (Fig. 4.6H and Sec. 4.5.3).

4.5.3 Cooperative equilibria

Cooperativity in ternary complex formation can be described with a parameter α. For positive cooperative interaction between the proteins on the RNA

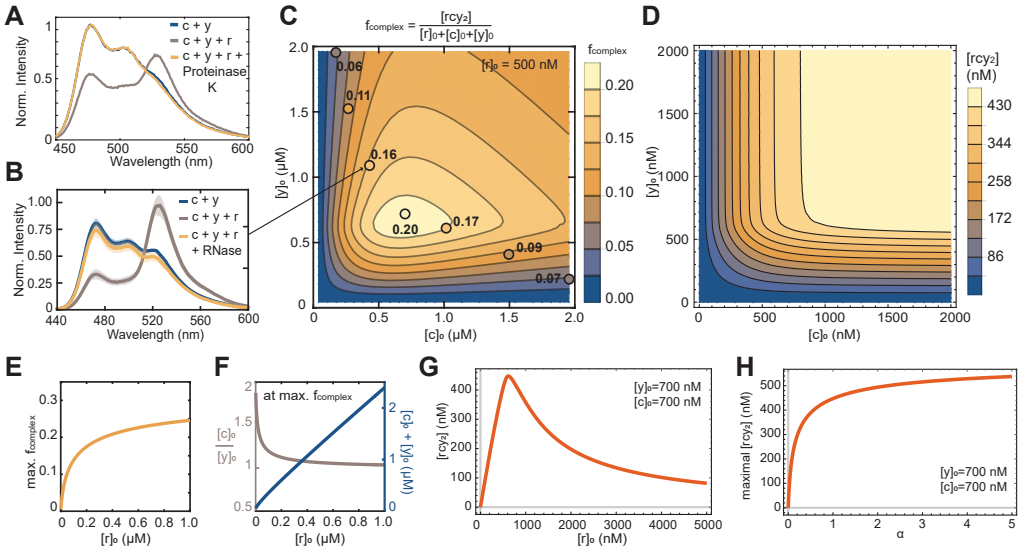

Figure 4.6: **A** Emission spectra of purified proteins ($[c]_0 = 500$ nM and $[y]_0 = 700$ nM) before, after the addition of the RNA ($[r]_0 = 500$ nM and after 10 minutes of incubation with Proteinase K. **B** Normalized emission spectra of purified $[c]_0 = 430$ nM and $[y]_0 = 1070$ nM before and after the addition of the RNA ($[r]_0 = 500$ nM) and after the degradation of the RNA by RNase (solid lines: mean value of n=3; transparent area represents the SD). **C** The fraction of fully assembled complexes $f_{complex}$, as a function of $[c]_0$ and $[y]_0$ for $[r]_0 = 500$ nM. The heat map shows calculated values of $f_{complex}$, whereas the circles represent values obtained from experiments (such as the spectra shown in B.). **D** Calculated heat map of the complex concentration $[rcy_2]$ as a function of the protein concentrations $[c]_0$ and $[y]_0$ using the constants from Table 4.1. **E** Calculation of the maximum $f_{complex}$ as a function of $[r]_0$. $f_{complex}$ approaches 1/3 (dotted line) as the concentration of the components approach infinity. **F** At maximum $f_{complex}$, the ratio of $[c]_0$ to $[y]_0$ approaches 1 as $[r]_0$ increases. On the right y-axis: Sum of protein concentrations $[c]_0 + [y]_0$ as a function of $[r]_0$ at maximal $f_{complex}$. **G** The concentration of fully assembled complexes $[rcy_2]$ decreases again if the RNA concentration exceeds the concentration of the proteins, due to the fact that the probability of having an RNA bound with only one protein increases, effectively diluting the fully assembled complexes (known as the "prozone" effect) – curve calculated with Eq. 4.9. **H** The strength of additional interaction between the proteins, quantified by $\alpha > 1$, increases the maximum of $f_{complex}$, here shown for $[r]_0 = 500$ nM, $[c]_0 = 700$ nM and $[y]_0 = 700$ nM and calculated with Eq. 4.15, whereas an repulsive interaction $\alpha < 1$ decreases $[rcy_2]_{max}$.

$\alpha > 1$ and for negative interactions $\alpha < 1$. Then derived from the law of mass action, the new equilibrium is given by:

$$\frac{K_c K_y}{\alpha} = \frac{[c][r][y_2]}{[rcy_2]} \tag{4.14}$$

It has been shown that an expression for $[rcy_2]$ cannot be obtained algebraically in terms of cooperativity α and known parameters such as dissociation constants and initial concentrations [252]. However, an expression for the maximal concentration of $[rcy_2]$ as a function of $[c]_0, [y]_0$ and α can be found:

$$\begin{aligned}
[rcy_2]_{max} &= \frac{1}{2}\Bigg[[c]_0 + \frac{1}{2}\left(2[y]_0 + K_{2y} - \sqrt{(2[y]_0 + K_{2y})^2 - 4[y]_0^2}\right) \\
&\quad + \frac{\sqrt{K_c} + \sqrt{K_y}}{\alpha}\Bigg] \\
&- \Bigg\{\Bigg[[c]_0 + \frac{1}{2}\left(2[y]_0 + K_{2y} - \sqrt{(2[y]_0 + K_{2y})^2 - 4[y]_0^2}\right) \\
&\quad + \frac{\sqrt{K_c} + \sqrt{K_y}}{\alpha}\Bigg]^2 \\
&- 4[c]_0 \frac{1}{2}\left(2[y]_0 + K_{2y} - \sqrt{(2[y]_0 + K_{2y})^2 - 4[y]_0^2}\right)\Bigg\}^{\frac{1}{2}}
\end{aligned} \tag{4.15}$$

Eq. 4.15 has been used to calculate the dependence of $[rcy_2]$ on α in Fig. 4.6H. Similarly, it has been shown that the concentration of the RNA scaffold at $[rcy_2]_{max}$ is given by:

$$\begin{aligned}
[r]_{max} &= \frac{\sqrt{K_y}}{\sqrt{K_c} + \sqrt{K_y}}([c]_0 + K_c) \\
&+ \frac{\sqrt{K_c}}{\sqrt{K_c} + \sqrt{K_y}}\left[\frac{1}{2}\left(2[y]_0 + K_{2y} - \sqrt{(2[y]_0 + K_{2y})^2 - 4[y]_0^2}\right) + K_y\right]
\end{aligned} \tag{4.16}$$

4.6 RNA stability in a cell-free expression system

In order to prototype expression and assembly of the RNA-protein nanostructure in a cell-free expression system, we used an *E. coli* based crude

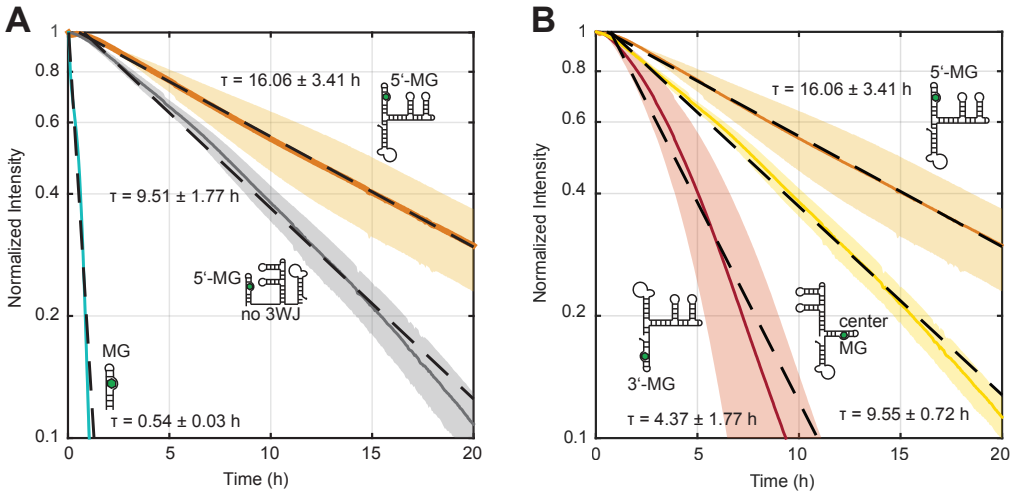

Figure 4.7: RNA degradation in cell extract. **A** Exponential degradation of different RNA structures ([r]=500 nM) by RNases present in the cell-free expression system reported by MG fluorescence. τ: Mean lifetimes for the RNA nanostructure, the four RNA aptamers without the 3WJ and the MG RNA aptamer alone. Mean and standard deviations of three replicates are indicated as full lines and shaded areas, respectively exponential fits are shown as broken lines). **B** MG fluorescence time traces in cell-free expression system for three different versions of the RNA nanostructure, for which the position of the MG aptamer was varied.

cell-extract [68], which had been previously used for the assembly of protein filaments [253]or of virus particles [64]. An important factor affecting the assembly yield in a one-pot expression reaction is the chemical stability of the RNA scaffold. Degradation of mRNA in the cell-extract based expression system was previously found to be well described by a mono-exponential decay with a half-life ranging from 12 to 18 minutes [189] [79]. While such short lifetimes would be unsuitable for applications in protein scaffolding, we surmised that the presence of strong secondary structure could increase the stability of our RNA scaffold considerably [254] [255].

To assess its stability, we studied the fluorescence of the MG aptamer for various aptamer positions on the scaffold. As shown in Fig. 4.7A, the half-life of an un-scaffolded MG aptamer was $\tau \approx 30$ min. In stark contrast, the fluorescence of the MG aptamer embedded into the RNA nanostructure decays with a half-life of $\tau \approx 16.0$ hours (Fig 4.7A). We checked the specific influence of the 3WJ motif by replacing it with a random sequence, which reduced the half-life to $\tau = 9.2$ hours. Experiments with different permutations of the four aptamers on the arms of the 3WJ (Fig. 4.7B) showed a faster fluorescence decay when the MG aptamer was positioned closer towards the

3' end of the RNA structure, suggesting that exonuclease degradation from the 3'end is a dominating factor [254]. Considering the production times of RNA and proteins, the observed stability of the RNA nanostructure is sufficient to serve as a protein scaffold (Fig. 4.9).

4.7 One-pot expression and assembly of the RNA-protein nanostructure

We next characterized the production of the individual components in the expression system. Expression levels of the fluorescent proteins were varied by changing the concentrations of their respective genetic templates. The sequences coding for mTurquoise2-Tat and YPet-PP7 were each placed into a pUC19-derived expression vector and put under the same constitutive E. coli RNA Polymerase (RNAP) promoter and the same ribosome binding site (Fig. 4.8). By changing the concentration of the plasmid for one of the proteins while fixing the other at 5 nM, we found that the optimum ratio of the protein end levels was reached for approximately equal concentrations of genetic templates (Fig. 4.8B). Even though this appears to be a relatively straightforward result, it is not trivial due to the complexity of the whole expression reaction [256]. Specifically, it is hard to make quantitative predictions of loading and crosstalk effects that result from the expression of multiple genes in parallel [79].

In order to verify the assembly of the fluorescent proteins on the RNA scaffold inside the expression system, we stopped protein translation after 14 hours by adding the ribosomal inhibitors spectinomycin and streptomycin (Fig. 4.9A) and only then added the gene construct of the RNA structure (Fig. 4.8B). To reliably observe fluorescence changes caused by FRET, we waited until the fluorescence signals had stabilized (Fig. 4.8C) and then supplemented the reaction with r-plasmids that contained the sequence for the RNA nanostructure under a T7 RNAP promoter, together with purified T7 RNAP and rNTPs. During the expression of the RNA scaffold, monitored by MG fluorescence, the donor signal in the CFP channel dropped whereas the signal in the FRET channel (excitation at 430 nm, emission at 540 nm) increased, indicating Förster transfer between the proteins that are scaffolded by the RNA.

We found that for optimal assembly of the RNA-protein nanostructure the protein concentrations were too low compared to the concentration of RNA scaffold. Increasing the amount of both genetic templates, the c-and y-plasmids, did not substantially increase the final concentrations of the pro-

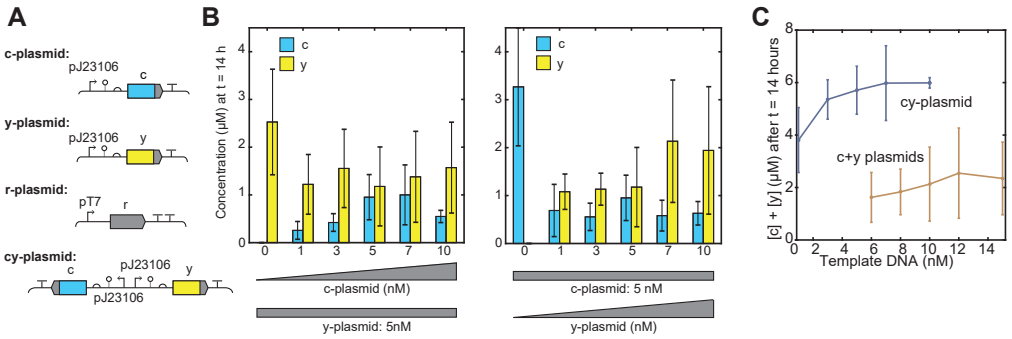

Figure 4.8: Optimization of expression and assembly of the RNA-protein nanostructure. **A** Genetic constructs encoding c, y and r. Both c and y are under a constitutive E. coli RNAP promoter (pJ23106) and have a self-cleaving ribozyme upstream of the RBS (PimJ for c and RiboJ for y). The RNA scaffold is transcribed from a T7 RNAP promoter. The cy-plasmid encodes the genes for both c and y. **B** Prototyping the expression of c and y in cell-free expression system by titration of either c or y-plasmid. Mean from three independent repeats with standard deviation shown as the error bars. In the presence of the c-plasmid, YPet-production is reduced with respect to the value obtained with only the y-plasmid. The concentration of mTurquoise2 is highest for $\approx$ 5-7 nM c-plasmid. On the other hand, when the y-plasmid is titrated, mTurquoise2 production is significantly reduced compared to expression from only the c-plasmid, and the YPet concentration is highest for 7-10 nM y-plasmid. The observed deviations are significantly higher than the standard error of the mean. **C** The mean total expression level ([c]+[y]) after 14 hours is increased if the cy-plasmid is used in comparison to expression from two different plasmids: c-plasmid and y-plasmid (shown are the mean and std. dev. from min. 3 repeats)

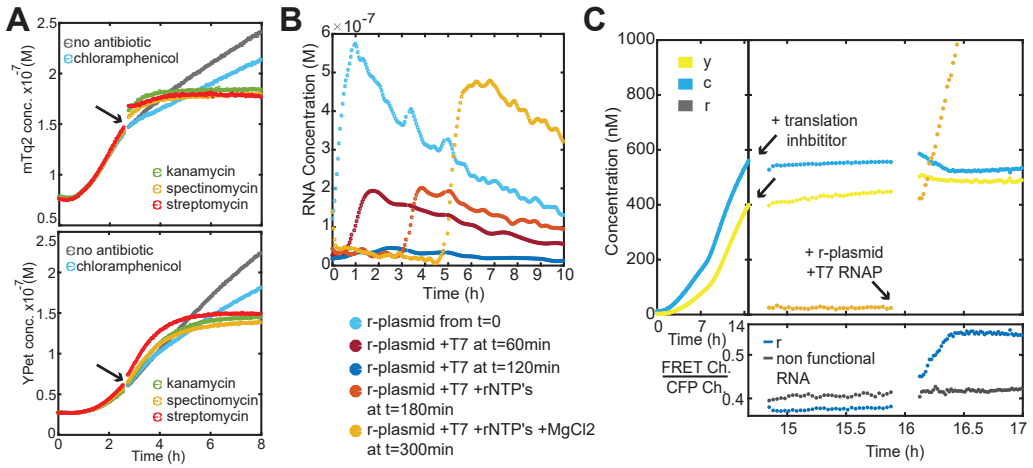

Figure 4.9: Optimization of expression and assembly of the RNA-protein nanostructure. **A** Testing the effect of different antibiotics on the expression of mTurquoise2-Tat (upper) and YPet-PCP (lower) by addition to the cell-free expression system after 150 minutes of expression (indicated by the arrow). The c-plasmid and y-plasmid both encode a chloramphenicol resistance. **B** Transcription of the RNA nanostructure from the r-plasmid in the cell-free expression system using T7 RNAP. If the plasmid is added 1 hour after the extract has been thawed and exposed to 29°C, the final expression level is reduced and after 2 hours there is no visible expression anymore. However, by adding purified T7 RNAP and rNTPs expression could be recovered again after 3h, even though not to the same level as expression at t=0. By supplementing purified T7 RNAP, rNTPs and $MgCl_2$ after 5 hours the expression level was recovered almost completely compared to t=0. **C** Fluorescence time traces reporting the expression of c-plasmid and y-plasmid (5 nM each) in the cell-free expression system. After 14 hours protein translation is stopped by the addition of spectinomycin and streptomycin. Approximately 45 minutes later, r is produced by the addition of the r-plasmid, T7 RNAP, rNTPs and $MgCl_2$. With the expression of RNA scaffold the donor signal decreases and the FRET signal increases, which indicates formation of the rcy_2 complex.

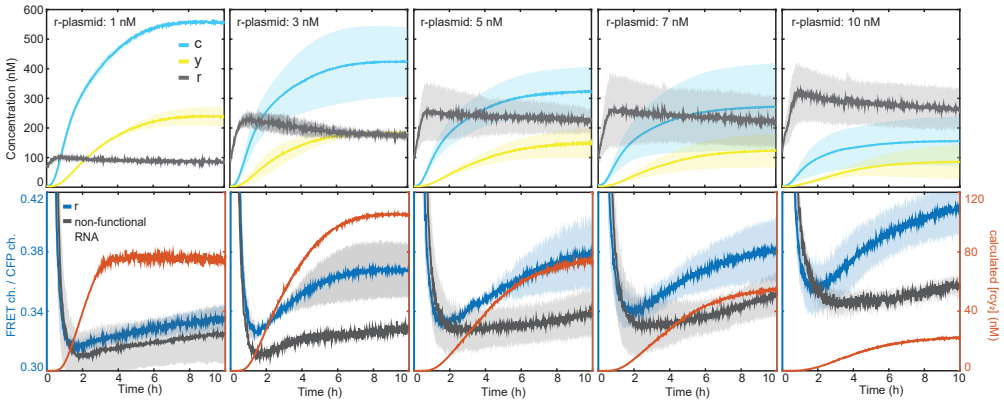

Figure 4.10: Upper plots: fluorescence time traces recorded during expression from cy-plasmid (5 nM) in the presence of varying concentrations of r-plasmid. Lower plots: Intensity of the FRET signal divided by the CFP signal for the sample with r-plasmid compared to a sample containing a plasmid with non-functional RNA (for all samples: mean and std. dev. from 3 repeats). The resulting concentration of $[rcy_2]$ complex calculated from the fluorescence values in the upper plot is shown in red.

teins (Fig. 4.8C). Furthermore, the strong expression of the RNA scaffold with T7 RNAP reduced the final concentrations of mTurquoise2-Tat and YPet-PP7 either below detection or below the levels required for maximal $f_{complex}$. In order to resolve this issue, the gene sequences for the two proteins were inserted into a single plasmid behind separate transcriptional promoters, which we found to increase the total amount of fluorescent proteins expressed by a factor of approximately 2 (Fig. 4.8C).

We then studied the expression and assembly of the RNA-protein nanostructure encoded on two plasmids – the cy-plasmid for the proteins and the r-plasmid for the RNA scaffold – in a one-pot reaction (Fig. 4.10). We screened the genetic template concentration of the r-plasmid for a constant cy-plasmid concentration to find the optimum relative expression levels. Based on the fluorescence levels of the single components we calculated that the largest concentration of complexes, nM, was expected to form for 5 nM of cy-plasmid and 3 nM of r-plasmid. Indeed, the background subtracted FRET signal obtained for this sample was found to be the highest. The results shown in Fig. 4.10 also demonstrate that the 'load' on the expression system caused by RNA expression is approximately equal to that by the expression of proteins, which justifies the assumption initially made for the definition of $f_{complex}$.

4.8 Co-expression and localization of RNA-protein complexes on streptavidin beads

We utilized the fourth aptamer function on the RNA structure by employing the streptavidin-binding aptamer for the localization of RNA-protein nanostructures onto streptavidin-coated beads. To this end, we encapsulated purified proteins together with streptavidin-coated beads and either the full RNA nanostructure or only the protein scaffolding TAR-PP7 aptamer part (i.e., without streptavidin aptamer) inside of water-in-oil emulsion droplets (Fig. 4.11A). As desired, co-localization of the proteins in the CFP, YFP and FRET channel on the microbeads is observed only in the presence of the streptavidin aptamer (Fig. 4.11B). We then co-encapsulated the r- and cy-plasmids, cell-free expression system and the streptavidin coated beads into water-in-oil emulsion droplets. Also in this case we were able to follow the expression and assembly of the RNA-protein nanostructure on the beads (Fig. 4.11C). Finally, we investigated the interaction of the RNA nanoscaffold with streptavidin alone.

4.9 Expression of the RNA-protein structure inside bacteria

Cell-extract based gene expression systems have been previously used successfully to prototype synthetic gene circuits, which were still functional when operated inside of bacterial cells [77]. We therefore transformed the two-plasmid system encoding the RNA-protein nanostructure into *E. coli* cells (strain BL21 STAR (DE3)), containing an IPTG inducible T7 RNA polymerase. We first verified that each component was expressed in bacteria (Fig. 4.12A). As shown in Fig. 4.12B, we were able to follow expression of the RNA nanostructure in bacteria by monitoring the fluorescence of the MG aptamer after induction with IPTG. Along with the production of the RNA structure, the relative fluorescence intensity of the FRET-acceptor (YPet-PCP) increased compared to the donor intensity (Fig. 4.13A and B), suggesting *in vivo* assembly of the peptide-tagged fluorescent proteins on the RNA scaffold. We found that the FRET signal observed *in vivo* was strongly reduced compared to the *in vitro* system, and the FRET effect also varied considerably between experiments. In contrast to the *in vitro* experiments, however, the plasmid copy numbers were not optimized. Furthermore, bacterial growth has a strong influence on the expression levels of the components of the nanostructure. We found that mTurquoise2-Tat expression had a negative

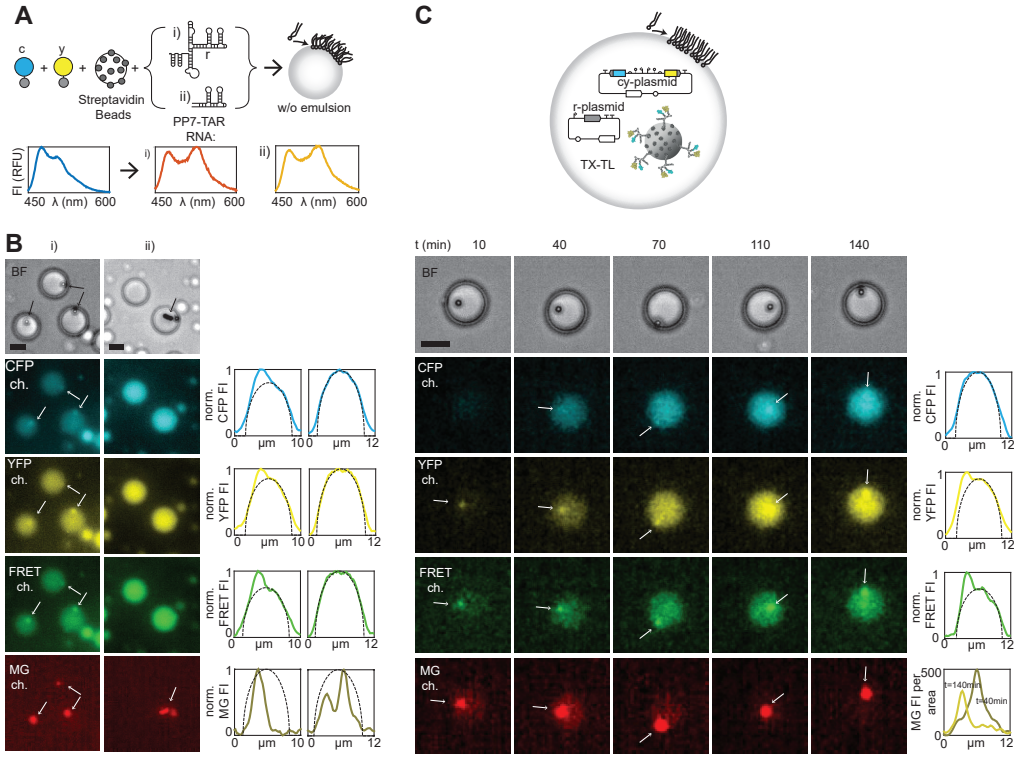

Figure 4.11: Expression, assembly and spatial localization in emulsion droplets. **A** Purified components c and y are co-encapsulated in water-in-oil emulsion droplets with streptavidin coated magnetic beads (SCMB) and i) the complete RNA scaffold r or ii) RNA with only the PP7-TAR aptamer (FRET-arm). Lower plots: emission spectra from the samples before encapsulation of the purified proteins ($[c]_0 = [y]_0 = 700$ nM), i) with complete RNA scaffold (500 nM) added or ii) with FRET-arm (500 nM) added. **B** Epi-fluorescence microscopy images of encapsulated samples from (A) in water-in-oil emulsion droplets in bright field and four different fluorescence channels CFP, YFP, FRET and MG. i) with the complete RNA scaffold an increased fluorescence and FRET signal is co-localized with the beads (arrows). ii) with the FRET-arm only there is no co-localization with the beads. The fluorescence in the MG channel indicates non-specific binding of MG to the beads. Scale bar: 5 μm. **C** Scheme of the multi-component nanostructure expressed from two plasmids within an emulsion droplet containing cell-free expression system and SCMBs. (D) Time series of epi-fluorescence microscopy images showing the expression of c, y and r and assembly of cry_2 on the SCMB (arrow) in bright field and CFP, YFP, FRET and MG channel. Scale bar: 5 μm. Some binding of the fluorescent proteins to the beads might be due to unspecific interactions from components in the cell-free expression system.

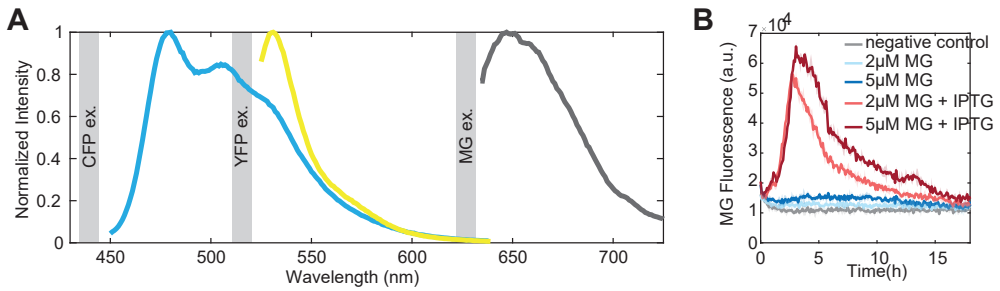

Figure 4.12: Expression of the RNA-protein nanostructure in bacterial cells. **A** Fluorescence emission spectra of mTurquoise2-Tat, YPet-PCP and the RNA nanostructure (MG aptamer) expressed in *E. coli* (BL21 STAR (DE3)) in M9 minimal medium at OD=0.5, excited at the indicated wavelengths. **B** Corresponding fluorescence time traces of the bacteria with the r-plasmid and without the r-plasmid (negative control, but with the cy-plasmid) (for all samples: mean and std. dev. from 3 repeats).

influence on cell growth. Removing the self-cleaving ribozyme PlmJ in front of the coding region improved the expression of mTurquoise2-Tat. This also indicates that optimized conditions for *in vitro* assembly of an RNA-protein nanostructure cannot be directly transferred to the *in vivo* case, for which further adjustments would be required.

4.10 Conclusion

Our results demonstrate that a multi-functional RNA-protein nanostructure can be expressed, assembled and spatially organized in a one-pot cell-free reaction, which is a crucial step for the implementation of genetically encoded multi-component nanostructures in artificial cellular systems. Our study highlights several of the challenges which have to be addressed for efficient production and assembly of such structures such as binding equilibria, stoichiometry, chemical stability, expression strength and competition for resources. We show that a cell-free expression system facilitates the quantitative study of these aspects under well-controlled conditions, which is essential for the development and optimization of *in vitro* expressed nanostructures, as well as for prototyping structures for potential applications in living cells.

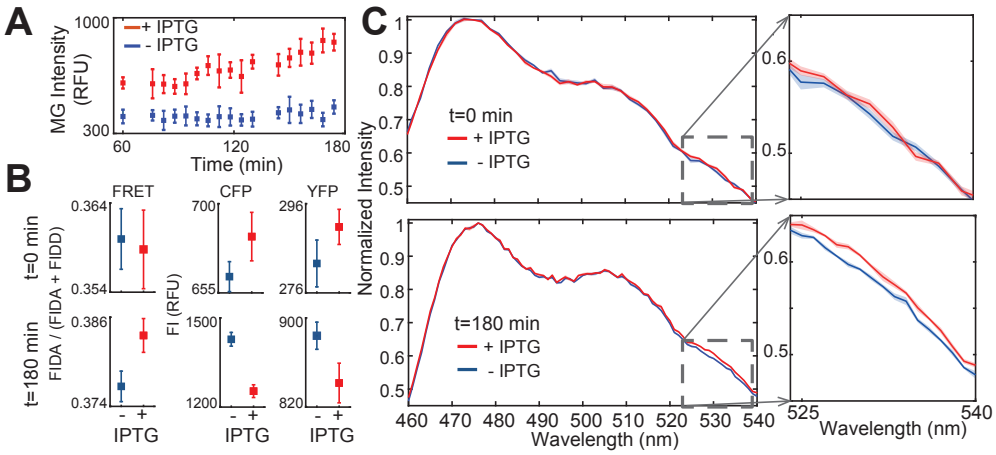

Figure 4.13: **A** Fluorescence intensity time traces of the transcription of the RNA nanostructure in bacteria reported by MG with and without the induction of T7 RNAP via IPTG at t=0 min in M9 minimal medium at OD=0.5. Mean and SD from min. 4 technical repeats are shown. **B** Upper row: FRET output, absolute value of the mean fluorescence in the CFP channel (λ_{ex}=440 nm, λ_{em}=474 nm) and YFP channel (λ_{ex}=500 nm, λ_{em}=530 nm) for the bacteria without and with IPTG at the time of induction (t=0 min). Lower row: FRET output and absolute value of the mean fluorescence in the CFP and YFP channel for bacteria without and with IPTG after t=180 min. The mean and 95% confidence interval for min. 4 technical repeats are shown. **C** Fluorescence emission spectra of the fluorescent proteins (c and y) at t=0 min (top panel) without and with IPTG for induction of the RNA nanostructure and after t=180 min (bottom panel). The spectra are normalized to the intensity of the FRET-donor protein (λ=474 nm. Mean (solid line) and s.e.m. (shaded area) for min. 4 technical repeats are shown.

4.11 Materials and Methods

4.11.1 In vitro transcription

Linear DNA templates were purchased from IDT (USA) and amplified with primers from Biomers (Germany) by a polymerase chain reaction (PCR) using Phusion®High-Fidelity PCR Master Mix (NEB, USA) followed by QIAquick PCR Purification kit (Qiagen, USA). A typical transcription reaction has a volume of 100 μl, is kept at 37°C overnight and contains: 100 nM template DNA, 4 mM of each rNTP (Epicentre, Singapore), 14 mM MgCl2 (Sigma Aldrich, USA), 1 x RNAPol Reaction Buffer (NEB, USA),1 U/μl RNase Inhibitor (NEB) and 2 U/μl T7 RNA Polymerase (NEB, USA). In order to digest the DNA template after transcription, the reaction was sup-

plemented with 1 x DNase I reaction buffer (NEB, USA) and 1 U/μl DNase I (NEB, USA) for 30 minutes.

4.11.2 RNA purification and quantification

The RNA was extracted from an in vitro transcription mix using Phenol Chloroform and 5PRIME Phase Lock GelTM tubes (Quantabio, USA) for improved phase separation, followed by ethanol precipitation. RNA was stored in nuclease free water at -20°C for about one month. RNA concentrations were estimated by denaturing polyacrylamide gel electrophoresis (PAGE). 2 μl of purified RNA in loading gel buffer (0.25% bromophenol blue (BPB), 0.25% xylene cyanol (XC), 30% glycerol) was loaded on a 10% (w/v) of 27.5:1 acrylamide/bisacrylamide gel with 8M urea in 1 x TBE buffer for 90 minutes with 9 V/cm at 25°C. The gel was stained with SYBR Green II (Invitrogen, S7564, USA) and the concentration was determined by comparing the fluorescence intensity of the relevant band to the standard molecular weight ladder Low Range Riboruler (Bio-Rad, USA) with ImageJ.

4.11.3 Protein purification

The bacterial expression vector (pJ431) for YPet-PCP and mTurquoise2-Tat provides a T7 promoter, a His-tag (6xHis) and Kanamycin resistance (DNA 2.0, USA). The proteins were expressed in E. coli strain BL21(DE3) pLysS (Promega, USA) in 25 ml LB medium and induced with 1 mM IPTG at OD=0.6. mTurquoise2-Tat was expressed at 37°C for 2 hours and YPet-PCP at 20°C for 6 hours. Cells were pelleted and lysed by bead beating in a Minilys device (Peqlab, Germany). Here 0.25 %(w/v) of 0.1 mm glass beads are added to the cells which are then shaken at 5000 rpm for 30 sec and relaxed on ice for 30 sec five times in a row. Then both proteins were purified by gravity-flow chromatography with Ni-NTA Agarose Beads (Qiagen, USA). After purification, the proteins were dialyzed overnight at room temperature using a dialysis cassette (MWCO 7 kDa, Thermo Fisher Scientific, USA) in RNA binding buffer (10mM HEPES/KOH (pH 7.5), 100 mM KCl and 1 mM $MgCl_2$). mTurquoise2-Tat was stored at 4°C while YPet-PCP was stored at room temperature.

4.11.4 Electrophoretic mobility shift assay (EMSA)

For the EMSA in Figure 2B, a gradient polyacrylamide gel 4-20 % (Mini-PROTEAN ®TGX Precast Protein Gel, #4561096, Bio-Rad, USA) was loaded with purified RNA (100 ng), mTurquoise2-Tat (5x molar excess to

RNA) and YPet-PCP (10x molar excess to RNA) were incubated in RNA binding buffer supplemented with 2 U/μl Murine RNase Inhibitor (NEB) for about 10 minutes at room temperature. The running buffer consisted of 25mM Tris, 192 mM glycine, pH 8.3, 10 mM KCl and 2 mM $MgCl_2$ and 1x Purple (no SDS) Gel Loading Dye (NEB, USA) was used. The gel was run at 200 V for 35 minutes. Gels were first scanned on a Typhoon FLA 9500 Gel Laser Scanner (General Electrics, USA) to detect the proteins fluorescence, then stained either with SYBR Green II (Invitrogen, S7564, USA) or 1 mM Malachite Green (Sigma Aldrich, USA) and scanned again. For the EMSA in Figure S2, 4 % (w/v) agarose gel was used with 7 V/cm for 135 min. The running buffer contained 2 mM $MgCl_2$ and the loading buffer consisted of 7 % (w/v) sucrose and 4 mM Bromphenol Blue. The first lane was loaded with 80 ng Low Molecular Weight DNA Ladder (NEB, USA).

4.11.5 Plasmid construction and purification

The genes for mTurquoise2-Tat and YPet-PCP were ordered as linear fragments (IDT DNA) and cloned into the BioBrick backbones pSB1C3, and the RNA nanostructure similarly cloned into pSB4K5 using standard restriction and ligation protocols (NEB, USA). The DNA sequences used can be found in the appendix. Fluorescence Spectra. Fluorescence experiments with purified RNA and fluorescent proteins were performed on a Horiba/Jobin Yvon Fluorolog 3 system in 45 μl quartz glass cuvettes in 1x RNA binding buffer. Conversion from fluorescence intensity to concentration was done using the following relationship

$$F(\lambda = 474) = m_{474} c_{tot} \tag{4.17}$$

where c_{tot} is the concentration of the fluorescence protein and m_{474} is the slope of the linear fit to the calibration data at excitation wavelength 440 nm and emission wavelength 474 nm (Fig. 4.14). For the degradation of the RNA in the presence of the fluorescent proteins 1.5 U/μl RNase 1f (NEB, USA) were added to the cuvette and the spectrum was taken about 10 min after incubation at 37°C. For the protease control, 0.01 U/μl of Proteinase K (NEB, USA) was added to the sample with 10 min incubation at 37°C.

4.11.6 Cell-free expression system

The cell-free gene expression system was prepared according to the protocol provided by Sun *et al.* [68]. Briefly, crude S30 cell extract was obtained by bead beating a BL21- Rosetta2(DE3) mid-log phase culture with 0.1 mm

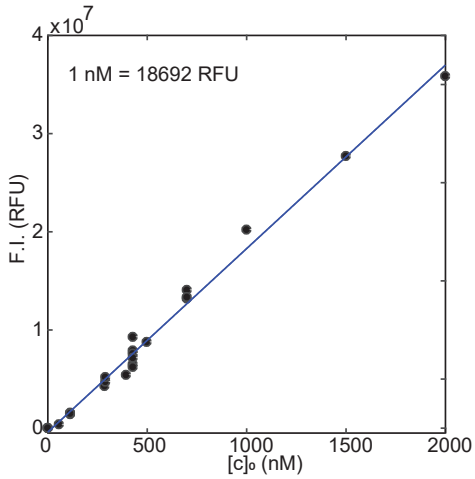

Figure 4.14: Calibration curve relating the concentration of purified mTurquoise2-Tat (c) to the fluorescence signal in the fluorescence spectrometer.

glass beads in a Minilys device (Peqlab, Germany). After centrifugation, the supernatant was incubated at 37°C for 80 minutes and again centrifuged in order to remove endogenous DNA. The supernatant was dialyzed for four hours against a buffer (pH 8.2, 2mM DTT, 14 mM Mg-glutamate, 60 mM K-glutamate, 5 mM Tris) in a cassette with a molecular weight cut-off of 10kDa. Using a Bradford assay, the protein concentration was determined to be 30 mg/ml. For the reaction buffer we used phosphoenolpyruvate (PEP) as an energy source instead of 3-phosphoglyceric acid (3-PGA) [213]. The reaction buffer thus contains 50 mM Hepes pH 8, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/ml tRNA, 26 mM coenzyme A, 0.33 mM NAD, 0.75 mM cAMP, 68 mM folinic acid, 1 mM spermidine, 30 mM PEP, 1 mM DTT, 2 % PEG-8000. All components were flash-frozen in liquid nitrogen and stored at -80° before thawing on ice. The composition of a single cell-free reaction was: 33 % (v/v) S30 cell extract mixed with 42 % (v/v) buffer and 25 % (v/v) DNA plus additives. The plasmids were purified using a Midi-Prep kit (Macherey-Nagel, Germany).

4.11.7 Fluorescence intensity time traces

To follow the expression in the cell-free expression system we loaded 15 μl of our samples in a 384-well plate (IBIDI, Germany) into a plate reader (ClarioStar, BMG Labtech) at 29°C. To stop the translation, 50 μg/ml of Spectinomycin (Sigma-Aldrich, USA) and 10 μg/ml of Streptomycin sulfate (Sigma-Aldrich, USA) were added after 14 hours to the sample. For the other

antibiotics, we supplemented the cell-free expression system with 20 μg/ml Chloramphenicol (Sigma-Aldrich, USA) and 50 μg/ml Kanamycin (Sigma-Aldrich, USA). In order to start a transcription reaction on top of finished cell-free expression reaction, we added 5 nM of plasmid, 200 nM of T7 RNAP (purified), 2 mM of each rNTP (Epicentre, USA) and 10 mM $MgCl_2$.

4.11.8 Encapsulation study using streptavidin coated beads

The samples containing purified components consisted of $[c]_0 = 700$ nM and $[y]_0 = 500$ nM and either RNA nanostructure (r) or the FRET-arm RNA in $1x$RNA binding buffer and 5 μM Malachite Green. The samples containing the cell-free expression system consisted of 5 nM cy-plasmid, 3 nM r-plasmid, 250 nM T7 RNAP (produced and purified in the lab), 5 μM Malachite Green (dissolved in DMSO) and 75 % (v/v) of the cell-free expression system. Both samples were then supplemented with micrometer-sized streptavidin coated magnetic beads (DynabeadsTM MyOneTM Streptavidin C1, Thermo Fisher Scientific, USA). 5 μl of the sample were then mixed with 45 μl of FC-40 Fluorinated oil (F9755, Sigma-Aldrich, USA) containing 2 % (w/w) E2k0660 non-ionic surfactant (RainDance Technologies, USA) in Protein LoBind tubes (Eppendorf, Germany). In order to create emulsion droplets, the sample was vortexed for 30 seconds (with 10 second intervals on ice). The droplets were then transferred to an observation chamber consisting of two melted Parafilm ®(Brand, Germany) strips between two microscopy cover slides using a pipette. The chamber was sealed with vacuum grease and nail polisher and placed in an inverse fluorescence microscope (Nikon TI-2). Time-lapse microscopy images for bright field and fluorescence (SOLA, Lumencor, USA) were performed at T=29°C using an incubator cage around the microscope (Okolab, Italy) and analyzed with ImageJ. The fluorescence intensity profiles were measured with ImageJ, normalized and fitted to the equation $FI = \frac{1}{r}\sqrt{r_s^2 - x^2}$ with MATLAB 2017a, which is the expected normalized intensity profile of a dye homogeneously distributed within a sphere of radius r_s.

4.11.9 TEM and AFM imaging

Purified RNA nanostructures (300 nM) diluted in Tris-HCl buffer (40 mM pH 7.2) with 10 mM MgCl2 and 100 mM KCl were applied to glow discharged formvar-supported carbon-coated Cu400 TEM grids (Science Service Munich, Germany), using a standard negative staining protocol. For the images with streptavidin (N7021S, NEB, USA), mTurquoise2-Tat and YPet-PCP, 300 nM

of each purified protein was added to the sample. AFM imaging was done on an AFM (Asylum Research Cypher, UK) and the RNA nanostructure was diluted in 15 mM Tris-HCl (pH 7,2) buffer containing 10 mM MgCl2 and 50 mM KCl. The samples were placed on a specially modified mica surfaces (APS mica) [257] and imaged in air in tapping mode.

4.11.10 Bacterial cultures

All experiments were performed in *E. coli* strain BL21 StarTM (DE3) (Thermo Fisher Scientific, USA). The r-plasmid and cy-plasmid were transformed into electrocompetent cells, grown in SOC-Medium for two hours and plated on LB-Agar plates containing Kanamycin (50 μg/ml) for the r-plasmid and Chloramphenicol (25 μg/ml) for the cy-plasmid overnight at 37°C. Single colonies were sequenced (Sanger sequencing, GATC, Germany) before and after they were used for glycerol stocks (25 % w/v glycerol) in LB medium at -80°C. Overnight cultures were grown from glycerol stocks in M9 minimal medium supplemented with 20 mM glucose and 300 μg/ml thiamine hydrochloride (Sigma-Alrich, #$T1270$) and Kanamycin (25 μg/ml) for the r-plasmid and Chloramphenicol (15 μg/ml) for the cy-plasmid. The next morning the cells were diluted and re-grown to an OD of 0.5 and then again diluted to an OD between 0.2-0.4 and 300 μl per well were added in a 96-well plate (IBIDI, Germany), sealed with sealing tape (Bio-Rad #2239444, USA) and placed in the plate reader (ClarioStar, BMG Labtech, Germany) at 37°C.

4.12 Contributions

The idea for a RNA-mediated FRET pair assembly was brought up by Fabio Chizzolini (F.C.) and Sheref Mansy (S.M.) and was further developed by Matthaeus Schwarz-Schilling (M.S.) and Friedrich C. Simmel (F.S.). M.S. and F.C. designed the RNA nanostructure and performed preliminary experiments to express and assemble the RNA-protein structure. Guido Grossi and M.S. designed and purified the fluorescent proteins and performed initial binding assays. Swati Krishnan acquired the transmission electron microscopy (TEM) images and Ali Aghebat Rafat performed measurements on the AFM. Analysis of TEM images was done by M.S. Fabrication of genetic constructs was done by Aurore Dupin (A.D.) and M.S. Experiments on chemical stability of the RNA in the cell-free expression system were designed and performed by A.D. and M.S. Preparation of the cell-free expression system was done by A.D. and M.S. Optimization of the expression and assembly in

the cell-free expression system was performed by M.S. Experiments on the expression of the RNA-protein structure in bacteria were done by M.S.

Chapter 5

CRISPR interference: filamentation and restoration of normal growth in bacteria

5.1 Summary

The contents of this chapter have been published in the article: "Filamentation and restoration of normal growth in Escherichia coli using a combined CRISPRi sgRNA/antisense RNA approach" by Andrea Mückl, Matthaeus Schwarz-Schilling, Katrin Fischer, Friedrich C. Simmel in *PLOS ONE*, 2018, 13(9): e0198058.

Gene regulation is a crucial tool to enable cells to respond to changing external conditions. Synthetic biologists use natural gene regulation elements to build synthetic genetic constructs in addition to new molecular tools created by design. CRISPR interference (introduced in section 2.6) is a new tool to control the expression of genes by either blocking transcription elongation or inhibiting transcription initiation.

In contrast to other transcription factors, which bind directly to operator sites with a sequence specific to that protein. The binding of dCas9 only requires the sequence "GGN" adjacent to the target site. This sequence is called the proto-adjacent motif (PAM) and can vary from different Cas proteins that bind to DNA. The rest of the binding site is determined by the sequence of the sgRNA molecule that is bound to dCas9.

This mechanism opens up the possibility to knockdown genes on the chromosome of bacteria without having to change the DNA sequence of the chromosome itself. Any sequence next to a PAM can be targeted.

It has been shown that a knockdown of the *ftsZ* gene results in an arrest of

cell division in *E. coli* but not of cell growth. The chromosome still replicates and while the width of the bacteria does not change largely, the length can increase significantly due to the absence of division events [258]. By having an inducible version of *ftsZ* in *E. coli*, Sanchez-Gorostiaga *et al.* were able to show that a recovery of FtsZ levels can rescue bacterial division [259]. Without restoration of FtsZ the filamentous bacteria will lyse. This target opens the possibility to study the reversibility of CRISPRi on bacterial genes.

Hence we used an inducible CRISPRi construct to target the *ftsZ* gene. Three different sgRNAs were targeted at the template strand of three different promoters of *ftsZ*. The expression of the sgRNAs was controlled via the induction of T7 RNAP and a pLysS plasmid. There were, however, no transcriptional terminator between the different sgRNA sequences.

When the CRISPRi plasmid was first cloned in the bacteria, they grew filamentous even without induction of the CRISPRi. This problem was solved by introducing decoy-binding sites for the sgRNA-dCas9 complex on a high copy plasmid. Stopping cell division was then achieved by inducing both sgRNA and dCas9, and resulted in filamentous bacteria.

The reversibility of CRISPRi was tested, first by introducing an antisense RNA that is complementary to the sgRNA. The approach has been shown to work in bacteria for fluorescent proteins [141]. We first tested the inhibitory effect of anti-sgRNA on the CRISPRi process in a cell-free expression system with the expression of a fluorescent protein (mVenus). The non-template strand within the coding region was targeted. Here dCas9, sgRNA, anti-sgRNA and the genetic template for mVenus were added to the cell-free expression reaction as purified components. The ratio of anti-sgRNA to sgRNA added to the reaction was 5:1 and resulted in full inhibition of the sgRNA-dCas9 assembly. However, when the anti-sgRNA was added 15 minutes or 30 minutes later than the rest of the components, the de-repression was significantly reduced.

In bacteria, the anti-sgRNA, which was inducible by the quorum-sensing molecule AHL, was placed on a plasmid with a high copy number. We explored whether filamentous cells can be switched back to normal growth by inducing the anti-sgRNA while having the CRISPRi induced. In microfluidic chambers which allow to follow individual cells, we found that cells did not revert to normal cell growth but lysed. It was possible to revert to normal cell growth only by transiently inducing the filamentation (for about 2 hours), and then removing the CRISPRi inducers. About 1-3 % of the analyzed cells reverted to normal growth with a delay of about 5 hours after the CRISPRi inducers were removed. The addition of AHL (which induces the anti-sgRNA) after the removal of the CRISPRi inducers allowed the bacteria to revert to normal growth slightly faster. With filamentation, heterogene-

ity between cells in regard to cell length and expression of two fluorescent reporter proteins increased.

In addition, the sensitivity to the inducer concentration of CRISPRi varied between two different glycerol stocks. The glycerol stock of the bacteria, which used for all the results in this chapter unless stated otherwise, was sensitive to low CRISPRi inducer concentration. The other glycerol stock was not as sensitive, meaning that low CRISPRi induction did not completely arrest cell division for all cells.

5.2 Introduction

Before an *E. coli* cell divides into two identical daughter cells, proteins of the cell division machinery accumulate at its center with an accuracy of about 2 % and form the so-called 'Z ring' [260].The Z ring serves as a scaffold at the future division site for the other over 20 known proteins that constitute the divisome [261]. The Z ring itself is assembled from at least six proteins, including the filament forming protein FtsZ and its membrane anchoring proteins FtsA and ZipA [262] [263] [264] [265]. The cellular content of FtsZ needs to be regulated, as under- or over-expression leads to filamentous bacteria or minicells without a genome [266] [267]. Moreover, an imbalance between FtsZ and FtsA results in cell division arrest and bacterial filaments that contain multiple copies of the bacterial chromosome [258] [268]. Since FtsZ is one of the earliest proteins to initiate the assembly of the division machinery, mutations or knock-downs of FtsZ result in filaments with stalled constriction sites or partially divided regions [269]. One of the earliest studied strains with a mutation in FtsZ, *ftsZ84*, displays filamentous growth at 42°C and can be switched back to normal growth at 30°C [270] [271]. Aborted constriction sites do not seem to be continued after temporary upshifts from 30°C to 42°C for 2 minutes, but newly formed division sites are used [272]. However, the *ftsZ84* mutant lyses or loses its viability after about 3 hours after the temperature shift.

Recently, Sánchez-Gorostiaga *et al.* studied the response of *E. coli* when FtsZ falls below a critical level, followed by its restoration using an IPTG inducible promoter in front of the *ftsZ* gene [259]. They show that besides forming filaments, FtsZ deprived cells are more prone to improper chromosome segregation, show global changes in transcription levels and lose integrity of their membrane. A convenient technique for the regulation of chromosomal gene expression has been recently provided through CRISPR interference (CRISPRi) using dCas9, a non-cleaving mutant of the CRISPR associated nuclease Cas9 [53]. The protein can be directed to any position

on the chromosome by a single guide RNA (sgRNA) molecule, provided that the target sequence neighbors a protospacer adjacent motif (PAM) with the canonical sequence NGG. Depending on the binding site, the dCas9-sgRNA complex can repress transcription by either preventing transcription initiation by RNA polymerase or by acting as a roadblock for transcriptional elongation.

Elhadi *et al.* have recently used the CRISPRi mechanism to change the morphology of *E. coli* by targeting the *ftsZ* gene and the *mreB* gene, a gene found to control the width of the cells [138] [273]. Focusing on microbial bioproduction of plastics, they found that the morphologically altered cells provide a larger volume for the accumulation of intracellular polyhydroxybutyrate (PHB) inclusion bodies. However, the CRISPRi mediated gene knockdown had a negative impact on bacterial growth rate.

In order establish a reversible knockdown system based on CRISPRi, Lee *et al.* utilized antisense RNA to target the sgRNA [141]. They were able to show that the expression of fluorescent proteins could be successfully suppressed and reactivated in bacterial cells, with each step having a response time of ≈ 3 hours. Using RNA instead of proteins to regulate gene expression can have several advantages. They are straightforward to design since RNA-RNA interactions can often be reduced to base-pair interactions and their secondary structure can be predicted with software tools [274] [275]. Moreover, RNA is expressed faster than proteins and requires less of the cell's resources [77].

In the present work, we reversibly induce filamentation in *E. coli* by targeting FtsZ using the CRISPR/dCas9 interference mechanism combined with an antisense RNA strategy. We first switch bacteria into the filamentous cell growth mode and subsequently reverse this process, such that the cells return into a normal growth phenotype. We quantify the switching process and dynamically control the system by the inducible expression of antisense RNAs that are complementary to the sgRNAs engaged in *ftsZ* knockdown. We use single-cell fluorescence microscopy experiments from which we derive bacterial length distributions [276], growth and division rates, as well as reporter gene expression levels. This allows us to identify factors that affect the switching process. In particular, we find that bacteria strongly respond to the CRISPRi knockdown of ftsZ, but only few bacteria revert to normal cell division several hours after termination of CRISPRi induction. Antisense-sgRNA expression supported the recovery process and facilitated considerably faster switching than in the absence of the antisense RNAs. Antisense-sgRNA thus provides a relatively straightforward means to control and adjust the kinetics of CRISPRi de-repression, which is of great interest for the realization of gene circuitry, in which cellular processes have to be dynamically switched

on or off in response to endogenous signals or external cues.

5.3 Results

5.3.1 Experimental design with decoy-binding sites

In order to reversibly switch bacteria to filamentous growth, we disturbed the FtsZ/FtsA ratio through CRISPR interference using appropriate sgRNAs (Fig. 5.1A). Three of at least six known promoters for *ftsZ* (ftsZ2p, ftsZ3p, ftsZ4p) lie within the *ftsA* coding region [277]. We targeted the template strand of these three promoters for *ftsZ* transcriptional initiation blockage. In contrast to tunable CRISPRi (tCRISPRi) with an inducible chromosome-integrated dCas9 [140], we here use plasmid-encoded inducible dCas9 and sgRNA. The sgRNAs are induced by isopropyl β-D-thiogalactopyranoside (IPTG) via T7 RNA polymerase and were encoded on the 'CRISPRi plasmid' together with TetR-controlled dCas9 and mVenus reporter protein.

We then restored cell division by removing CRISPRi inducers aTc and IPTG and thus turning off dCas9 and sgRNA expression or by removing CRISPRi inducers and adding anti-sgRNA inducer (the 'active' approach). The anti-sgRNAs were designed to absorb the sgRNAs via duplex formation (Fig. 5.1A) in a similar manner as previously demonstrated [141]. In our approach, the anti-sgRNAs are transcribed from the 'anti-sgRNA plasmid' and induced by AHL (acyl homoserine lactone (HSL) 3-oxo-C6-HSL) from pLux promoters.

Transformation of bacteria with the CRISPRi plasmid resulted in filamentous cells even without induction of the expression of dCas9 and sgRNA. In order to create a threshold for dCas9-sgRNA below which *ftsZ* is not regulated down, we introduced decoy-binding sites ('sponges') for the dCas9-sgRNA complex. However, the addition of sponges to the CRISPRi plasmid did not stop filamentation in the absence of inducers (Fig. 5.2). Only after the introduction of additional sponge elements on the high copy number plasmid (about 500-700 per cell [13]) that encodes the anti-sgRNA, the bacteria exhibited the intended normal growth morphology.

Expression of dCas9 in the absence of sgRNA or with sgRNAs of different sequence does not lead to filamentous growth under our experimental conditions (5.2C). This can be important, since changes in cell morphology were reported previously for high level expression of the dCas9 [142].

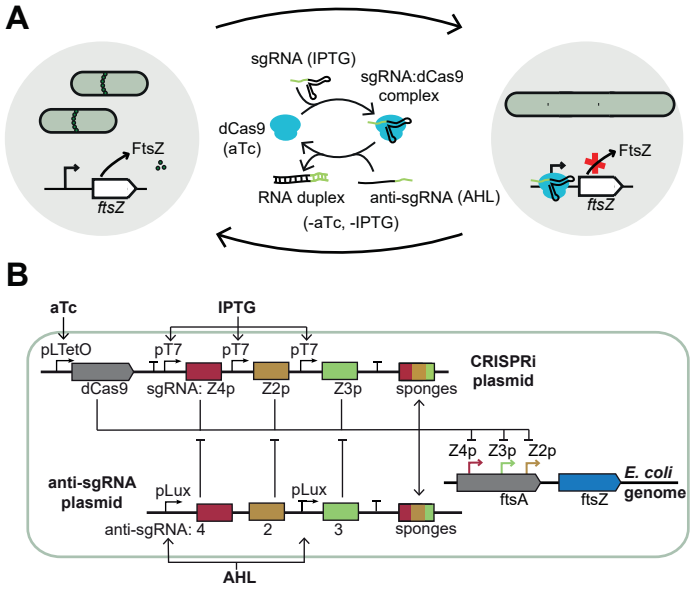

Figure 5.1: **A** Schematic representation of E. coli switching into filamentous growth after induction of single guide RNA (sgRNA) via IPTG and dCas9 via aTc. The dCas9-sgRNA complex blocks the expression of FtsZ, stopping the formation of the septal ring that is essential for cell division in E. coli. Cell division can be rescued by inducing appropriate antisense sgRNAs ('anti-sgRNA') with AHL and by removing the inducers for the dCas9 and sgRNA. **B** Details of the genetic constructs involved: the CRISPRi plasmid codes for dCas9 under aTc-inducible promoters and three different sgRNAs under T7 promoters which target three different promoters of the *ftsZ* gene on the genome of the *E. coli*. T7 RNA polymerase is inducible with IPTG. The anti-sgRNA plasmid codes for anti-sgRNAs under the control of an AHL-inducible promoter. The sponge elements on the plasmids act as decoy binding sites for the corresponding dCas9-sgRNA complexes.

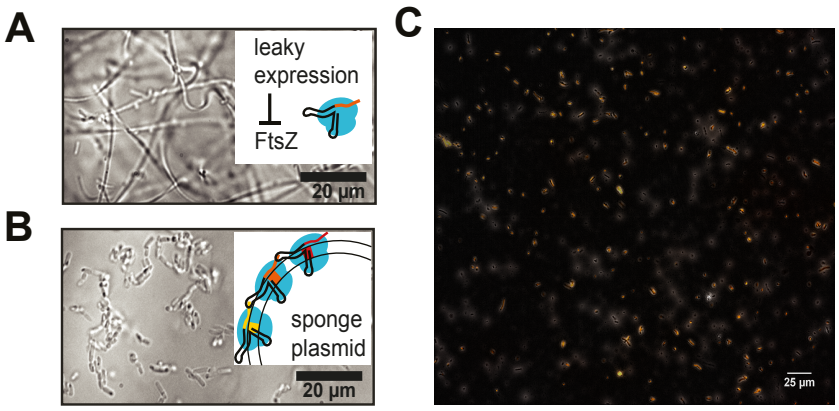

Figure 5.2: **A** Filamentation control tests. First, we tested the decoy-binding site strategy under the non-induced state. The liquid culture bacteria contained the CRISPRi-plasmid without **A** and with **B** a high copy number sponge plasmid. Only with additional sponge elements, the cells grow and divide. **C** In we tested dCas9 influence on cell morphology. The cells with CRISPRi and anti-sgRNA plasmids grow normal under dCas9 induction (107 nM aTc) for 3 hours. The corresponding reporter gene mVenus is also expressed from a tet-promoter as well as constitutive mRFP. The image is an overlay of phase contrast, mRFP and mVenus fluorescent channels.

5.3.2 Reversal of CRISPR interference using anti-sense RNA

To test the effect of anti-sgRNA on CRISPRi, we initially performed a series of *in vitro* experiments in homemade bacterial cell extract [68]. First, we tested two different designs for the anti-sgRNAs, which differed in their lengths and bound to different complementary regions of the sgRNA, one creating a 42 bp and the other a 56 bp duplex with the sgRNA (Fig. 5.3A). We regulated the expression of mVenus using purified dCas9 together with sgRNA and anti-sgRNA. As shown in Fig. 5.3B, mVenus expression is efficiently suppressed in the presence of dCas9 and sgRNA, while the addition of the anti-sgRNA variants inhibits the repression of mVenus expression (cf. Appendix for the corresponding sequences). Our experiments confirmed that sequestration of the spacer region and only part of the dCas9 handle by the short anti-sgRNA is sufficient to de-activate the CRISPRi mechanism (Fig. 5.3B). In a second experiment, we delayed the addition of anti-sgRNA relative to dCas9-sgRNA. We found that mVenus fluorescence is recovered compared to the knockdown case, however the recovered level is already halved for a delay time of 15 minutes (Fig 5.3C).

5.3.3 Single-cell analysis of filamentation

After induction of the CRISPRi mechanism with aTc and IPTG, bacterial cells rapidly stopped division and started the expression of mVenus (which also was under the control of a pLTetO promoter). The cell length distribution of *ftsZ*-knockdown bacteria broadens and shifts towards greater lengths (Fig. 5.4). The mean cell length increases from $< L >= 3$ μm to $< L >= 21$ μm in three hours with 500 μM IPTG and 107 nM aTc (which we defined as the 100 % induction level).

5.3.4 Efficiency of restoration of normal cell division

We performed time-lapse video microscopy studies (see Sec. 5.5) to observe the growth and fluorescence of individual filamentous bacteria over time. Microfluidic trap chambers were connected to fresh medium supply channels (Fig. 5.5A). Filamentous growth of *E. coli* proceeded up to 10 hours after which all bacteria burst [97].

We investigated the process of reverting filamentous cells back to normal growth by stopping the production of dCas9-sgRNAs via removal of CRISPRi inducers aTc and IPTG (passive switching) or by removal of CRISPRi inducers and addition of anti-sgRNA inducer (the active approach). Inducing the

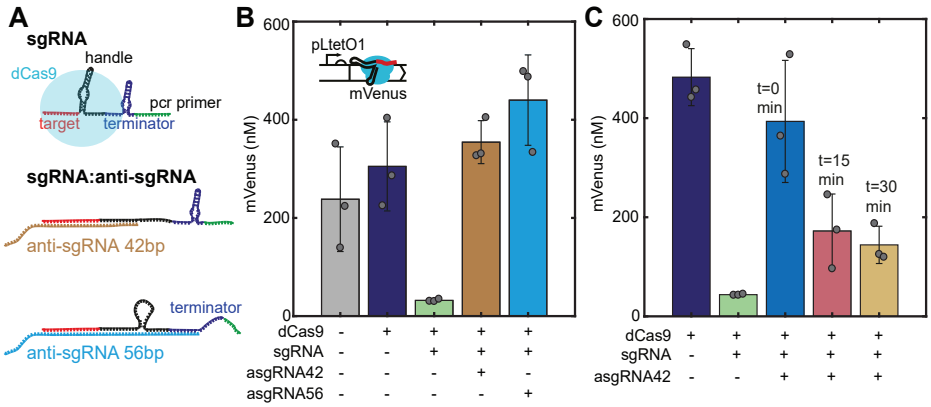

Figure 5.3: Experimental validation of the anti-sgRNA strategy. (A) Secondary structures of free and complexed sgRNA and anti-sgRNA variants. The dCas9 handle is destroyed by duplex formation, which prevents dCas9 from binding. (B) Prototyping of the CRISPRi knockdown and rescue system in a cell-free gene expression system using mVenus as a fluorescent reporter protein. The restoration of mVenus expression is shown with truncated anti-sgRNA versions (the number indicates the number of base pairs in the resulting sgRNA:anti-sgRNA duplex). The expression of mVenus ([template DNA] = 5 nM) is blocked by the supplementation of purified dCas9 (70 nM) and sgRNA (100 nM) and is re-activated upon addition of anti-sgRNA (0.5 μM). Fluorescence levels for three different samples are taken at t = 15.5 hours or t = 11 h. Error bars are plotted as SD from 3 individual replicates. (C) Delaying the time of anti-sgRNA addition relative to dCas9-sgRNA results in lower mVenus fluorescence intensities. (taken at t = 12 hours). Error bars are plotted as SD from 3 replicates (c(sgRNA) = 250 nM, c(anti-sgRNA) = 1 μM).

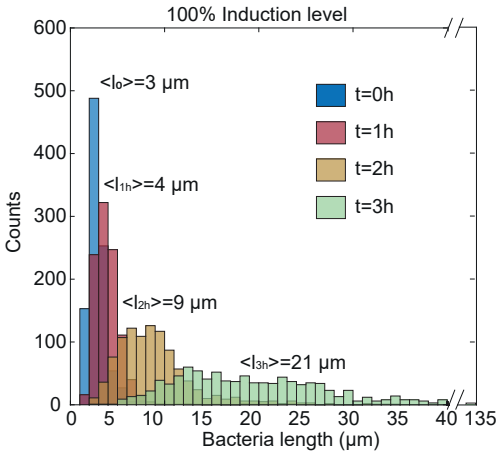

Figure 5.4: Filamentous bacteria. Histogram of the cell length (calculated from 1000 cells each) at different time points at 100 % induction level (corresponding to 500 μM IPTG and 107 nM aTc). Over time the distribution broadens and the whole population shifts to larger lengths.

anti-sgRNA with AHL in the presence of CRISPRi inducers did not result in arrest of filamentation in microfludic cell trap experiments.

The removal of the CRISPRi inducers – both with or without AHL – allowed a fraction of about 1.5-3 % of all analyzed cells (about 2000 for each condition) to resume normal cell division (Fig. 5.6). However, in the presence of AHL some cells started to divide again after filamentous growth faster than in the absence of AHL. Furthermore, full induction of the CRISPRi compared to 43 % induction level for about 2 hours delayed the timing of the first division event even in the presence of AHL.

Our experiments in microfluidic cell traps revealed a considerable phenotypic heterogeneity upon induction of filamentation, which was reversed once normal growth was restored (Fig. 5.7A and B). In addition to their size, we characterized the cells with respect to their fluorescence intensity. After 2 hours of continuous induction with IPTG/aTc, the coefficient of variation (CV=standard deviation/mean) of the cell size as well as of mVenus and mRFP expression increased considerably (Fig. 5.7B). In contrast to aTc-induced mVenus (on the CRISPRi plasmid), the constitutively expressed mRFP (on the anti-sgRNA plasmid) level dropped during filamentation (Fig. 5.7C).

We characterized the bacteria from experiments using the active approach to revert to normal growth in respect to the fluorescence of the reporter proteins.

First, the mRFP expression of the cells can be fitted by a Gaussian dis-

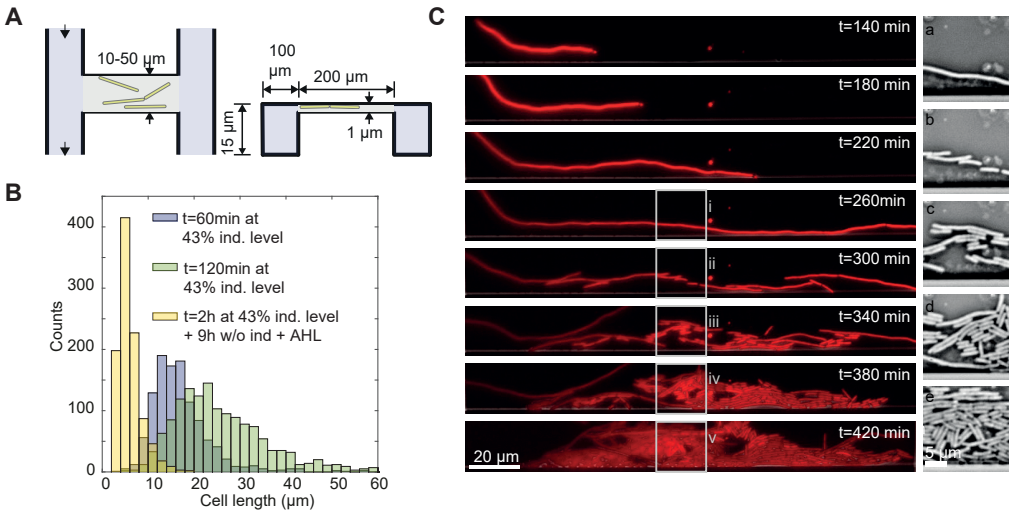

Figure 5.5: Restoration of cell division. **A** Microfluidic trap dimensions for single cell measurements. The medium inlet direction is indicated with arrow marks. The exchange of nutrients and waste products occurs via diffusion. The channel width and trap dimensions are given in the layout. The trap height is 1 μm. The traps are incorporated in a microfluidic gradient mixer [278]. **B** Histogram of cell lengths. Actively switched cells regain normal cell length distributions similar to t = 0 hours of starting induction of filamentous growth. **C** Left column: Fluorescence microscopy time series in the mRFP channel of an *E. coli* cell in a microfluidic chamber in the absence of inducers. Prior to the first image (top), the bacterium was grown filamentous for two hours at 43% induction level followed by a 140-minute exposure to growth medium without any inducers. At 300 minutes of growth, the bacterium resumes cell division. The new daughter cells quickly approach the normal cell size. Right column: images (a)-(e) show bright field images corresponding to the boxed regions on the left.

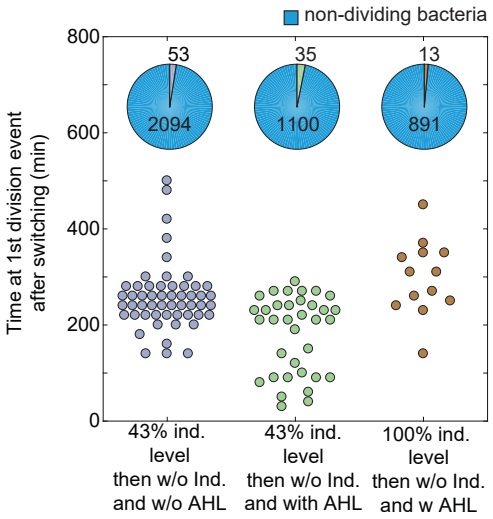

Figure 5.6: Single cell analysis of cell division switching. The time delay between the removal of IPTG/aTc with and without AHL at the first cell division event for different inducer concentrations of IPTG/aTc. The pie charts show the fraction of cells that divided compared to the cells that did not divide for each condition.

tribution on a logarithmic scale, which corresponds to lognormal distribution on a linear scale (Fig. 5.8A). As stated in Sec. 2.3, the lognormal distribution is frequently observed in biology, as any quantity that depends exponentially on the product of several independent rates can be well approximated by a lognormal distribution [279]. In theory, only the limiting distribution of the product of the several independent quantities will approach a Gaussian distribution (central limit theorem), but in practice this can already be applied in cases with several independent rates such as in gene expression.

Second, the mVenus expression could only be fitted with two Gaussian distributions using the Gaussian mixture fitting procedure on a logarithmic scale, indicating that there is a subpopulation of bacteria that has mVenus expression levels differing from those of the bulk of the population (Fig. 5.8B). Most of the bacteria that switch back after being filamentous have a mean mVenus fluorescence that is 0.2 times lower than the average of the main population that does not change back. Low mVenus fluorescence should correlate with low dCas9 concentration (mVenus and dCas9 on the CRISPRi plasmid are both under TetR-control), which provide an explanation to why this subpopulation is more prone to restoration.

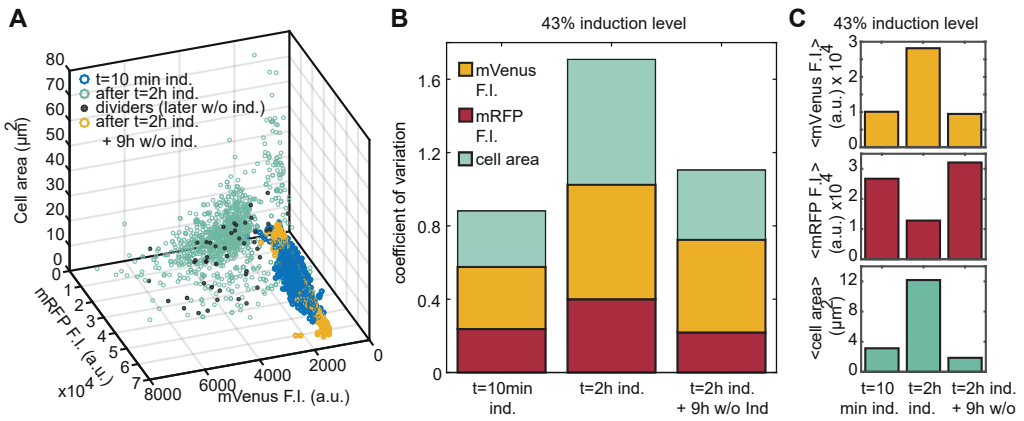

Figure 5.7: **A** 3D scatter plot with the mean mRFP and mVenus fluorescence levels of single bacteria plotted against the observed area of the corresponding cells for about 1000 cells at three different points in time: 10 minutes after induction (43% induction level), after 2 hours of continuous induction and after additional 9 hours without aTc/IPTG. About 5% of the analyzed filamentous cells divide ('dividers': black dots) again after the inducers have been removed from the growth medium. **B** The coefficient of variation was calculated for the fluorescence intensities of mVenus and mRFP and the cell area. The variability of the population increases during the 2 hours induction period and drops again for the new population emerging by switching, i.e., after removal of IPTG/aTc. **C** Upon aTc/IPTG induction, the cells shift towards higher mean mVenus fluorescence intensities, lower mean mRFP and higher mean cell area. After 9 hours without inducers the mean values return to the level of the first measurement (t=10 min ind.).

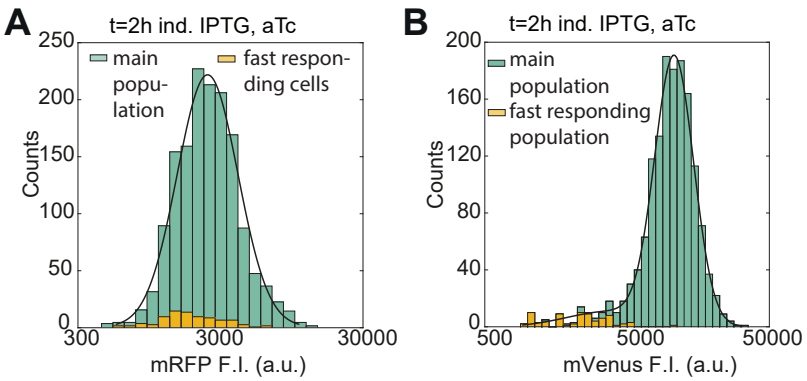

Figure 5.8: **A** Histogram of the mean mRFP fluorescence value for single cells after 2 hours of aTc/IPTG induction using a logarithmic scale on the x-axis. The bacteria that will later divide are differently colored. The mRFP expression levels can be fitted with a lognormal distribution (solid line). **B** Histogram of the mean mVenus fluorescence value for single cells after 2 hours of aTc/IPTG induction using a logarithmic scale on the x-axis. The mVenus expression levels could only be fitted using two lognormal distributions (solid line), indicating that there is a subpopulation with lower mVenus F.I. The bacteria that will later divide (differently colored) are predominately in this subpopulation.

5.3.5 Glycerol stocks

The behavior for intermediate induction level (below 100% induction level) differed between two different glycerol stocks. In the one glycerol stock, which was used for all the experiments in this chapter, except for Fig. 5.9B and C, an induction level of 30% already stopped cell division in nearly all cells so that the length distribution is the same as for 100 % induction level. For the other glycerol stock, a medium induction level (43 %) results in a constant average cell length which means that there is still cell division occurring. The mean cell division rate is reduced in a continuous fashion between zero and full induction (Fig. 5.9C). As found by Li *et al.* [140], at low induction levels there is a co-existence of subpopulations of normal growing and filamentous cells.

5.4 Discussion

In *E. coli* cells, a knockdown of the *ftsZ* gene stops cell division while still allowing cell growth, resulting in long filamentous cells. Restoration of FtsZ levels, in principle, allows cells to return back to normal growth. This opens up the possibility to implement a genetic switch that can turn cell divi-

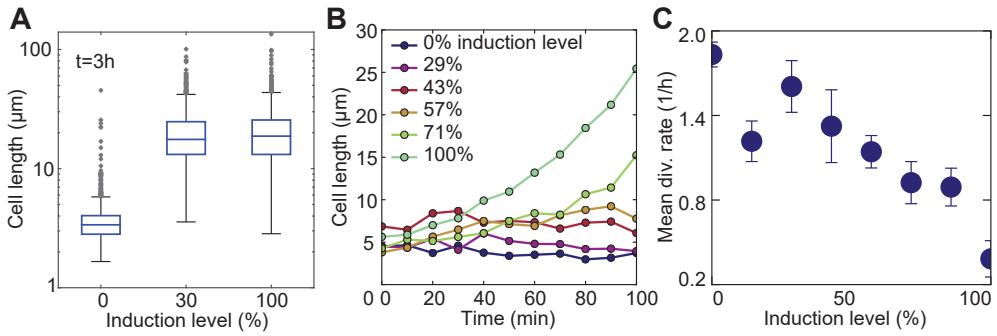

Figure 5.9: **A** Here bacteria come from the glycerol stock that was also used for the other cell-division switching experiments in the chapter. Even a induction level of 30% cells stop to divide and the cell length distribution does not differ from the one at full induction level (t=3h). **B** A different glycerol stock and a different experiment. Tracking the cell length at different induction levels within microfluidic traps. **C** Division rates decrease smoothly with increasing inducer concentrations. Results are mean values over 100 min. Upon 100 % induction of the CRISPRi mechanism, we were able to switch to filamentous cell growth (with a division rate approaching zero).

sion in *E. coli* off and on again. We here used a plasmid-based inducible CRISPRi system to reduce the FtsZ levels in *E. col*i and stop cell division. To implement the circuit in viable bacteria, decoy-binding sites for the FtsZ-suppressing dCas9-sgRNA complexes had to be introduced as genetic buffer elements. We then reverted the knockdown by removing the CRISPRi inducers and counter-acted CRISPRi utilizing inducible antisense-sgRNA molecules that could be used to adjust the threshold for bacterial filamentation and thus modulate the switching process. Due to the critical role of cell division in the bacterial life cycle, however, our CRISPRi-based approach strongly interfered with the physiology of the bacterial host chassis including slowing down cell division restoration in the majority of the cells.

We used microscopy time lapse experiments to follow thousands of cells through the switching process to understand the response to the genetic switch on the single cell level. We found that at best only about 1.5-3 % of the analyzed cells revert back to normal growth whereas the rest of the cells lyse. However, expression of anti-sgRNAs can rescue cell division faster than in experiments, in which only the CRISPRi inducers are removed. Our experiments demonstrate both the possibilities and limitations of using dCas9 as a switchable repressor for genes such as ftsZ that affect bacterial fitness.

5.4.1 Switching cell division off

The CRISPRi-mediated knockdown of ftsZ was a surprisingly efficient method to stop cell division. The microfluidic time-lapse videos of the filamentation process revealed that the distribution of cell lengths broadens with continuing CRISPRi induction. We were able to identify at least three contributions to such a broadening distribution: filamentous cells do not necessarily divide symmetrically [280], larger cells can produce cell mass faster than smaller cells [281] and filamentous cells start to slow down their growth rate at some point and eventually lyse.

It was found that when the FtsZ level falls below ≈ 28 % of its initial value, the cell is not able to form FtsZ rings [259]. We targeted the template strand of the promoters (ftsZ2p, ftsZ3p, ftsZ4p) that, however, only account for about 14 % of the total transcription of ftsZ in the cell [282]. Hence, a major fraction of the FtsZ level reduction may be caused by dCas9-sgRNA blocking the transcriptional elongation also from upstream promoters, although it was previously found that targeting the template strand with the sgRNA does not result in strong repression of transcriptional elongation [53].

In this context, dCas9 off-target kinetics have been previously studied [283]. Off-target binding strongly affects dissociation and association kinetics and thus occupancy of DNA sites, which could further contribute to gene expression heterogeneity. In addition, Cho et al. reported dCas9 binding to several genes even in the absence of guide RNA [142].

To our surprise, the introduction of the plasmid-based inducible CRISPRi-system without decoy-binding sites led to filamentous growth even in the absence of inducers (Fig. 5.2). This stands in contrast to *ftsZ* knockdown using tunable CRISPRi (tCRISPRi) or other previously employed plasmid-based CRISPRi systems [138] [140]. One major difference is that we used an inducible T7 RNAP for the expression of the sgRNAs rather than constitutive E. coli promoters. We assume that the leaky expression of the T7 RNAP together with the leaky expression of dCas9 produces enough dCas9-sgRNA complexes to already sufficiently knock down ftsZ on the chromosome. Previously, strongly binding TFs such as TetR were buffered by the introduction of DNA sponges [20]. In a similar way, we here set the threshold for CRISPRi-based *ftsZ* knockdown via the sponge copy number, solving the problem of non-induced filamentation (Fig. 5.2). Thus, cellular filamentation in the present study could be clearly linked to the CRISPRi. Potential CRISPRi off-target effects have not been studied in the context of this work, however.

5.4.2 Restoration of *E. coli* to normal growth

Microfluidic time lapse videos showed that switching cell division back on was not very efficient. Methods that lack an appropriate time resolution do not allow to distinguish between cells that have just switched back and daughter cells from already restored cells. This could lead to an overestimation of the switching efficiency, in particular when the new state of the bacteria is associated with increased fitness (which is the case here).

In all our switching strategies, restoration of normal growth for most cells occurred within a few hours, which is slower than in studies where *ftsZ* was directly put under the control of an IPTG-inducible promoter [259]. In our case, the removal of the inducers stops new dCas9-sgRNA complexes from being expressed and assembled, but does not influence already bound complexes. In particular, it has been shown that activated dCas9 binds to its recognized DNA target site very strongly ($K_d \approx 0.5$ nM) [283] [284]. In general, transcription factors (TFs) with strong affinity for their DNA-binding site such as TetR do not unbind from their target site until an extremely low amount of only a few molecules per cell (corresponding to low nanomolar concentrations) is reached. In agreement with this consideration, Boyle *et al.* found that dissociation of dCas9 from on-target sites with strong affinity can take many hours [135]. Recovery of FtsZ levels is thus expected to be governed by the dilution of the dCas9-sgRNA complexes which is automatically provided by cell growth. In addition, it has been shown that RNAP or DNA replication machinery knock the dCas9-sgRNA complex from the target site [137]. Thus, cell division can resume at potential division sites as soon as FtsZ levels recover. Other studies aiming at the reversal of CRISPRi due to dilution, reported full restoration of expression levels of targeted fluorescent proteins after washing away CRISPRi inducers at around 350-480 min [53] [140]. Interestingly, with our approach, it took only about 260 minutes on average to observe the first division events after the removal of CRISPRi inducers, suggesting that cell division resumes even before the *ftsZ* levels are fully recovered. However, in microfluidic devices we also observe that the main part of the population does not survive the slow knockdown and recovery process (Fig. 5.6). Importantly, the expression of anti-sgRNA allows some cells to resume cell division faster than without anti-sgRNA. As suggested by the results from the *in vitro* experiment (Fig. 5.3), the anti-sgRNA inhibits complex formation of dCas9 and sgRNA by binding to the sgRNA. Hence, the FtsZ levels is probably recovered faster in some cells with dCas9-sgRNA being kicked off the target site in combination with the expression of anti-sgRNA.

Another CRISPRi anti-sgRNA approach was recently conducted by Lee

et al. [141]. The anti-sense RNA in this study was shown to fully recover the expression of a fluorescent reporter within 360 minutes, even in the presence of continuously expressed sgRNA. Furthermore, they were able to establish a relationship between the de-repression efficiency and the binding affinity of the anti-sgRNA to the sgRNA, supporting the hypothesis that de-repression by anti-sense RNA can be partially explained by equilibrium thermodynamics. In contrast to the simple strand invasion approach taken in our study, Lee *et al.* augmented their antisense-sgRNA with binding sites for the RNA chaperone Hfq, which apparently increased its de-repression efficiency.

Rather than increasing sgRNA-anti-sgRNA interactions in this manner, alternatively the binding strength of dCas9-sgRNA to its target may be weakened. For instance, Vigouroux *et al.* recently demonstrated that reduced complementarity between target DNA and sgRNA increases the probability of the dCas9-sgRNA complex being kicked off by RNA or DNA polymerase [137]. This could help in the release of kinetically trapped dCas9-sgRNA complexes on the DNA target site to the cytosol where they might be more accessible to anti-sgRNA.

5.4.3 Heterogeneity in gene expression

After switching to filamentous growth, we observed an increase in population heterogeneity, which is reflected by altered reporter fluorescence levels, an increase in their coefficients of variation, and by varying bacterial length distributions (Fig. 5.7). The measured decrease in constitutively expressed mRFP fluorescence intensity could be caused by plasmid dilution, a reduced protein expression rate or both.

Although all the plasmids copy number controls are based on a negative feedback mechanism that measures the concentration of the plasmids in the cells, it is unclear what happens to the copy number in filamentous cells [285] [286]. Furthermore, only the copy number of the sum of the pLysS and CRISPRi plasmid should remain fixed, allowing for larger ratios between the two. We continuously selected with the antibiotics of the CRISPRi plasmid, to guarantee the presence of the plasmid. We assume that the subpopulation of bacteria that switches faster than the rest of the cells, and which can be characterized by low mVenus fluorescence (Fig. 5.8), has a low copy number of the CRISPRi plasmid. After switching back to normal cell growth, the new population of growing cells displayed CV values similar to the starting population.

5.5 Materials and Methods

5.5.1 Plasmids

We constructed plasmids by Gibson assembly of synthetic DNA fragments into the target vector (pSB1K3 and CRISPRi plasmid 44249 from addgene). The final plasmid sequences can be found in the appendix. The sender strain was constructed in an earlier study and contains the gene for LuxI synthase (BioBrick part BBa C0261). The sgRNA, anti-sgRNA and sponge element sequences are listed in Appendix.

5.5.2 Bacterial cell culture

Experiments with filamentous cells were performed in Escherichia coli BL21 (DE3) pLysS. Although the pLysS plasmid has the same origin of replication as the CRISPRi plasmid we select with the antibiotics (Carbenicillin) for the CRISPRi plasmid to ensure its presence in the cells. Furthermore, we detect the inducible mVenus expression from the CRISPRi plasmid in the cells. Cells from glycerol stock were grown in 5 ml Luria-Bertani medium containing antibiotics selecting for both plasmids (CRISPRi and asgRNA plasmid) and incubated over night at 37°C and 250 rpm. The following day, cells were diluted 1:1000 and incubated for additional 4 h. Optical density (OD 600nm) values between 0.4-0.6 were obtained. From this batch, 1 ml of the culture was centrifuged and the pellet resuspended in 300 μl growth medium. The concentrated cells were immediately loaded on a microfluidic chamber until single or few bacteria were captured in the traps. In such a microfluidic chemostat with defined bacterial trap dimensions, we supplied the bacterial suspension constantly with fresh nutrients (LB medium, antibiotics and/or inducer chemicals or dyes) using a pressure flow controller (OB1, Elveflow).

5.5.3 Fluorescence time-lapse microscopy

The microfluidic PDMS (Sylgard 182, Dow Corning) device was fabricated using standard soft lithography as previously described [287]. The microfluidic device is a combination of a gradient mixer [278] and bacterial traps designed with dimensions of 200 μm x 10-50 μm x 1 μm as a H-shaped chemostat (Fig. 5.5). Time-lapse microscopy measurements were conducted on a Nikon Ti-Eclipse epi-fluorescence microscope controlled with NIS-Elements Imaging Software. The microscope was equipped with a sCMOS camera (Zyla, Andor), an automated x-y-stage (Prior Scientific, Cambridge, UK) and an incubator box (Okolab) to maintain an operation temperature of

37°C. All videos were recorded with 40x apochromatic magnification objectives. Every 5 to 20 min, images in phase contrast mode, YFP as well as RFP fluorescence mode (in combination with the appropriate filter sets) were taken for a total run time of up to 20 hours. The exposure times were automatically adjusted.

5.5.4 Cell-free expression

For the cell-free assay, the sgRNA sequences were designed complementary to the non-template strand (sequences can be found in the appendix. Anti-sgRNA and sgRNA where transcribed in vitro by T7 RNA Polymerase (NEB) from linear DNA (IDT DNA) overnight and then extracted with Phenol-Chloroform. The concentration of the RNA was determined by comparing a SYBR Green II stained band in a denaturing PAGE (8M Urea at 45 °C) to the RNA Ladder (NEB, N0364S). The plasmid with mVenus was purified using Phenol-Chloroform prior the reaction in the cell extract. The crude S30 cell extract was obtained by beat beating of a BL21- Rosetta2(DE3) mid-log phase culture with 0.1 mm glass beads in a Minilys device (Peqlab) and supplemented with an energy mix and reaction buffer as described in ref. [68]. Instead of 3-phosphoglyceric acid (3-PGA), phosphoenolpyruvate (PEP) was utilized as an energy source. dCas9 was His-tagged and purified by gravity- flow chromatography with Ni-NTA Agarose Beads (Qiagen). The fluorescence intensity was measured with a FLUOstar Omega plate reader (BMG) in 96-well plate (ibidi) at 37°C. The composition of a single cell-free reaction was: 33 % (v/v) S30 cell extract mixed with 42 % (v/v) buffer and 25 % (v/v) DNA plus inducers.

5.5.5 Data analysis

Image analysis was performed using NIS-Elements (Nikon) and customized MATLAB software.

5.6 Contributions

The experiments were conceptualized by Andrea Mückl (A.M.), Friedrich C. Simmel (F.S.) and Matthaeus Schwarz-Schilling (M.S.). The plasmids were designed and constructed by A.M. and Katrin Fischer. The microscopy images of the bacteria with/without decoy binding sites and the control with aTc and without IPTG were acquired by A.M. As well as the experiments and the analysis of the Fig. 5.9B and C. Unless specifically stated otherwise,

the experiments in the cell-free exression system or with bacteria that are presented in this chapter were conducted and analyzed by M.S.

Chapter 6

Conclusion and outlook

In summary, we have used cell-free gene expression to prototype functional aspects of genetic constructs. We have shown that cell-free expression systems provide a simpler and more controllable environment for gene expression and regulation than bacteria. Furthermore, limitations of the proclaimed goal to use cell-free expression as a prototyping platform for *in vivo* applications have been demonstrated.

In the first project, we compared the function of a plasmid-based sender-receiver system in bacteria and a cell-free gene expression system. We encapsulated both expression systems in water-in-oil droplets, which we then loaded into a glass capillary. This linear arrangement of droplets allowed us to study gene expression based on the diffusion of two different chemical inducers, AHL and IPTG, in a quantitative manner. We here found that for characterization of the response of the receiver plasmid to chemical inducers the cell-free expression system is better suited than bacteria. The sender plasmid was successfully implemented in the cell-free expression system, so that we were able to produce sufficient amounts of AHL to establish and demonstrate chemical communication between the cell-free senders and bacterial receivers. The direction of the communication could be reversed, with cell-free receivers listening to bacterial senders. The possibility of communication between cells and artificial non-living cells opens up a field for many applications from sensing of pathogenic bacteria to inhibiting biofilm formation. These experiments show that artificial cellular systems are easier to control in certain aspects than living organisms. The chemical interface between artificial cells and cells could be used to create multicellular hybrid system, in which different modules that are easier to implement in one or the other can be combined.

In a second project, the design, expression and assembly of a RNA-protein hybrid structure in buffer solutions, a cell-free expression system and bacteria

was demonstrated. We found that assembly of the protein-based FRET pair on the RNA structure was more efficient in buffer solution than in a cell-free expression system. However the cell-free expression system proved superior to bacteria. Unwanted cross-talk with other molecules or sub-optimal concentration of components are common reasons for less efficient assembly in more complex environments. We show that a cell-free expression system allows to address the challenges of designing and expressing artificial multi-component nanostructures in a systematic fashion. We believe that there will be more new functional nanostructures made from assembling different kind of molecules. Intricate molecular machines like the ribosome, which consists of three RNA molecules and about 52 proteins demonstrate the potential [103]. We believe that cell-free expression systems are an important tool in bottom-up synthetic biology to develop and create these genetically encoded multi-component nanostructures.

In the third project, CRISPRi and an anti-sgRNA were used in a cell-free expression system to knock-down the expression of a fluorescent protein. Reversibility of CRISPRi was tested in bacteria by targeting the expression of *ftsZ*, which halts cell division if its concentration falls below 28 % of its initial value [259]. We successfully implemented sponge elements to buffer the effect of leaky CRISPRi in an uninduced state. The anti-sense RNA strategy proved to not be efficient in de-repressing gene expression, neither in the cell-free expression system nor in cells. The CRISPRi mediated knockdown of *ftsZ* strongly interfered with the physiology and gene expression levels of reporter proteins in the cell. Hence a cell-free expression system will not be able to predict the response of bacteria to a genetic switch. For synthetic gene switches that perturb the cell cycle, the response of the cells is often complex. This directs the focus of the research away from the synthetic switch to the response of the organism. We believe that cell-free synthetic biology allows to re-direct the focus on the development of tools by providing a controlled and open environment. As the field of cell-free synthetic biology is growing, we can envision that more tools that are designed in cell-free environments will not be transferred to cells but will actually be applied in a cell-free context. In effect, as applications of cell-free synthetic biology are on the rise, we can foresee new biology and biochemistry happening outside of cells.

List of publications

Functional Surface-immobilization of Genes Using Multistep Strand Displacement Lithography
G. Pardatscher*, M. Schwarz-Schilling*, S. Sagredo, F. C. Simmel,
J. Vis. Exp. **140**, e58634, (2018).

Filamentation and restoration of normal growth in Escherichia coli using a combined CRISPRi sgRNA/antisense RNA approach
A. Mückl*, M. Schwarz-Schilling*, K. Fischer, and F. C. Simmel,
PLOS ONE **13**, e0198058 (2018).

Optimized Assembly of a Multifunctional RNA-Protein Nanostructure in a Cell-Free Gene Expression System
M. Schwarz-Schilling, A. Dupin, F. Chizzolini, S. Krishnan, S. S. Mansy, and F. C. Simmel,
Nano Letters **18**, 2650–2657 (2018)

Gene Expression on DNA Biochips Patterned with Strand Displacement Lithography
G. Pardatscher, M.Schwarz-Schilling, S. S. Daube, R. H Bar-Ziv, and F. C. Simmel,
Angewandte Chemie Int. Ed. **57**, 4783-4786 (2018)

Chemical communication between bacteria and cell-free gene expression systems within linear chains of emulsion droplets
M. Schwarz-Schilling*, L. Aufinger*, A. Mückl, F. C. Simmel,
Integrative Biology **8**, 564-570 (2016)

Building a Synthetic Transcriptional Oscillator
M. Schwarz-Schilling, J. Kim, C. Cuba, M. Weitz, E. Franco, F.C. Simmel,
Cell Cycle Oscillators. Methods in Molecular Biology, vol 1342. Humana Press, New York, NY (2016)

*= Authors have contributed equally to this work.

Bibliography

1. Agapakis, C. M. Designing Synthetic Biology. *ACS Synthetic Biology* **3,** 121–128 (2014).
2. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351** (2016).
3. Weiss, M. *et al.* Sequential bottom-up assembly of mechanically stabilized synthetic cells by microfluidics. *Nature Materials* **17,** 89 (2017).
4. *MIT Synthetic Biology Center: About the SBC* `http://synbio.mit.edu`. Accessed: 2018-02-20.
5. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology* **21,** 796–802 (2003).
6. Groves, B. *et al.* Computing in mammalian cells with nucleic acid strand exchange. *Nature Nanotechnology* **11,** 1–9 (2015).
7. Soloveichik, D., Seelig, G. & Winfree, E. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 5393 (2010).
8. Davies, J. A. Real-World Synthetic Biology: Is It Founded on an Engineering Approach, and Should It Be? *Life* **9** (2019).
9. Schmidt, M. Synthetic Biology. The Technoscience and its Societal Consequences: Chapter 6: Do I understand what I can create? Biosafety issues in synthetic biology. **107,** 5393 (2009).
10. Rollie, S., Mangold, M. & Sundmacher, K. Designing biological systems Systems Engineering meets Synthetic Biology. *Chemical Engineering Science* **69,** 1–29 (2012).
11. Simmel, F. C. Physics of Life, Lecture Notes of the 49^{th} IFF Spring School 2018, B9: Synthetic Biology - Science and Engineering of Synthetic Biological Systems. **158** (2018).

12. Knight, T. *Idempotent Vector Design for Standard Assembly of Biobricks* tech. rep. (MIT Synthetic Biology Working Group, 2003).

13. Shetty, R. P., Endy, D. & Knight, T. F. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering* **2,** 5 (2008).

14. IGEM Foundation. *Registry of Standard Biological Parts* Accessed: 2018-04-05. <`http://parts.igem.org/Main_Page`> (2018).

15. Kwok, R. Five hard truths for Synthetic Biology. *Nature* **463,** 288–290 (2010).

16. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods* **10.** Article, 354 (2013).

17. Keren, L. *et al.* Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166,** 1282–1294.e18 (2016).

18. Michaels, Y. S. *et al.* Precise tuning of gene expression levels in mammalian cells. *Nature Communications* **10,** 818 (2019).

19. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).

20. Potvin-Trottier, L., Lord, N. D., Vinnicombe, G. & Paulsson, J. Synchronous long-term oscillations in a synthetic gene circuit. *Nature* **538,** 514–517 (2016).

21. Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie International Edition* **57,** 4143–4148 (2018).

22. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).

23. Balaji, S., Iyer, L. M., Aravind, L. & Babu, M. M. Uncovering a Hidden Distributed Architecture Behind Scale-free Transcriptional Regulatory Networks. *Journal of Molecular Biology* **360,** 204–212 (2006).

24. Kashtan, N. & Alon, U. Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 13773 (2005).

25. Meyer, A. J., Ellefson, J. W. & Ellington, A. D. Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for in Vivo or in Vitro Synthetic Circuitry. *ACS Synthetic Biology* **4,** 1070–1076 (2015).

26. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nature Reviews Microbiology* **16,** 277 (2018).

27. Choudhary, S., Quin, M. B., Sanders, M. A., Johnson, E. T. & Schmidt-Dannert, C. Engineered Protein Nano-Compartments for Targeted Enzyme Localization. *PLOS ONE* **7,** 1–11 (2012).

28. Lau, Y. H., Giessen, T. W., Altenburg, W. J. & Silver, P. A. Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nature Communications* **9,** 1311 (2018).

29. Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nature Biotechnology* **31,** 335 (2013).

30. Zawada, J. F. *et al.* Microscale to manufacturing scale-up of cell-free cytokine production–a new approach for shortening protein production development timelines. *Biotechnology and Bioengineering* **108,** 1570–1578.

31. Niederholtmeyer, H. *et al.* Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife* **4** (2015).

32. Tirabassi, R. *Foundations of Molecular Cloning - Past, Present and Future* Accessed: 2019-04-07. <`https://www.neb.com/tools-and-resources/feature-articles/foundations-of-molecular-cloning-past-present-and-future`> (2019).

33. Luria, S. E. & Human, M. L. A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology* **64,** 557–569 (1952).

34. Linn, S. & Arber, W. Host Specificity of DNA Produced by Escherichia coli, X. In vitro Restriction of Phage fd Replicative Form. *Proceedings of the National Academy of Sciences of the United States of America* **59,** 1300–1306 (1968).

35. Bode, V. C. & Kaiser, A. D. Changes in the structure and activity of l DNA in a superinfected immune bacterium. *Journal of Molecular Biology* **14,** 399–417 (1965).

36. Cozzarelli, N., Melechen, N., Jovin, T. & Kornberg, A. Polynucleotide cellulose as a substrate for a polynucleotide ligase induced by phage T4. *Biochemical and biophysical research communications* **28,** 578–586 (1967).

37. Gefter, M. L., Becker, A. & Hurwitz, J. The enzymatic repair of DNA. I. Formation of circular lambda-DNA. *Proceedings of the National Academy of Sciences of the United States of America* **58,** 240–247 (1967).

38. Gellert, M. Formation of covalent circles of lambda DNA by E. coli extracts. *Proceedings of the National Academy of Sciences of the United States of America* **57,** 148–155 (1967).

39. Olivera, B. M. & Lehman, I. R. Linkage of polynucleotides through phosphodiester bonds by an enzyme from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **57,** 1426–1433 (1967).

40. Weiss, B. & Richardson, C. C. Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America* **57,** 1021–1028 (1967).

41. Cohen, S. N., Chang, A. C. Y. & Hsu, L. Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of Escherichia coli by R-Factor DNA. *Proceedings of the National Academy of Sciences of the United States of America* **69,** 2110–2114 (1972).

42. Mullis, K. B. & Faloona, F. A. in *Recombinant DNA Part F* 335–350 (Academic Press, 1987).

43. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74,** 5463–5467 (1977).

44. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3,** 318–356 (1961).

45. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat Methods* **11,** 508–520 (2014).

46. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).

47. Weiss, R. *et al.* Genetic circuit building blocks for cellular computation, communications, and signal processing. *Natural Computing* **2,** 47–84 (2003).

48. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology* **10,** 410 (2009).

49. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).

50. Kramer, B. P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnology* **22,** 867 (2004).

51. Weiss, R. & Knight Jr., T. F. *Engineered Communications for Microbial Robotics* in *Revised Papers from the 6th International Workshop on DNA-Based Computers: DNA Computing* (Springer-Verlag, Berlin, Heidelberg, 2001), 1–16.

52. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352** (2016).

53. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).

54. Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell* **159,** 925–939 (2014).

55. Nielsen, A. A. & Voigt, C. A. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular Systems Biology* **10,** 763–763 (2014).

56. Green, A. A. *et al.* Complex cellular logic computation using ribocomputing devices. *Nature* **548,** 117 (2017).

57. Riglar, D. T. & Silver, P. A. Engineering bacteria for diagnostic and therapeutic applications. *Nature Reviews Microbiology* **16,** 214 (2018).

58. Kim, S. *et al.* Quorum Sensing Can Be Repurposed To Promote Information Transfer between Bacteria in the Mammalian Gut. *ACS Synthetic Biology* **7,** 2270–2281 (2018).

59. Chong, S. Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications. *Current Protocols in Molecular Biology* **108,** 16.30.1–16.30.11 (2014).

60. Buntru, M., Vogel, S., Spiegel, H. & Schillberg, S. Tobacco BY-2 cell-free lysate: an alternative and highly-productive plant-based in vitro translation system. *BMC Biotechnology* **14,** 37 (2014).

61. Nirenberg, M. W. & Matthaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **47,** 1588–1602 (1961).

62. Iizuka, N., Najita, L., Franzusoff, A. & Sarnow, P. Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae. *Molecular Cell Biology* **14,** 7322–7330 (1994).

63. Algire, M. A. *et al.* Development and characterization of a reconstituted yeast translation initiation system. *RNA* **8,** 382–397 (2002).

64. Shin, J., Jardine, P. & Noireaux, V. Genome Replication, Synthesis, and Assembly of the Bacteriophage T7 in a Single Cell-Free Reaction. *ACS Synthetic Biology* **1,** 408–413 (2012).

65. Hodgman, C. E. & Jewett, M. C. Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering* **14,** 261–269 (2012).

66. Smith, M. T., Wilding, K. M., Hunt, J. M., Bennett, A. M. & Bundy, B. C. The emerging age of cell-free synthetic biology. *FEBS Letters* **588,** 2755–2761 (2014).

67. Kigawa, T. *et al.* Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *Journal of Structural and Functional Genomics* **5,** 63–68 (2004).

68. Sun, Z. Z. *et al.* Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *Journal of Visualized Experiments,* e50762 (2013).

69. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19,** 751 (2001).

70. Isalan, M., Lemerle, C. & Serrano, L. Engineering Gene Networks to Emulate Drosophila Embryonic Pattern Formation. *PLoS Biology* **3,** e64 (2005).

71. Karzbrun, E., Tayar, A. M., Noireaux, V. & Bar-Ziv, R. H. Programmable on-chip DNA compartments as artificial cells. *Science* **345,** 829 (2014).

72. Tayar, A. M., Karzbrun, E., Noireaux, V. & Bar-Ziv, R. H. Propagating gene expression fronts in a one-dimensional coupled system of artificial cells. *Nature Physics* **11,** 1037 (2015).

73. Bayley, H. *et al.* Droplet interface bilayers. *Molecular Biosystems* **4,** 1191–1208 (2008).

74. Dupin, A. & Simmel, F. C. Signalling and differentiation in emulsion-based multi-compartmentalized in vitro gene circuits. *Nature Chemistry* **11,** 32–39 (2019).

75. Noireaux, V., Bar-Ziv, R. & Libchaber, A. Principles of cell-free genetic circuit assembly. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 12672–12677 (2003).

76. Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V. & Murray, R. M. Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an Escherichia coli Based TX-TL Cell-Free System. *ACS Synthetic Biology* **3,** 387–397 (2014).

77. Takahashi, M. K. *et al.* Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. *ACS Synthetic Biology* **4,** 503–15 (2015).

78. Niederholtmeyer, H., Stepanova, V. & Maerkl, S. J. Implementation of cell-free biological networks at steady state. *Proceedings of the National Academy of Sciences of the United States of America,* 1–6 (2013).

79. Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V. & Murray, R. M. Gene circuit performance characterization and resource usage in a cell-free "breadboard". *ACS Synthetic Biology* **3,** 416–25 (2014).

80. Nieß, A., Failmezger, J., Kuschel, M., Siemann-Herzberg, M. & Takors, R. Experimentally Validated Model Enables Debottlenecking of in Vitro Protein Synthesis and Identifies a Control Shift under in Vivo Conditions. *ACS Synthetic Biology* **6,** 1913–1921 (2017).

81. Litschel, T., Ramm, B., Maas, R., Heymann, M. & Schwille, P. Beating Vesicles: Encapsulated Protein Oscillations Cause Dynamic Membrane Deformations. *Angewandte Chemie International Edition* **57,** 16286–16290 (2018).

82. Pardee, K. *et al.* Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* **165,** 1255–1266 (2016).

83. Wen, K. Y. *et al.* A Cell-Free Biosensor for Detecting Quorum Sensing Molecules in P.aeruginosa-Infected Respiratory Samples. *ACS Synthetic Biology* **6,** 2293–2301 (2017).

84. Kim, D.-M. & Swartz, J. R. Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of Escherichia coli. *Biotechnology and Bioengineering* **85,** 122–129 (2004).

85. Kanter, G. *et al.* Cell-free production of scFv fusion proteins: an efficient approach for personalized lymphoma vaccines. *Blood* **109,** 3393–3399 (2007).

86. Opgenorth, P. H., Korman, T. P., Iancu, L. & Bowie, J. U. A molecular rheostat maintains ATP levels to drive a synthetic biochemistry system. *Nature Chemical Biology* **13,** 938 (2017).

87. Schwille, P. *et al.* MaxSynBio: Avenues Towards Creating Cells from the Bottom Up. *Angewandte Chemie International Edition* **57,** 13382–13392 (2018).

88. Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S. & Erb, T. J. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **354,** 900–904 (2016).

89. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nature Reviews Materials* **3,** 17068 (2017).

90. Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).

91. Adleman, L. Molecular computation of solutions to combinatorial problems. *Science* **266,** 1021–1024 (1994).

92. Woods, D. *et al.* Diverse and robust molecular algorithms using reprogrammable DNA self-assembly. *Nature* **567,** 366–372 (2019).

93. Padirac, A., Fujii, T., Estévez-Torres, A. & Rondelez, Y. Spatial waves in synthetic biochemical networks. *Journal of the American Chemical Society* **135,** 14586–14592 (2013).

94. Zadorin, A. S. *et al.* Synthesis and materialization of a reaction-diffusion French flag pattern. *Nature Chemistry* **9.** Article, 990 (2017).

95. Schwarz-Schilling, M., Aufinger, L., Mückl, A. & Simmel, F. C. Chemical communication between bacteria and cell-free gene expression systems within linear chains of emulsion droplets. *Integrative Biology* **8,** 564–570 (2016).

96. Schwarz-Schilling, M. *et al.* Optimized Assembly of a Multifunctional RNA-Protein Nanostructure in a Cell-Free Gene Expression System. *Nano Letters* **18,** 2650–2657 (2018).

97. Mückl, A., Schwarz-Schilling, M., Fischer, K. & Simmel, F. C. Filamentation and restoration of normal growth in Escherichia coli using a combined CRISPRi sgRNA/antisense RNA approach. *PLoS ONE* **13,** 1–19 (2018).

98. Crick, F. H. On protein synthesis. *Symposia of the Society for Experimental Biology* **12,** 138–163 (1958).

99. Hartvig, L. & Christiansen, J. Intrinsictermination ofT7 RNA polymerase mediated by either RNA or DNA. *EMBO Journal* **15,** 4767–4774 (1996).

100. Arnold, S. *et al.* Kinetic modeling and simulation of in vitro transcription by phage T7 RNA polymerase. *Biotechnology and Bioengineering* **72,** 548–561 (2001).

101. Macdonald, L. E., Durbin, R. K., Dunn, J. J. & McAllister, W. T. Characterization of Two Types of Termination Signal for Bacteriophage T7 RNA Polymerase. *Journal of Molecular Biology,* 145–158 (1994).

102. Alberts B Johnson A, L. J. *Molecular Biology of the Cell. Chapter 6, From DNA to RNA* 5th edition (Garland Science, New York, USA, 2002).

103. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **289,** 905–920 (2000).

104. Kurland, C. G. Molecular characterization of ribonucleic acid from Escherichia coli ribosomes: I. Isolation and molecular weights. *Journal of Molecular Biology* **2,** 83–91 (1960).

105. Wikipedia Contributors (2019-04-08). *Translation (biology) - Wikipedia, The Free Encyclopedia* Accessed: 2019-04-21. <`https://en.wikipedia.org/wiki/Translation_(biology)#/media/File:Ribosome_mRNA_translation_en.svg`>.

106. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* 1st edition (Chapman and Hall/CRC, New York, 2006).

107. Van Assche, E., Van Puyvelde, S., Vanderleyden, J. & Steenackers, H. P. RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Frontiers Microbiology* **6,** 141–141 (2015).

108. Bertram, R. & Schuster, C. F. Post-transcriptional regulation of gene expression in bacterial pathogens by toxin-antitoxin systems. *Frontiers in Cellular Infection Microbiology* **4,** 6–6 (2014).

109. Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences* **29,** 11–17 (2004).

110. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 12795 (2002).

111. Elowitz, M. B. Stochastic Gene Expression in a Single Cell. *Science* **297,** 1183–1186 (2002).

112. Tsimring, L. S. Noise in biology. *Reports on Progress in Physics* **77,** 026601 (2014).

113. Taniguchi, Y. *et al.* Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329,** 533 (2010).

114. Beal, J. Biochemical complexity drives log-normal variation in genetic expression. *Engineering Biology* **1,** 55–60 (2017).

115. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135,** 216–226 (2008).

116. Furusawa, C. & Kaneko, K. Zipf's Law in Gene Expression. *Physical Review Letters* **90,** 4 (2003).

117. Furusawa, C., Suzuki, T., Kashiwagi, A., Yomo, T. & Kaneko, K. Ubiquity of Log-normal Distributions in Intra-cellular Reaction Dynamic. *Biophysics* **1,** 25–31 (2005).

118. Powers, D. M. W. *Applications and Explanations of Zipf's Law* in *Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning* (Association for Computational Linguistics, Sydney, Australia, 1998), 151–160.

119. Kirkwood, T. B. L. Geometric Means and Measures of Dispersion. *Biometrics,* 908–909 (1979).

120. Limpert, E., Stahel, W. A. & Abbt, M. Log-normal Distributions across the Sciences : Keys and Clues. *BioScience* **51,** 341–352 (2001).

121. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37,** 67–78 (2017).

122. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).

123. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics* **46,** 505–529 (2017).

124. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–821 (2012).

125. DiCarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Research* **41,** 4336–4343 (2013).

126. Li, J.-F. *et al.* Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nature Biotechnology* **31,** 688 (2013).

127. Shan, Q. *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology* **31,** 686 (2013).

128. Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G. & Kamoun, S. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nature Biotechnology* **31,** 691 (2013).

129. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–823 (2013).

130. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* **31,** 227 (2013).

131. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339,** 823–826 (2013).

132. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology* **31,** 230 (2013).

133. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2,** e00471 (2013).

134. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155,** 1479–1491 (2013).

135. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* **41,** 7429–7437 (2013).

136. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* **8,** 2180–2196 (2013).

137. Vigouroux, A., Oldewurtel, E., Cui, L., Bikard, D. & van Teeffelen, S. Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes. *Molecular Systems Biology* **14,** e7899–e7899 (2018).

138. Elhadi, D., Lv, L., Jiang, X. R., Wu, H. & Chen, G. Q. CRISPRi engineering E. coli for morphology diversification. *Metabolic Engineering* **38,** 358–369 (2016).

139. Jones, D. L. *et al.* Kinetics of dCas9 target search in Escherichia coli. *Science* **357,** 1420 (2017).

140. Li, X. T. *et al.* tCRISPRi: tunable and reversible, one-step control of gene expression. *Scientific Reports* **6,** 39076 (2016).

141. Lee, Y. J., Hoynes-O'Connor, A., Leong, M. C. & Moon, T. S. Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Research* **44,** 2462–2473 (2016).

142. Cho, S. *et al.* High-Level dCas9 Expression Induces Abnormal Cell Morphology in Escherichia coli. *ACS Synthetic Biology* **7,** 1085–1094 (2018).

143. Cui, L. *et al.* A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. *Nature Communications* **9,** 1912 (2018).

144. Seeman, N. C. Nucleic acid junctions and lattices. *Journal of Theoretical Biology* **99,** 237–247 (1982).

145. Seeman, N. C. Nanomaterials based on DNA. *Annual Review of Biochemistry* **79,** 65–87 (2010).

146. Grossi, G., Dalgaard Ebbesen Jepsen, M., Kjems, J. & Andersen, E. S. Control of enzyme reactions by a reconfigurable DNA nanovault. *Nature Communications* **8,** 992 (2017).

147. Langecker, M. *et al.* Synthetic Lipid Membrane Channels Formed by Designed DNA Nanostructures. *Science* **338,** 932–936 (2012).

148. Wagenbauer, K. F., Sigl, C. & Dietz, H. Gigadalton scale shape programmable DNA assemblies. *Nature* **552,** 78–83 (2017).

149. Praetorius, F. & Dietz, H. Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes. *Science* **355** (2017).

150. Jaeger, L. & Leontis, N. B. Tecto-RNA: One-Dimensional Self-Assembly through Tertiary Interactions. *Angewandte Chemie International Edition* **39,** 2521–2524 (2000).

151. Chworos, A. & Jaeger, L. Building Programmable Jigsaw Puzzles with RNA. *Science* **306,** 2068–2072 (2004).

152. Afonin, K. A. *et al.* In vitro assembly of cubic RNA-based scaffolds designed in silico. *Nature Nanotechnology* **5,** 676–82 (2010).

153. Endo, M., Takeuchi, Y., Emura, T., Hidaka, K. & Sugiyama, H. Preparation of Chemically Modified RNA Origami Nanostructures. *Chemistry - A European Journal* **20,** 15330–15333 (2014).

154. Geary, C., Chworos, A., Verzemnieks, E., Voss, N. R. & Jaeger, L. Composing RNA Nanostructures from a Syntax of RNA Structural Modules. *Nano Letters* **17,** 7095–7101 (2017).

155. Geary, C., Rothemund, P. W. & Andersen, E. S. RNA nanostructures. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **345,** 799–804 (2014).

156. Jaeger, L. & Chworos, A. The architectonics of programmable RNA and DNA nanostructures. *Current Opinion in Structural Biology* **16,** 531–43 (2006).

157. Sachdeva, G., Garg, A., Godding, D., Way, J. C. & Silver, P. A. In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic Acids Research* **42,** 9493–503 (2014).

158. Shu, Y., Shu, D., Haque, F. & Guo, P. Fabrication of pRNA nanoparticles to deliver therapeutic RNAs and bioactive compounds into tumor cells. *Nature Protocols* **8,** 1635–1659 (2013).

159. Klug, S. & Famulok, M. All you wanted to know about SELEX. *Molecular Biology Reports* **20,** 97–107 (1994).

160. Paige, J. S., Nguyen-Duc, T., Song, W. & Jaffrey, S. R. Fluorescence imaging of cellular metabolites with RNA. *Science* **335,** 1194 (2012).

161. Lim, F., Downey, T. P. & Peabody, D. S. Translational Repression and Specific RNA Binding by the Coat Protein of the Pseudomonas Phage PP7. *Journal of Biological Chemistry* **276,** 22507–22513 (2001).

162. Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333,** 470–4 (2011).

163. Schwartz, C. & Guo, P. Ultrastable pRNA hexameric ring gearing hexameric phi29 DNA-packaging motor by revolving without rotating and coiling. *Current Opinion in Biotechnology* **24,** 581–590 (2013).

164. Zhang, H. *et al.* Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. *RNA* **19,** 1226–37 (2013).

165. Khisamutdinov, E. F., Jasinski, D. L. & Guo, P. RNA as a Boiling-Resistant Anionic Polymer Material To Build Robust Structures with Defined Shape and Stoichiometry. *ACS Nano* **8,** 4771–4781 (2014).

166. Shu, Y. *et al.* Fabrication of 14 different RNA nanoparticles for specific tumor targeting without accumulation in normal organs. *RNA* **19,** 767–777 (2013).

167. Gilbert, S. *Developmental Biology, Multicellularity: The Evolution of Differentiation.* 6th edition (Sinauer Associates, Sunderland (MA), USA, 2000).

168. Gurdon, J. B. & Bourillot, P.-Y. Morphogen gradient interpretation. *Nature* **413,** 797–803 (2001).

169. Wolpert, L. Positional information and patterning revisited. *Journal of Theoretical Biology* **269,** 359–365 (2011).

170. Driever, W. & Nüsslein-Volhard, C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54,** 95–104 (1988).

171. Hol, F. J. H. & Dekker, C. Zooming in to see the bigger picture: Microfluidic and nanofabrication tools to study bacteria. *Science* **346,** 1251821–1251821 (2014).

172. Locke, J. C. W. & Elowitz, M. B. Using movies to analyse gene circuit dynamics in single cells. *Nature Reviews Microbiology* **7,** 383 (2009).

173. Balagadde, F. K., You, L., Hansen, C. L., Arnold, F. H. & Quake, S. R. Long-Term Monitoring of Bacteria Undergoing Programmed Population Control in a Microchemostat. *Science* **309,** 137–140 (2005).

174. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123,** 1025–1036 (2005).

175. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463,** 326–330 (2010).

176. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in Escherichia coli. *Cell* **113,** 597–607 (2003).

177. Xavier, K. B. & Bassler, B. L. LuxS quorum sensing: more than just a numbers game. *Current Opinion in Microbiology* **6,** 191–197 (2003).

178. Nealson, K. H., Platt, T. & Hastings, J. W. Cellular Control of the Synthesis and Activity of the Bacterial Luminescent System. *Journal of Bacteriology* **104,** 313–322 (1970).

179. Fuqua, C. & Greenberg, E. P. Listening in on bacteria: acyl-homoserine lactone signalling. *Nature Reviews Molecular Cell Biology* **3.** Review Article, 685 (2002).

180. Davies, D. G. *et al.* The Involvement of Cell-to-Cell Signals in the Development of a Bacterial Biofilm. *Science* **280,** 295–298 (1998).

181. Derzelle, S., Duchaud, E., Kunst, F., Danchin, A. & Bertin, P. Identification, Characterization, and Regulation of a Cluster of Genes Involved in Carbapenem Biosynthesis in Photorhabdus luminescens. *Applied and Environmental Microbiology* **68,** 3780–3789 (2002).

182. Whiteley, M., Lee, K. M. & Greenberg, E. P. Identification of genes controlled by quorum sensing in Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 13904 (Nov. 1999).

183. Perego, M. & Hoch, J. A. Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 1549–1553 (1996).

184. Galloway, W. R. J. D., Hodgkinson, J. T., Bowden, S. D., Welch, M. & Spring, D. R. Quorum Sensing in Gram-Negative Bacteria: Small-Molecule Modulation of AHL and AI-2 Quorum Sensing Pathways. *Chemical Reviews* **111,** 28–67 (2011).

185. Fuqua, C. & Greenberg, E. P. Signalling: Listening in on bacteria: acyl-homoserine lactone signalling. *Nature reviews. Molecular Cell Biology* **3,** 685–695 (2002).

186. Weitz, M. *et al.* Communication and Computation by Bacteria Compartmentalized within Microemulsion Droplets. *Journal of the American Chemical Society* **136,** 72–75 (2014).

187. Holtze, C. *et al.* Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab on a Chip* **8,** 1632 (2008).

188. Kubitschek, H. E. & Friske, J. A. Determination of bacterial cell volume with the Coulter Counter. *Journal of Bacteriology* **168,** 1466–1467 (1986).

189. Karzbrun, E., Shin, J., Bar-Ziv, R. H. & Noireaux, V. Coarse-grained dynamics of protein synthesis in a cell-free system. *Physical Review Letters* **106,** 048104 (2011).

190. Del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M. & Diaz-Orejas, R. Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews* **62,** 434–64 (1998).

191. Lin, J. & Amir, A. Homeostasis of protein and mRNA concentrations in growing cells. *Nature Communications* **9,** 4496 (2018).

192. Kawaguchi, T., Chen, Y. P., Norman, R. S. & Decho, A. W. Rapid Screening of Quorum-Sensing Signal N-Acyl Homoserine Lactones by an In Vitro Cell-Free Assay. *Applied and Environmental Microbiology* **74,** 3667–3671 (2008).

193. Yang, Y.-H. *et al.* Cell-Free Escherichia coli-Based System To Screen for Quorum-Sensing Molecules Interacting with Quorum Receptor Proteins of Streptomyces coelicolor. *Applied and Environmental Microbiology* **75,** 6367–6372 (2009).

194. Urbanowski, M. L., Lostroh, C. P. & Greenberg, E. P. Reversible Acyl-Homoserine Lactone Binding to Purified Vibrio fischeri LuxR Protein. *Journal of Bacteriology* **186,** 631–637 (2004).

195. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology* **26,** 787 (2008).

196. Halleran, A. D. & Murray, R. M. Cell-Free and In Vivo Characterization of Lux, Las, and Rpa Quorum Activation Systems in E. coli. *ACS Synthetic Biology* **7,** 752–755 (2018).

197. Kamaraju, K. *et al.* Effects on Membrane Lateral Pressure Suggest Permeation Mechanisms for Bacterial Quorum Signaling Molecules. *Biochemistry* **50,** 6983–6993 (2011).

198. Li, X. *et al.* Development and application of a method for the analysis of N-acylhomoserine lactones by solid-phase extraction and ultra high pressure liquid chromatography. *Journal of Chromatography A* **1134,** 186–193 (2006).

199. Gruner, P. *et al.* Controlling molecular transport in minimal emulsions. *Nature Communications* **7,** 1–9 (1).

200. Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie* **99,** 162–168 (2014).

201. Crank, J. *The mathematics of diffusion* (Oxford University Press, 1979).

202. Choi, W. S., Ha, D., Park, S. & Kim, T. Synthetic multicellular cell-to-cell communication in inkjet printed bacterial cell systems. *Biomaterials* **32,** 2500–2507 (2011).

203. Dilanji, G. E., Langebrake, J. B., De Leenheer, P. & Hagen, S. J. Quorum Activation at a Distance: Spatiotemporal Patterns of Gene Regulation from Diffusion of an Autoinducer Signal. *Journal of the American Chemical Society* **134,** 5618–5626 (2012).

204. Missner, A. & Pohl, P. 110 Years of the Meyer Overton Rule Predicting Membrane Permeability of Gases and Other Small Compounds. *ChemPhysChem* **10,** 1405–1414 (2009).

205. Llinàs, M., Caldero, G., García-Celma, M. J., Patti, A. & Solans, C. New insights on the mechanisms of drug release from highly concentrated emulsions. *Journal of Colloid and Interface Science* **394,** 337–345 (2013).

206. Tompkins, N. *et al.* Testing Turing's theory of morphogenesis in chemical cells. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 4397 (2014).

207. Crick, F. Diffusion in Embrygenesis. *Nature* **225,** 1–3 (1970).

208. Gruner, P. *Molecular Transport in Emulsions: From Permeation to Controlled Delivery using Microfluidics* PhD Thesis (Georg-August-Universitaet Goettingen, 2015).

209. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434,** 1130–1134 (2005).

210. Hense, B. A. *et al.* Does efficiency sensing unify diffusion and quorum sensing? *Nature reviews. Microbiology* **5,** 230–239 (2007).

211. Iizuka, R., Yamagishi-Shirasaki, M. & Funatsu, T. Kinetic study of de novo chromophore maturation of fluorescent proteins. *Analytical Biochemistry* **414,** 173–178 (2011).

212. Sklodowska, K. & Jakiela, S. Enhancement of bacterial growth with the help of immiscible oxygenated oils. *RSC Advances* **7,** 40990–40995 (2017).

213. Kim, D.-M. & Swartz, J. R. Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnology and Bioengineering* **74,** 309–316 (2001).

214. Kaneda, S. *et al.* Modification of the Glass Surface Property in PDMS-Glass Hybrid Microfluidic Devices. *Analytical Sciences* **28,** 39–39 (2012).

215. Skhiri, Y. *et al.* Dynamics of molecular transport by surfactants in emulsions. *Soft Matter* **8,** 10618 (2012).

216. Kapsner, K. & Simmel, F. C. Partitioning Variability of a Compartmentalized In Vitro Transcriptional Thresholding Circuit. *ACS Synthetic Biology* **9,** 1136–1143 (2015).

217. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **333,** 642–646 (2011).

218. Greenbaum, N. L. How Tat targets TAR: structure of the BIV peptide-RNA complex. *Structure* **4,** 5–9 (1996).

219. Scott, B. L. & Hoppe, A. D. Optimizing fluorescent protein trios for 3-Way FRET imaging of protein interactions in living cells. *Scientific Reports* **5,** 10270 (2015).

220. Shu, D., Khisamutdinov, E. F., Zhang, L. & Guo, P. Programmable folding of fusion RNA in vivo and in vitro driven by pRNA 3WJ motif of phi29 DNA packaging motor. *Nucleic Acids Research* **42,** e10 (2014).

221. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338,** 1177–1183 (2012).

222. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523,** 441–4 (2015).

223. Haque, F. *et al.* Ultrastable synergistic tetravalent RNA nanoparticles for targeting to cancers. *Nano Today* **7,** 245–257 (2012).

224. Liu, D. *et al.* Synthesizing topological structures containing RNA. *Nature Communication* **8,** 14936 (2017).

225. Gradi Scaron Ar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nature Chemical Biology* **9,** 362–366 (2013).

226. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537,** 320–7 (2016).

227. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340,** 595–9 (2013).

228. Ohno, H. *et al.* Synthetic RNA-protein complex shaped like an equilateral triangle. *Nature Nanotechnology* **6,** 116–20 (2011).

229. Ponchon, L. *et al.* Co-expression of RNA-protein complexes in Escherichia coli and applications to RNA biology. *Nucleic Acids Research* **41,** e150 (2013).

230. Janas, T., Janas, T. & Yarus, M. Human tRNA(Sec) associates with HeLa membranes, cell lipid liposomes, and synthetic lipid bilayers. *RNA* **18,** 2260–8 (2012).

231. Den Hamer, A. *et al.* Small-Molecule-Induced and Cooperative Enzyme Assembly on a 14-3-3 Scaffold. *ChemBioChem* **18,** 331–335 (2017).

232. Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5 A resolution. *Science* **310,** 827–34 (2005).

233. Will, C. L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology* **3,** 1–23 (2011).

234. Reiter, N. J. *et al.* Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468,** 784–791 (2010).

235. Guo, P., Zhang, C., Chen, C., Garver, K. & Trottier, M. Inter-RNA interaction of phage phi29 pRNA to form a hexameric complex for viral DNA transportation. *Molecular Cell* **2,** 149–55 (1998).

236. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160,** 339–50 (2015).

237. Chakraborty, S., Mehtab, S. & Krishnan, Y. The Predictive Power of Synthetic Nucleic Acid Technologies in RNA Biology. *Accounts of Chemical Research* **47,** 1710–1719 (2014).

238. Grate, D. & Wilson, C. Laser-mediated, site-specific inactivation of RNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 6131–6 (1999).

239. Jepsen, M. D. E. *et al.* Development of a genetically encodable FRET system using fluorescent RNA aptamers. *Nature Communication* **9,** 18 (2018).

240. Babendure, J. R., Adams, S. R. & Tsien, R. Y. Aptamers switch on fluorescence of triphenylmethane dyes. *Journal of the American Chemical Society* **125,** 14716–7 (2003).

241. Bayer, T. S., Booth, L. N., Knudsen, S. M. & Ellington, A. D. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA* **11,** 1848–57 (2005).

242. Chao, J. A., Patskovsky, Y., Almo, S. C. & Singer, R. H. Structural basis for the coevolution of a viral RNA-protein complex. *Nature Structural Molecular Biology* **15,** 103–5 (2008).

243. Srisawat, C. & Engelke, D. R. Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* **7,** 632–41 (2001).

244. Iost, I., Guillerez, J. & Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *Journal of Bacteriology* **174,** 619–22 (1992).

245. Mairhofer, J., Wittwer, A., Cserjan-Puschmann, M. & Striedner, G. Preventing T7 RNA polymerase read-through transcription-A synthetic termination signal capable of improving bioprocess stability. *ACS Synthetic Biology* **4,** 265–73 (2015).

246. Xu, X., Zhao, P. & Chen, S.-J. Vfold: A Web Server for RNA Structure and Folding Thermodynamics Prediction. *PLoS ONE* **9,** 1–7 (2014).

247. Goedhart, J. *et al.* Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature Communication* **3,** 751 (2012).

248. Nguyen, A. W. & Daugherty, P. S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nature Biotechnology* **23,** 355–60 (2005).

249. Fábián, Á. I., Rente, T., Szöllősi, J., Mátyus, L. & Jenei, A. Strength in Numbers: Effects of Acceptor Abundance on FRET Efficiency. *ChemPhysChem* **11,** 3713–3721 (2010).

250. Bray, D. & Lay, S. Computer-based analysis of the binding steps in protein complex formation. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 13493–8 (1997).

251. Wu, B., Chao, J. A. & Singer, R. H. Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. *Biophysical Journal* **102,** 2936–44 (2012).

252. Douglass E. F., J., Miller, C. J., Sparer, G., Shapiro, H. & Spiegel, D. A. A comprehensive mathematical model for three-body binding equilibria. *Journal of the American Chemical Society* **135,** 6092–9 (2013).

253. Heyman, Y., Buxboim, A., Wolf, S. G., Daube, S. S. & Bar-Ziv, R. H. Cell-free protein synthesis and assembly on a biochip. *Nature Nanotechnology* **7,** 374 (2012).

254. Deutscher, M. P. Degradation of Stable RNA in Bacteria. *Journal of Biological Chemistry* **278,** 45041–45044 (2003).

255. Filonov, G. S., Kam, C. W., Song, W. & Jaffrey, S. R. In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chemistry and Biology* **22,** 649–60 (2015).

256. Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength. *ACS Synthetic Biology* **3,** 363–371 (2014).

257. Lyubchenko, Y. L. & Shlyakhtenko, L. S. AFM for analysis of structure and dynamics of DNA and protein–DNA complexes. *Methods* **47,** 206–213 (2009).

258. Dai, K. & Lutkenhaus, J. The proper ratio of FtsZ to FtsA is required for cell division to occur in Escherichia coli. *Journal of Bacteriology* **174,** 6145–6151 (1992).

259. Sanchez-Gorostiaga, A. *et al.* Life without Division: Physiology of Escherichia coli FtsZ-Deprived Filaments. *mBio* **7,** e01620–16 (2016).

260. Yu, X. C. & Margolin, W. FtsZ ring clusters in min and partition mutants: Role of both the Min system and the nucleoid in regulating FtsZ ring localization. *Molecular Microbiology* **32,** 315–326 (1999).

261. Den Blaauwen, T., Hamoen, L. W. & Levin, P. A. The divisome at 25: the road ahead. *Current Opinion in Microbiology* **36,** 85–94 (2017).

262. Bi, E. F. & Lutkenhaus, J. FtsZ ring structure associated with division in Escherichia coli. *Nature* **354,** 161–4 (1991).

263. Hale, C. A. & de Boer, P. A. J. Direct Binding of FtsZ to ZipA, an Essential Component of the Septal Ring Structure That Mediates Cell Division in E. coli. *Cell* **88,** 175–185 (1997).

264. Pichoff, S. & Lutkenhaus, J. Unique and overlapping roles for ZipA and FtsA in septal ring assembly in Escherichia coli. *EMBO Journal* **21,** 685–93 (2002).

265. Du, S. & Lutkenhaus, J. Assembly and activation of the Escherichia coli divisome. *Molecular Microbiology* **105,** 177–187 (2017).

266. Ward J. E., J. & Lutkenhaus, J. Overproduction of FtsZ induces minicell formation in E. coli. *Cell* **42,** 941–9 (1985).

267. Begg, K. J. & Donachie, W. D. Cell shape and division in Escherichia coli: experiments with shape and division mutants. *Journal of Bacteriology* **163,** 615–22 (1985).

268. Dai, K. & Lutkenhaus, J. ftsZ is an essential cell division gene in Escherichia coli. *Journal of Bacteriology* **173,** 3500–3506 (1991).

269. Addinall, S. G., Bi, E. & Lutkenhaus, J. FtsZ ring formation in fts mutants. *Journal of Bacteriology* **178,** 3877–3884 (1996).

270. Ricard, M. & Hirota, Y. Process of cellular division in Escherichia coli: physiological study on thermosensitive mutants defective in cell division. *Journal of Bacteriology* **116,** 314–22 (1973).

271. Lutkenhaus, J. F., Wolf-Watz, H. & Donachie, W. D. Organization of genes in the ftsA-envA region of the Escherichia coli genetic map and identification of a new fts locus (ftsZ). *Journal of Bacteriology* **142,** 615–20 (1980).

272. Addinall, S. G., Cao, C. & Lutkenhaus, J. Temperature shift experiments with an ftsZ84(Ts) strain reveal rapid dynamics of FtsZ localization and indicate that the Z ring is required throughout septation and cannot reoccupy division sites once constriction has initiated. *Journal of Bacteriology* **179,** 4277–4284 (1997).

273. Van den Ent, F., Amos, L. A. & Lowe, J. Prokaryotic origin of the actin cytoskeleton. *Nature* **413,** 39–44 (2001).

274. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31,** 3406–15 (2003).

275. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry* **32,** 170–3 (2011).

276. Athale, C. A. & Chaudhari, H. Population length variability and nucleoid numbers in Escherichia coli. *Bioinformatics* **27,** 2944–8 (2011).

277. Aldea, M., Garrido, T., Pla, J. & Vicente, M. Division genes in Escherichia coli are expressed coordinately to cell septum requirements by gearbox promoters. *EMBO Journal* **9,** 3787–94 (1990).

278. Dertinger, S. K. W., Chiu, D. T., Jeon, N. L. & Whitesides, G. M. Generation of Gradients Having Complex Shapes Using Microfluidic Networks. *Analytical Chemistry* **73,** 1240–1246 (2001).

279. Koch, A. L. The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly. *Journal of Theoretical Biology* **12,** 276–290 (1966).

280. Wehrens, M. *et al.* Size Laws and Division Ring Dynamics in Filamentous Escherichia coli cells. *Current Biology* **28,** 972–979 e5 (2018).

281. Zheng, H. *et al.* Interrogating the Escherichia coli cell cycle by cell dimension perturbations. *Proceedings of the National Academy of Sciences of the United States of America* **113,** 15000–15005 (2016).

282. Flardh, K., Palacios, P. & Vicente, M. Cell division genes ftsQAZ in Escherichia coli require distant cis-acting signals upstream of ddlB for full expression. *Molecular Microbiology* **30,** 305–15 (1998).

283. Boyle, E. A. *et al.* High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proceedings of the National Academy of Sciences of the United States of America* **114,** 5461–5466 (2017).

284. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–67 (2014).

285. Tomizawa, J. Control of ColE1 plasmid replication: binding of RNA I to RNA II and inhibition of primer formation. *Cell* **47,** 89–97 (1986).

286. Stueber, D. & Bujard, H. Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. *EMBO Journal* **1,** 1399–1404 (1982).

287. Ramalho, T. *et al.* Single Cell Analysis of a Bacterial Sender-Receiver System. *PLoS One* **11,** e0145829 (2016).

Appendix

DNA, RNA and protein sequences for Ch. 4

The sequences used in the constructs for the experiments shown in Chapter 4.

RNA sequences

MG aptamer	GGAUCCCGA CUGGCGAGAG CCAGGUAACG AAUGGAUCC
TAR aptamer	GGCUCGUGUAGCUCAUUAGCUCCGAGCC
PP7 Aptamer	GGCACAGAAGAUAUGGCUUCGUGCC
MS2 Aptamer	GCGCACAUGAGGAUCACCCAUGUGC
Streptavidin Aptamer	AUGCGGCCGCCGACCAGAAUCAUGCAAGUG CGUAAGAUAGUCGCGGGUCGGCGGCCGCAU
pRNA-3WJ	1st Part: TTGTCATGTGTATGTTGGG 2nd Part: CCCACATACTTTGTTGATCC 3rd Part: GGATCAATCATGGCAA
RNA nanostructure	GGGTTGTCATGTGTATGTTGGGGGATCCCGAC TGGCGAGAGCCAGGTAACGAATGGATCCCCC ACATACTTTGTTGATCCCGCAGGATTCGGCTC GTGTAGCTCATTAGCTCCGAGCCGAGTCCTCGAA TACGAGCTGGGCACAGAAGATATGGCTTCGTGCCCA GGAGGTGTTCGCACTTCTCTCGTGTTC GATTGTGGGATCAATCATGGCAAATGCGGCCG CCGACCAGAATCATGCAAGTGCGTAAGATAGTCG CGGGTCGGTGGTCGCAT
TAR-PP7 RNA	GGGCGCAGGATTCGGCTCGTGTAGCT CATTAGCTCCGAGCCGAGTCCTCGAATACGA GCTGGGCACAGAAGATATGGCTTCGTGCCCAG GAAGTGTTCGCACTTCTCTCGTATTCGATTG CGTGATACCTACGAC
MS2-PP7 RNA	GGGCGCAGGACTCGCGCACATGAGGAT CACCCATGTGCGAGTCCTCGAATACGAGCTGG GCACAGAAGATATGGCTTCGTGCCCAGGAAGT GTTCGCACTTCTCTCGTATTCGATTGCGTGA TACCTACGAC

Peptide sequences

PP7 Coat Protein	MSKTIVLSVGEATRTL
	TEIQSTADRQIFEEKVGPLVGRL
	RLTASLRQNGAKTAYRVNLKLDQAD
	VVDSGLPKVRYTQVWSHDVTIVANS
	TEASRKSLYDLTKSLVATSNVE
	DLVVNLVPLGRK
Tat peptide	SGPRPRGTRGKGRRIRR
MS2 Coat Protein	MASNFTQFVLVDNGGT
	GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQ
	SSAQNRKYTIKVEVPKGAWRSYLNMEL
	TIPIFATNSDCELIVKAMQGLL
	KDGNPIPSAIAANSGIY
Peptide-Linker	TRGGGGS
mTurquoiuse2	MVSKGEELFTGVVPIL
	VELDGDVNGHKFSVSGEGEGDATYG
	KLTLKFICTTGKLPVPWPTLVTTLSWGVQC
	FARYPDHMKQHDFFKSAMPEGYVQERTIFF
	KDDGNYKTRAEVKFEGDTLVNRIELKGIDF
	KEDGNILGHKLEYNYFSDNVYITADKQKNGI
	KANFKIRHNIEDGGVQLADHYQQNTPIG
	DGPVLLPDNHYLSTQSKLSKDPNEKRDHMV
	LLEFVTAAGITLGMDELYK
Ypet	MSKGEELFTGVVPILVELDGD
	VNGHKFSVSGEGEGDATYGKLTLKLLCTTG
	KLPVPWPTLVTTLGYGVQCFARYP
	DHMKQHDFFKSAMPEGYVQERTIFFKD
	DGNYKTRAEVKFEGDTLVNRIE
	LKGIDFKEDGNILGHKLEYNYNSHNVYITA
	DKQKNGIKANFKIRHNIEDGGVQL
	ADHYQQNTPIGDGPVLLPDNHYLSYQS
	ALFKDPNEKRDHMVLLEFLTAAGI
	TEGMNELYK

Cy-plasmid

cgccccgccctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcaca
aacggcatgatgaAcctgAatcgccagcggcatcagcaccttgtcgccttgcgtataatatttgcccatgg
tgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccaggga
ttggctgagacgaaaaacatattctcaataaacccctttaggGaaataggccaggttttcaccgtaacacgc
cacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccAgagcgatgaaaacg
tttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctttcA
ttgccatacgaaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttg
tgcttatttttctttAcggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgag
caactgactgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtga
ttttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgatctt
atttcattatggtgaaagttggaacctcttacgtgcccgatcaactcgagtgccacctgacgtctaagaaac
cattattatcatgacattaacctataaaaataggcgtatcacgaggcagaatttcagataaaaaaaatcctt
agctttcgctaaggatgatttctggaattcgagtaagccgcttctagagcaccgattgagaagagaaaaga
aaaccgccgatcctgtccaccgcattactgcaaggtagtggacaagaccggcggtcttaagtttttggctg
aaagtagtaccattaacgacgaatgcgacgacccttaccgcgggtaccacgaggacgcggaccgctgctgc
cgccgccaccacgggtcttgtacagttcgtccatacccagcgtgatacccgccgcggtaacaaattccagca
ggaccatgtggtcacgtttttcgttcggatctttggacagtttggattgcgtgctcaagtagtgattatctgg
cagcaggaccggaccatcgccaatcggggtgttctgttgatagtgatcggccaattgcacgccgccatcttc
aatattgtgacggatcttgaagtttgctttaatgccgtttttctgtttatcagcggtgatgtacacattgtcgc
tgaagtaattatactccagtttgtgacccaggatgttaccgtcctctttaaagtcaatacctttcagctcaat
acgattcaccagcgtgtcaccctcgaacttgacttctgcgcgggtcttatagttaccgtcgtctttgaaaaag
atcgtacgttcctgaacgtagccttccggcacgcgctcttaaagaagtcatgctgtttcatgtgatccggata
gcgcgcgaaacattgcacgccccacgacagagtcgtcaccagggcaccacgggtaaattgccggttccgtg
gccgaccctggtcacgacgctgggttatggtgtacaatgttttgcacgctatccggaccacatgaaacagca
cgatttcttcaagagcgcgatgccggaaggctatgttcaggaacgtaccatctttttcaaagatgatggtaa
ttacaaaacccgcgcagaagtgaagttcgagggtgacaccctggtgaaccgtattgagctgaagggtattg
acttcaaggaagatggcaatattctgggtcacaaactggagtacaactataacagccataacgtctacatc
accgcggataagcaaaaaaatggtatcaaagcaaatttcaagattcgccacaacatcgaagatggcggcg
tgcaactggccgatcattatcagcagaataccccaatcggtgacggtccggtgctgttgccggataaccact
acctgagctatcaaagcgcgttgttcaaagacccgaatgaaaaacgtgaccacatggttctgctggaatttc
tgaccgctgcgggcatcactgaaggcatgaatgaactgtacaagacgcgtggtggcggcggttcgatgagc
aagactatcgttttgtccgtcggcgaggctacccgtaccttgaccgaaattcaatccaccgcggaccgtcaa
atttttgaggaaaaagtcggtcctctggtgggtcgtctgcgtctgaccgcgagcctgcgccagaacggtgcc
aaaacggcataccgtgttaatctgaaactggatcaggccgacgttgtggacagcggtctgccgaaagtccg
ctacacccaggtgtggagccacgatgtgacgatcgttgcgaatagcaccgaagcgagccgcaagagcctgt
acgacctgaccaagagcctggtggcaacgtcccaagttgaagatctggttgttaacctggtgccgctgggtc
gttaaagcatgccggaggaaacacagaaaaaagcccgcacctgacagtgcgggctttttttttcgaccaaa
ggggtgcatactagtagcggccgctgcagtccggcaaaaaagggcaaggtgtcaccaccctgccctttttct
ttaaaaccgaaaagattacttcgcgttatgcaggcttcctcgctcactgactcgctgcgctcggtcgttcggc

tgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcagg
aaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttc
cacaggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacag
gactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttac
cggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagt
tcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttа
tccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaac
aggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacac
tagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttg
atccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaa
aaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaag
ggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatc
aatctaaagtatatatgagtaaacttggtctgacagctcgaggcttggattctcaccaataaaaaacgcccg
gcggcaaccgagcgttctgaacaaatccagatggagttctgaggtcattactggatctatcaacaggagtc
caagcgagctcgatatcaaattagtcggccacggaaccggcagcttaccggtcgtgcagataaacttcagg
gtcaatttaccgtaggtggcgtcgccctcgccctcaccgctaacgctaaacttgtgaccgttgacatcaccat
ccagctcaaccagaatcggcacgacgccggtgaacagttcttcacccttgctaaccatgtctatttctcctct
ttactgactttaaacaaaattatttgtagaggtcataagtttcgtcgcatttcagcgactcatcagtgggctt
agcccagacttatgactgaagcttcgctagcactatacctaggactgagctagccgtaaagatcctgaatct
cgagattggtgccttgcttggattatgaaatgccagcctgatcatacaacgttggacgaagtagctttaccg
tccacacggtcggaacttggtgtgccgaatccaagtcgagataatcactagatgtcctgaggtgcatactag
aggaccacgcatcgtgatgcctatgcgcggtagtcccaccttgtccactagaatggaagattggcacgtatc
aagactttggagtagtaccataacgccgtttacggctagctcagtcctaggtatagtgctagctagcgcagc
gctcaacgggtgtgcttcccgttctgatgagtccgtgaggacgaaagcgcctctacaaataattttgtttaat
catgagaaagaggagaaaactagatgtctaaaggtgaagaactgtttacgggtgtcgtgccgattctggtc
gagttggacggcgacgtgaacggtcacaaattcagcgtgagcggcgagggcgagggtgacgcgacgtacg
gtaagctgactctgaagctgctgt

r-plasmid

ggtagttattttcaagcattaacatgaacttaaattcatcaaggctaatctctatatttgccttgtgagttttc
ttttgtgttagttcttttaataaccactcataaatcctcatagagtatttgttttcaaaagacttaacatgttc
cagattatattttatgaatttttttaactggaaaagataaggcaatatctcttcactaaaaactaattctaat
ttttcgcttgagaacttggcatagtttgtccactggaaaatctcaaagcctttaaccaaaggattcctgattt
ccacagttctcgtcatcagctctctggttgctttagctaatacaccataagcattttccctactgatgttcatc
atctgagcgtattggttataagtgaacgataccgtccgttctttccttgtagggttttcaatcgtggggttga
gtagtgccacacagcataaaattagcttggtttcatgctccgttaagtcatagcgactaatcgctagttcat
ttgctttgaaaacaactaattcagacatacatctcaattggtctaggtgattttaatcactataccaattgag
atgggctagtcaatgataattacatgtccttttcctttgagttgtgggtatctgtaaattctgctagacctttg
ctggaaaacttgtaaattctgctagaccctctgtaaattccgctagacctttgtgtgttttttttgtttatattc
aagtggttataatttatagaataaagaaagaataaaaaaagataaaaagaatagatcccagccctgtgta

taactcactactttagtcagttccgcagtattacaaaaggatgtcgcaaacgctgtttgctcctctacaaaac
agaccttaaaaccctaaaggcttaagtagcaccctcgcaagctcgggcaaatcgctgaatattccttttgtc
tccgaccatcaggcacctgagtcgctgtctttttcgtgacattcagttcgctgcgctcacggctctggcagtg
aatgggggtaaatggcactacaggcgccttttatggattcatgcaaggaaactacccataatacaagaaaa
gcccgtcacgggcttctcagggcgttttatggcgggtctgctatgtggtgctatctgactttttgctgttcagc
agttcctgccctctgattttccagtctgaccacttcggattatcccgtgacaggtcattcagactggctaatg
cacccagtaaggcagcggtatcatcaacaggcttacccgtcttactgtccctagtgcttggattctcaccaat
aaaaaacgcccggcggcaaccgagcgttctgaacaaatccagatggagttctgaggtcattactggatcta
tcaacaggagtccaagcgagctcgtaaacttggtctgacagctctagctccggcaaaaaaacgggcaaggt
gtcaccaccctgccctttttctttaaaaccgaaaagattacttcgcgtttgccacctgacgtctaagaaaagg
aatattcagcaatttgcccgtgccgaagaaaggcccacccgtgaaggtgagccagtgagttgattgctacg
taattagttagttagcccttagtgactcgaattcgcggccgcttctagagcacaggccacagtaatacgact
cactatagggttgtcatgtgtatgttgggggatcccgactggcgagagccaggtaacgaatggatccccca
catactttgttgatcccgcaggattcggctcgtgtagctcattagctccgagccgagtcctcgaatacgagct
gggcacagaagatatggcttcgtgcccaggaggtgttcgcacttctctcgtgttcgattgtgggatcaatca
tggcaaatgcggccgccgaccagaatcatgcaagtgcgtaagatagtcgcgggtcggtggtcgcattactt
agcataaccccgcggggcctcttcgggggtctcgcggggtttttgctgaaacggctgctaacaaagcccga
aaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggt
cttgaggggtttttgctgaaaggaggaactacactgtctgcaggagtcactaagggttagttagttagatt
agcagaaagtcaaaagcctccgaccggaggcttttgactaaaacttcccttggggttatcattggggctcac
tcaaaggcggtaatcagataaaaaaaatccttagctttcgctaaggatgatttctgctagtattattagaaa
aactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagc
cgtttctgtaatgaaggagaaaactcaccgaggcagttccaaagaatggcaaggtcctggtaacggtctgc
gattccgacccgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaa
atcaccatgagtgacgactgaatccggtgagaatggcaagagcttgtgcatttctttccagacttgttcaac
aggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcatgcgtgattgcgcctg
agcaagacgaaatacacgatcgctgttaaaaggacaattacaaacaggaatcgaatgtaaccggcgcagg
aacacggccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaaggctgttttcc
caggaatcgcggtggtgagtaaccacgcatcatcaggagtacggataaaatgcttgatggtcgggagagg
cataaactccgtcagccagttgagacggaccatctcatctgtaacatcattggcaacgctacctttgccatgt
ttcagaaacaactctggcgcatcgggcttcccatacaagcgatagattgtcgcacctgattgcccgacatta
tcgcgagcccatttatacccatataaatcagcgtccatgttggagtttaagcgcggacgggagcaagacgtt
tcccgttgaatatggctcataacacccttgtattactgtttatgtaagcagacagttttattgttcatgatga
tatatttttatcttgtgcaatgtaacatcagagattttgagacacaacgtggctttgttgaataaatcgaact
tttgctgagttgaaggatcagctctagtagttacattgtcgatctgttcatggtgaacagctttgaatgcacc
aaaaactcgtaaaagctctgatgtatctatctttttacaccgttttcatctgtgcatatggacagttttccct
ttgatatgtaacggtgaacagttgttctacttttgtttgttagtcttgatgcttcactgatagatacaagagcc
ataagaacctcagatccttccgtatttagccagtatgttctctagtgtggttcgttgtttttgcgtgagccatg
agaacgaaccattgagatcatacttactttgcatgtcactcaaaaattttgcctcaaaactggtgagctgaa
tttttgcagttaaagcatcgtgtagtgtttttcttagtccgttatgtaggtaggaatctgatgtaatggttgtt
ggtattttgtcaccattcatttttatctggttgttctcaagttcggttacgagatccatttgtctatctagttca

acttggaaaatcaacgtatcagtcgggcggcctcgcttatcaaccaccaatttcatattgctgtaagtgttta
aatctttacttattggtttcaaaacccattggttaagccttttaaactcat

Sequences for Ch. 5

Plasmid maps

CRISPRi plasmid

attcgcaaacgccctctaatcgaaactaatggggaaactggagaaattgtctgggataaagggcgagattt
tgccacagtgcgcaaagtattgtccatgccccaagtcaatattgtcaagaaaacagaagtacagacaggcg
gattctccaaggagtcaattttaccaaaaagaaattcggacaagcttattgctcgtaaaaaagactgggat
ccaaaaaaatatggtggttttgatagtccaacggtagcttattcagtcctagtggttgctaaggtggaaaaa
gggaaatcgaagaagttaaaatccgttaaagagttactagggatcacaattatggaaagaagttcctttga
aaaaaatccgattgacttttagaagctaaaggatataaggaagttaaaaaagacttaatcattaaactac
ctaaatatagtctttttgagttagaaaacggtcgtaaacggatgctggctagtgccggagaattacaaaaa
ggaaatgagctggctctgccaagcaaatatgtgaatttttatatttagctagtcattatgaaaagttgaag
ggtagtccagaagataacgaacaaaaacaattgtttgtggagcagcataagcattatttagatgagattat
tgagcaaatcagtgaattttctaagcgtgttattttagcagatgccaatttagataaagttcttagtgcatat
aacaaacatagagacaaaccaatacgtgaacaagcagaaaatattattcatttatttacgttgacgaatct
tggagctcccgctgcttttaaatattttgatacaacaattgatcgtaaacgatatacgtctacaaaagaagt
tttagatgccactcttatccatcaatccatcactggtctttatgaaacacgcattgatttgagtcagctagga
ggtgactaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttt
cgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctg
cgtttatacctagggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcgag
cggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgag
agggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaat
cagtggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcggctccctcgtgcgctct
cctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacact
cagttccgggtaggcagttcgctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgc
cttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactg
gtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggt
gactgcgctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccc
tgcaaggcggttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatctt
attaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccca
tacgatataagttgttactagttcttacccgcaaaaaaccccgcttcggcggggttttttcgcgcttttaattt
accaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcccc
gtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccc
acgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctg
caactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatag
tttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagct
ccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtc

ctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctct
tactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgt
atgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaa
agtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttc
gatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaa
acaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcc
ttttcaatcatgattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttaga
aaaataaacaaattaattaataagccggcactaaggtaattccttatgctggatatatacgtgttccagttc
ttatcctatatcgaaagttcaattatacatgtctgcgtcgtcgatatcggtggcgtaccagtgtgttcgtgtc
atctaggaggggcgcgtaggataaataatttttcatacgcaagtgcgtatgccggcatacgcacttgcgtat
gaaaaattatttatatggtaatacgactcactatagggtttcatacgcaagtgcgtatgtttaagagctatg
ctggaaacagcatagcaagtttaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtg
ctttttttaagtctctagactcgcacatgagcaaaacggtagagcgtcactttcactagtgtcgcggccgcta
atacgactcactatagggactaaggtaattccttatgcgtttaagagctatgctggaaacagcatagcaagt
ttaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttttaagtctctagagtt
gatccggattgactctatgacggtacctgctgactagtgcgcctgcaggatggtaatacgactcactatagg
ggtctgcgtcgtcgatatcgggtttaagagctatgctggaaacagcatagcaagtttaaataaggctagtc
cgttatcaacttgaaaaagtggcaccgagtcggtgctttttttttgcgctgcagccaggcatcaaataaaac
gaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtca
cactggctcaccttcgggtgggcctttctgcgtttatatttctagagtatggactagtgcttggattctcacca
ataaaaaacgcccggcggcaaccgagcgttctgaacaaatccagatggagttctgaggtcattactggatc
tatcaacaggagtccaagcgagctcgatatcaaaTCCCTATCAGTGATAGAGATTGA
CATCCCTATCAGTGATAGAGATACTGAGCACGGGTCTATACTAG
AGAAAGAGGAGAAATACTAGATGAGCAAAGGCGAAGAACTGTTC
ACGGGTGTGGTTCCGATCCTGGTTGAACTGGATGGCGATGTGAA
CGGTCATAAATTTAGCGTGTCTGGTGAAGGCGAAGGTGATGCGA
CCTACGGCAAACTGACGCTGAAACTGATTTGCACCACGGGTAAA
CTGCCGGTTCCGTGGCCGACCCTGGTGACCACGCTGGGTTATGG
TCTGATGTGTTTCGCACGTTACCCGGATCACATGAAACGCCATG
ATTTCTTTAAATCTGCGATGCCGGAAGGCTATGTGCAGGAACGT
ACCATCTTTTTCAAAGATGATGGTAACTACAAAACCCGCGCGGA
AGTTAAATTTGAAGGCGATACGCTGGTGAACCGTATTGAACTGA
AAGGTATCGATTTCAAAGAAGATGGCAATATTCTGGGTCACAAA
CTGGAATACAACTACAACAGTCATAACGTGTACATTACCGCCGA
TAAACAGAAAAACGGTATCAAAGCAAACTTCAAAATCCGTCACA
ACATCGAAGATGGCGGTGTTCAGCTGGCCGATCATTACCAGCAG
AACACCCCGATTGGCGATGGTCCGGTGCTGCTGCCGGATAATCA
TTATCTGAGTTACCAGAGCAAACTGTCTAAAGATCCGAATGAAA
AACGCGATCACATGGTTCTGCTGGAATTTGTGACCGCGGCCGGC
ATTACGCATGGTATGGATGAACTGTATAAAtaacgtctcattttcgccagatat
cgacgtcttaagacccactttcacatttaagttgtttttctaatccgcatatgatcaattcaaggccgaataa

gaaggctggctctgcaccttggtgatcaaataattcgatagcttgtcgtaataatggcggcatactatcagt
agtaggtgtttccctttcttctttagcgacttgatgctcttgatcttccaatacgcaacctaaagtaaaatgcc
ccacagcgctgagtgcatataatgcattctctagtgaaaaaccttgttggcataaaaaggctaattgatttt
cgagagtttcatactgtttttctgtaggccgtgtacctaaatgtacttttgctccatcgcgatgacttagtaaa
gcacatctaaaacttttagcgttattacgtaaaaaatcttgccagctttccccttctaaagggcaaaagtga
gtatggtgcctatctaacatctcaatggctaaggcgtcgagcaaagcccgcttatttttacatgccaataca
atgtaggctgctctacacctagcttctgggcgagtttacgggttgttaaaccttcgattccgacctcattaag
cagctctaatgcgctgttaatcactttacttttatctaatctagacatcattaattcctaatttttgttgacact
ctatcgttgatagagttattttaccactccctatcagtgatagagaaaagaattcaaaagatctaaagagga
gaaaggatctatggataagaaatactcaataggcttagctatcggcacaaatagcgtcggatgggcggtg
atcactgatgaatataaggttccgtctaaaaagttcaaggttctgggaaatacagaccgccacagtatcaa
aaaaaatcttataggggctcttttatttgacagtggagagacagcggaagcgactcgtctcaaacggacag
ctcgtagaaggtatacacgtcggaagaatcgtatttgttatctacaggagatttttttcaaatgagatggcga
aagtagatgatagtttctttcatcgacttgaagagtctttttttggtggaagaagacaagaagcatgaacgtc
atcctatttttggaaatatagtagatgaagttgcttatcatgagaaatatccaactatctatcatctgcgaaa
aaaattggtagattctactgataaagcggatttgcgcttaatctatttggccttagcgcatatgattaagttt
cgtggtcatttttttgattgagggagatttaaatcctgataatagtgatgtggacaaactatttatccagttgg
tacaaacctacaatcaattatttgaagaaaaccctattaacgcaagtggagtagatgctaaagcgattcttt
ctgcacgattgagtaaatcaagacgattagaaaatctcattgctcagctccccggtgagaagaaaaatggc
ttatttgggaatctcattgctttgtcattgggtttgacccctaattttaaatcaaattttgatttggcagaaga
tgctaaattacagctttcaaaagatacttacgatgatgatttagataatttattggcgcaaattggagatca
atatgctgatttgtttttggcagctaagaatttatcagatgctattttactttcagatatcctaagagtaaata
ctgaaataactaaggctcccctatcagcttcaatgattaaacgctacgatgaacatcatcaagacttgactc
ttttaaaagctttagttcgacaacaacttccagaaaagtataaagaaatctttttgatcaatcaaaaaacg
gatatgcaggttatattgatgggggagctagccaagaagaattttataaatttatcaaaccaattttagaaa
aaatggatggtactgaggaattattggtgaaactaaatcgtgaagatttgctgcgcaagcaacggaccttt
gacaacggctctattccccatcaaattcacttgggtgagctgcatgctattttgagaagacaagaagacttt
tatccatttttaaaagacaatcgtgagaagattgaaaaaatcttgacttttcgaattccttattatgttggtc
cattggcgcgtggcaatagtcgttttgcatggatgactcggaagtctgaagaaacaattaccccatggaatt
ttgaagaagttgtcgataaaggtgcttcagctcaatcatttattgaacgcatgacaaactttgataaaaatc
ttccaaatgaaaaagtactaccaaaacatagtttgctttatgagtattttacggtttataacgaattgacaa
aggtcaaatatgttactgaaggaatgcgaaaaccagcatttctttcaggtgaacagaagaaagccattgtt
gatttactcttcaaaacaaatcgaaaagtaaccgttaagcaattaaaagaagattatttcaaaaaaataga
atgttttgatagtgttgaaatttcaggagttgaagatagatttaatgcttcattaggtacctaccatgatttg
ctaaaaattattaaagataaagatttttggataatgaagaaaatgaagatatcttagaggatattgtttta
acattgaccttatttgaagatagggagatgattgaggaaagacttaaaacatatgctcacctctttgatgat
aaggtgatgaaacagcttaaacgtcgccgttatactggttggggacgtttgtctcgaaaattgattaatggt
attagggataagcaatctggcaaaacaatattagatttttgaaatcagatggttttgccaatcgcaatttt
atgcagctgatccatgatgatagtttgacatttaaagaagacattcaaaaagcacaagtgtctggacaagg
cgatagtttacatgaacatattgcaaatttagctggtagccctgctattaaaaaaggtattttacagactgt
aaaagttgttgatgaattggtcaaagtaatggggcggcataagccagaaaatatcgttattgaaatggcac

gtgaaaatcagacaactcaaaagggccagaaaaattcgcgagagcgtatgaaacgaatcgaagaaggta
tcaaagaattaggaagtcagattcttaaagagcatcctgttgaaaatactcaattgcaaaatgaaaagctc
tatctctattatctccaaaatggaagagacatgtatgtggaccaagaattagatattaatcgtttaagtgat
tatgatgtcgatgccattgttccacaaagtttccttaaagacgattcaatagacaataaggtcttaacgcgt
tctgataaaaatcgtggtaaatcggataacgttccaagtgaagaagtagtcaaaaagatgaaaaactattg
gagacaacttctaaacgccaagttaatcactcaacgtaagtttgataatttaacgaaagctgaacgtggag
gtttgagtgaacttgataaagctggttttatcaaacgccaattggttgaaactcgccaaatcactaagcatg
tggcacaaattttggatagtcgcatgaatactaaatacgatgaaaatgataaacttattcgagaggttaaa
gtgattaccttaaaatctaaattagtttctgacttccgaaaagatttccaattctataaagtacgtgagatta
acaattaccatcatgcccatgatgcgtatctaaatgccgtcgttggaactgctttgattaagaaatatccaa
aacttgaatcggagtttgtctatggtgattataaagtttatgatgttcgtaaaatgattgctaagtctgagca
agaaataggcaaagcaaccgcaaaatatttcttttactctaatatcatgaacttcttcaaaacagaaattac
acttgcaaatggagag

anti-sgRNA

AGTTCTTATCCTATATCGAAAGTTCAATTATACATGTCTGCGTC
GTCGATATCGGTGGCGTACCAGTGTGTTCGTGTCATCTAGGAGG
GGCGCGTAGGATAAATAATTTTTCATACGCAAGTGCGTATCGGct
gcagtccggcaaaaaagggcaaggtgtcaccaccctgccctttttctttaaaaccgaaaagattacttcgcg
ttatgcaggcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactc
aaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccca
gcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccacaggctccgccccctgacgagc
atcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcc
ccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctccctt
cgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagct
gggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaa
cccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtag
gcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcg
ctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggta
gcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatct
tttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaa
ggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttg
gtctgacagctcgagtcccgtcaagtcagcgtaatgctctgccagtgttacaaccaattaaccaattctgatt
agaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaa
aaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcgg
tctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtg
agaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatttctttccagacttgtt
caacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcg
cctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcg
caggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgt

tttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaa
gaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgcc
atgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgac
attatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctggagcaaga
cgtttcccgttgaatatggctcataacacccttgtattactgtttatgtaagcagacagttttattgttcatg
atgatatatttttatcttgtgcaatgtaacatcagagattttgagacacaacgtggctttgttgaataaatcg
aactttgctgagttgaaggatcagctcgagtgccacctgacgtctaagaaaccattattatcatgacatta
acctataaaaataggcgtatcacgaggcagaatttcagataaaaaaaatccttagctttcgctaaggatga
tttctgGAATTCATAAAAAACGCCCGGCGGCAACCGAGCGTTACCTG
TAGGATCGTACAGGTTTACGCAAGAAAATGGTTTGTTATAGTCG
AATAAACACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAAA
CTTGCTATGCTGTTTCCAGCATAGCTCTTAAACGCATAAGGAAT
TACCTTAGTCCGCAAAAAACCCCGCTTCGGCGGGGTTTTTTCGC
CGACGGTGTAGGCACGTAAGAGGTTCCAACTTTCACCATAATGA
AACAGGGTCTATACTAGAGAAAGAGGAGAAATACTAGATGAAGA
ACATCAACGCGGATGATACGTACCGTATTATCAACAAGATTAAA
GCCTGCAGTTCGAACAACGATATCAACCAGTGCTTATCCGACAT
GACCAAAATGGTACATTGTGAGTACTACCTGCTGGCGATCATTT
ATCCGCATAGCATGGTAAAATCAGATATTAGCATTCTCGACAAC
TATCCCAAGAAATGGCGTCAGTATTATGATGACGCAAATCTCAT
CAAGTACGATCCGATCGTAGATTATTCTAACTCCAATCATTCAC
CCATTAACTGGAATATCTTTGAAAACAACGCTGTTAACAAAAAA
AGTCCCAATGTCATTAAAGAGGCCAAAACCTCGGGCCTGATTAC
CGGATTTTCCTTTCCAATCCATACGGCGAATAACGGCTTTGGCA
TGCTCTCTTTCGCACACAGTGAGAAGGACAACTATATTGATTCC
CTTTTTCTTCATGCGTGCATGAATATTCCCCTGATCGTCCCGTCT
CTGGTGGACAATTACCGCAAGATCAATATTGCTAATAACAAAAG
TAATAACGACCTTACGAAACGCGAGAAAGAGTGCCTGGCGTGGG
CGTGCGAAGGGAAAAGTTCGTGGGATATCTCTAAGATTCTTGGC
TGTTCGGAGCGCACTGTGACCTTCCATCTGACGAACGCCCAAAT
GAAATTAAACACGACGAATCGTTGCCAGTCCATCTCCAAAGCCA
TTCTGACGGGCGCAATCGACTGTCCGTATTTTAAAAATTAACCA
GGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTT
CGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCA
CACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATATAAGTA
GGGTCGACCAAGAACCGCAAGATGACCTGTAGGATCGTACAGGT
TTACGCAAGAAAATGGTTTGTTATAGTCGAATAAACAAGTAATT
GTCAACAAACTCCAGCATAGCTCTTAAACCCGATATCGACGACG
CAGACTTTTTTTTTTCCAGCATAGCTCTTAAACATACGCACTTGC
GTATGAAACAGAGAATATAAAAAGCCAGATTATTAATCCGGCTT
TTTTATTATTTTCCAGATGGAGTTCTGAGGTCATTACTGGATCT

AGAgcaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttc
ccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggct
ttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacatactagagaa
agaggagaaatactagatggcttcctccgaagacgttatcaaagagttcatgcgtttcaaagttcgtatgg
aaggttccgttaacggtcacgagttcgaaatcgaaggtgaaggtgaaggtcgtccgtacgaaggtacccag
accgctaaactgaaagttaccaaaggtggtccgctgccgttcgcttgggacatcctgtccccgcagttccag
tacggttccaaagcttacgttaaacacccggctgacatcccggactacctgaaactgtccttcccggaaggt
ttcaaatgggaacgtgttatgaacttcgaagacggtggtgttgttaccgttacccaggactcctccctgcaa
gacggtgagttcatctacaaagttaaactgcgtggtaccaacttcccgtccgacggtccggttatgcagaaa
aaaaccatgggttgggaagcttccaccgaacgtatgtacccggaagacggtgctctgaaaggtgaaatca
aaatgcgtctgaaactgaaagacggtggtcactacgacgctgaagttaaaaccacctacatggctaaaaa
accggttcagctgccgggtgcttacaaaaccgacatcaaactggacatcacctcccacaacgaagactaca
ccatcgttgaacagtacgaacgtgctgaaggtcgtcactccaccggtgcttaataacgctgatagtgctagt
gtagatcgctactagagccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgtttta
tctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttat
atactagtACTAAGGTAATTCCTTATGCTGGATATATACGTGTTCC

mVenus plasmid

ggggaattgtgagcggataacaattcccctgtagaaataattttgtttaactttaataaggagatataccat
gggcagcagccatcaccatcatcaccacagccaggatccgAATTGGCGTCCCTATCAGT
GATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCACG
GGTCTATACTAGAGAAAGAGGAGAAATACTAGATGAGCAAAGGC
GAAGAACTGTTCACGGGTGTGGTTCCGATCCTGGTTGAACTGGA
TGGCGATGTGAACGGTCATAAATTTAGCGTGTCTGGTGAAGGCG
AAGGTGATGCGACCTACGGCAAACTGACGCTGAAACTGATTTGC
ACCACGGGTAAACTGCCGGTTCCGTGGCCGACCCTGGTGACCAC
GCTGGGTTATGGTCTGATGTGTTTCGCACGTTACCCGGATCACA
TGAAACGCCATGATTTCTTTAAATCTGCGATGCCGGAAGGCTAT
GTGCAGGAACGTACCATCTTTTTCAAAGATGATGGTAACTACAA
AACCCGCGCGGAAGTTAAATTTGAAGGCGATACGCTGGTGAACC
GTATTGAACTGAAAGGTATCGATTTCAAAGAAGATGGCAATATT
CTGGGTCACAAACTGGAATACAACTACAACAGTCATAACGTGTA
CATTACCGCCGATAAACAGAAAAACGGTATCAAAGCAAACTTCA
AAATCCGTCACAACATCGAAGATGGCGGTGTTCAGCTGGCCGAT
CATTACCAGCAGAACACCCCGATTGGCGATGGTCCGGTGCTGCT
GCCGGATAATCATTATCTGAGTTACCAGAGCAAACTGTCTAAAG
ATCCGAATGAAAAACGCGATCACATGGTTCTGCTGGAATTTGTG
ACCGCGGCCGGCATTACGCATGGTATGGATGAACTGTATAAAAG
GCCTGCAGCAAACGACGAAAACTACGCTTTAGTAGCTTTAACCA
GGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTT

CGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCA
CACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAACTAGT
GCGaagcttgcggccgcataatgcttaagtcgaacagaaagtaatcgtattgtacacggccgcataatc
gaaattaatacgactcactataggggaattgtgagcggataacaattccccatcttagtatattagttaagt
ataagaaggagatatacatatggcagatctcaattggatatcggccggccacgcgatcgctgacgtcggta
ccctcgagtctggtaaagaaaccgctgctgcgaaatttgaacgccagcacatggactcgtctactagcgca
gcttaattaacctaggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtc
ttgaggggttttttgctgaaacctcaggcatttgagaagcacacggtcacactgcttccggtagtcaataaa
ccggtaaaccagcaatagacataagcggctatttaacgaccctgccctgaaccgacgacaagctgacgacc
gggtctccgcaagtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattc
aaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattca
tatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagt
tccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaattt
cccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggca
aaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatc
aaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcggtcgctgttaaaaggacaat
tacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatca
ggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcagga
gtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatct
gtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatc
gatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgtt
ggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcatactcttccttttcaatattatt
gaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatag
gcatgctagcgcagaaacgtcctagaagatgccaggaggatacttagcagagagacaataaggccggagc
gaagccgtttttccataggctccgccccctgacgaacatcacgaaatctgacgctcaaatcagtggtggcg
aaacccgacaggactataaagataccaggcgtttcccccctgatggctccctcttgcgctctcctgttcccgtc
ctgcggcgtccgtgttgtggtggaggctttacccaaatcaccacgtcccgttccgtgtagacagttcgctcca
agctgggctgtgtgcaagaaccccccgttcagcccgactgctgcgccttatccggtaactatcatcttgagtc
caacccggaaagacacgacaaaacgccactggcagcagccattggtaactgagaattagtggatttagat
atcgagagtcttgaagtggtggcctaacagaggctacactgaaaggacagtatttggtatctgcgctccact
aaagccagttaccaggttaagcagttccccaactgacttaaccttcgatcaaaccgcctccccaggcggttt
tttcgtttacagagcaggagattacgacgatcgtaaaaggatctcaagaagatcctttacggattcccgaca
ccatcactctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagt
tgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcat
cggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccag
tcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgg
gcgccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgag
agagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggc
gggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccg
gactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgat
gccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatc

ggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaa
tgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtc
ttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattag
tgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgtt
gcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgc
tggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactg
gaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattc
agctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcggg
aaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccc
tgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggat
ctcgacgctctcccttatgcgactcctgcattaggaaattaatacgactcactata

target sites and sgRNA/anti-sgRNA for ftsZ

pZ2 target site	tttcatacgcaagtgcgtatcgg
pZ3 target site	actaaggtaattccttatgctgg
pZ4 target site	gtctgcgtcgtcgatatcggtgg
pZ2 sgRNA	gggtttcatacgcaagtgcg
	tatgtttaagagctatgctg
	gaaacagcatagcaagttta
	aataaggctagtccgttatc
	aacttgaaaaagtggcaccg
	agtcggtgctttttttt
pZ3 sgRNA	gggactaaggtaattcc
	ttatgcgtttaagagct
	atgctggaaacagcata
	gcaagtttaaataaggct
	agtccgttatcaacttgaa
	aaagtggcaccgagtcggt
	gctttttttt
pZ4 sgRNA	ggggtctgcgtcgtcgatat
	cgggtttaagagctatg
	ctggaaacagcatagca
	agtttaaataaggctag
	tccgttatcaacttgaaa
	aagtggcaccgagtcggtg
	ctttttttt
pZ3 anti-sgRNA	cacttttcaagttgataa
	cggactagccttatttaaa
	cttgctatgctgtttccag
	catagctcttaaacgcata
	aggaattaccttagtccgca
	aaaaaccccgcttcggcggg
	gttttttcgc
pZ4-pZ2 anti-sgRNA	caagtaattgtcaacaaact
	ccagcatagctcttaaacccg
	atatcgacgacgcagactttt
	ttttttccagcatagctctta
	aacatacgcacttgcgtatga
	aacagagaatataaaaagcca
	gattattaatccggctttttt
	attattt

sgRNA/anti-sgRNA for mVenus

mVenus-sgRNA	GGGCCAUCCAGUUCAACCAGG
	AUGUUUAAGAGCUAUGCUGGAAA
	CAGCAUAGCAAGUUUAAAUAAG
	GCUAGUCCGUUAUCAACUUGAA
	AAAGUGGCACCGAGUCGGUCG
	UUUUUUUACACUGUCUGCAGUCC
mVenus anti-sgRNA(42)	GGGAUUUCCAGCAUAGCUCUU
	AAACAUCCUGGUUGAACUGGAUGG
	CCACACUGUCUGCAGUCCGGACA
	CGCACCAG
mVenus anti-sgRNA(56)	GGGCACUUUUUCAAGUUGAU
	AACGACUUGCUAUGCUGUUUCCAGC
	AUAGCUCUUAAACAUCCUGGUUG
	AACUGGAUGGCCACACUGUCUG
	CAGUCCGGACACGCACC

List of Figures

Acknowledgement

First of all, I would like to thank Professor Friedrich Simmel for providing me with the opportunity to be part of an exciting research group. I am thankful for all the support over the years, the helpful discussions and the excellent ideas.

I would like to thank all the past and current members of the E14 chair. I learned a lot from everyone of you and I am particularly grateful for the support and advice, also beyond the domain of science. I am very thankful for the support at the beginning of my thesis from Andrea Mückl, Maximilian Weitz and Korbinian Kapsner. I would also like to thank Lukas Aufinger and Günther Pardatscher for helpful discussions and great collaborations throughout the whole time. I would also like to thank Aurore Dupin and Daniela Ziegler for proof-reading my thesis.

I would like to thank Suannse Kinzel and Helene Budjarek for all the help with administrative and lab related questions.

I would like to thank all the members of the Graduate School for Molecular Principles of Synthetic Biology. In particular Dr. Beate Hafner, who was always available and constructive in regard to organizational and financial questions. Then, I would like to thank Karsten Miermans for his helpful advice and enthusiastic discussions.

I would like to thank the members of the Dietz lab and especially, Florian Praetorius for helpful discussions and advice.

I would like to thank the visiting scientist Fabio Chizzolini and Guido Grossi for their great ideas and assistance.

Last but not least, I would like to thank my family, Aurore, and my friends and flatmates Jan, Fonsi, Andi, Pau and Ria for all their support through-out the years.