Virus-inspired dna replication coupled with gene expression in a minimal cell framework

van Nies, Pauline

DOI
10.4233/uuid:f572409d-366d-4042-ba02-3bed8209b4a0

Publication date
2017

Document Version
Final published version

Citation (APA)
van Nies, P. (2017). *Virus-inspired dna replication coupled with gene expression in a minimal cell framework*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:f572409d-366d-4042-ba02-3bed8209b4a0

VIRUS-INSPIRED DNA REPLICATION COUPLED WITH GENE EXPRESSION IN A MINIMAL CELL FRAMEWORK

VIRUS-INSPIRED DNA REPLICATION COUPLED WITH GENE EXPRESSION IN A MINIMAL CELL FRAMEWORK

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 3 maart 2017 om 12:30 uur

door

Pauline Yvonne Brigitte VAN NIES

Master of Science in Applied Physics,
Technische Universiteit Delft, Nederland
geboren te Zoetermeer, Nederland.

This dissertation has been approved by the

promotor: Prof.dr.ir. S. Tans
copromotor: Dr. C.J.A. Danelon

Composition of the doctoral committee:

Rector Magnificus,	voorzitter
Prof. dr. ir. S. Tans	Technische Universiteit Delft
Dr. C. J. A. Danelon	Technische Universiteit Delft
Independent members:	
Prof. dr. M. Dogterom	Technische Universiteit Delft
Prof. dr. A.C. Forster	Uppsala University
Dr. P. Holliger	Cambridge MRC Laboratory of Molecular Biology
Dr. M. Mencía	Universidad Autònoma de Madrid
Prof. dr. C. Dekker	Technische Universiteit Delft, reservelid

Keywords: minimal cell, in vitro gene expression, liposomes, RNA detection, DNA replication, phi29

Printed by: Gildeprint

Front & Back: Cover art by Emma Heijerman

Casimir PhD Series, Delft-Leiden 2017-02

ISBN 978-90-8593-287-1

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

This intense feeling of fascination and love for nature,
for life,
let us share it now that we are part of it,
and pass it on to those who will be,
when we are long gone.

"What a wonderful and amazing Scheme have we here
of the magnificent Vastness of the Universe!
So many Suns, so many Earths, and every one of them stocked
with so many Herbs, Trees and Animals,
and adorned with so many Seas and Mountains!
And how must our wonder and admiration be encreased
when we consider the prodigious distance and multitude of the Stars?"

Christiaan Huygens, Cosmotheoros, 1695

CONTENTS

1

INTRODUCTION

Life has evolved on Earth for more than three billion years, and today we are here to admire its extreme variety and beautiful complexity. The question: how did life originate? is a most captivating one for humanity as it leaves plenty of room for imagination and speculation. It can be subdivided in two parts: one concerns the specific environment and molecules involved, in other words, the transition from non-living to living matter, or so to say from chemistry to biology. What molecules form the building blocks of life, and what are the environmental conditions to sustain chemical reactions that can create more complex macromolecules and interactions? This is the realm of prebiotic chemistry, whereby an important concept is the notion of self-assembly. The second part relates closely to the question what is life? Many different answers can be found in literature and definitions can differ from person to person. A slightly different question has emerged from the origins of life field: how would a cell that contains the minimal and sufficient number of components to perform the basic functions of cellular life look like? In this question, cellular life is defined as the capability to display the properties of self-maintenance, reproduction and evolution [1]. The search will gradually reveal an answer to the reduction of complexity of modern cells and will thus lead to the construction of a minimal cell. However, the construction of such an elementary cell will not necessarily teach us more about the origins of life, but it will provide fundamental insights into the basic principles of life and their concerted functioning in a cell.

1.1. MINIMAL CELL

A minimal cell should display the same essential functions of a living cell, that are mentioned below [2]:
1. maintenance, replication and processing of information (evolution)
2. metabolism
3. energy production
4. maintenance of cell structure and growth
5. cell division

How can we find such a minimal cell today? This question prompts us to investigate our contemporary organisms that have very small genomes, as the concept of a minimal cell is closely related to the minimal-gene-set concept. The latter aims to derive 'the minimal set of genes that are necessary and sufficient to sustain a functioning cell under ideal conditions, that is, in the presence of unlimited amounts of all essential nutrients and in the absence of any adverse factors, including competition' [3]. Whereas the smallest free-living bacteria is the *Pelagibacter ubique* bacteria (genome size 1.31 M base pairs (bp)), parasitic bacteria are known to have even smaller genomes, as they lack enzymes for the synthesis of some nutrients that they uptake from their environment. Accepting that a minimal cell thrives in an environment where all essential nutrients are available, the amount of functions that a cell has to perform is reduced with respect to an organism that needs to metabolize basic resources into useful building blocks. For example, amino acids and nucleotide triphosphates (NTPs) are considered nutrients, so the corresponding synthesis pathways are not required. The smallest known free-living parasite is the extracellular Mycoplasma bacteria, that lacks enzymes for the synthesis of amino acids, purine and pyrimidine bases and for fatty acids. The genome size of *Mycoplasma genitalium* is 0.58 Mbp, roughly twice smaller than the *Pelagibacter ubique* and six times smaller than *Escherichia coli* genome. Besides parasites, symbionts that form a mutual beneficial relation with their host likewise possess reduced genomes. The smallest archeabacterium is the *Nanoarchaeum equitans*, an obligate symbiont of the sulfur reducing species *Ignicoccus*, with genome size of 0.49 Mbp and a diameter of only 400 nm [4]. Another example of a reduced symbiont genome is the *Carsonella ruddii* bacteria living in plant-louse that lost genes for cell envelope biogenesis and metabolism of nucleotides and lipids. It has a circular genome of 0.16 Mbp encoding for 182 genes of which 90% of open reading frames (ORF) overlap with at least one adjacent ORF [5]. These small organisms form the starting point for studies that aim to find a minimal genome by the so called top-down approach inspired by comparative genomics as discussed in the next section.

1.1.1. COMPARATIVE GENOMICS

The idea of comparative genomics is to compare the protein sequences of, for example, (two) different bacteria belonging to different ancient lineages, to find a conserved set of genes that is thus essential for cellular function. The comparison of the 468 protein genes from *M. genitalium* with the 1703 genes of the *Haemophilus influenzae* bacteria[1] resulted in the finding of 240 genes that are orthogolous, leading finally to a theoretical set of 256 genes for a minimal genome [6] (Figure 1.1A). Importantly, a large fraction of these genes is involved in ribosomal protein synthesis. The ribosome itself consists already of around 50 proteins that stabilize the structure of the (3) RNA molecules that perform the catalytic peptidyltransferase activity. One can also question how much a minimal cell would resemble the hypothetical last common universal ancestor (LUCA), where the three domains, bacteria, archea and eukarya all descend from. An attempt to find the functional content of LUCA is to compare the complete set of proteins from archea, bacteria and eukarya and find the protein families that occur in all of them [7]. The resulting full list of universal functions contained 246 unique biochemical functions

[1] These bacteria were the first ones to have their genome completely sequenced.

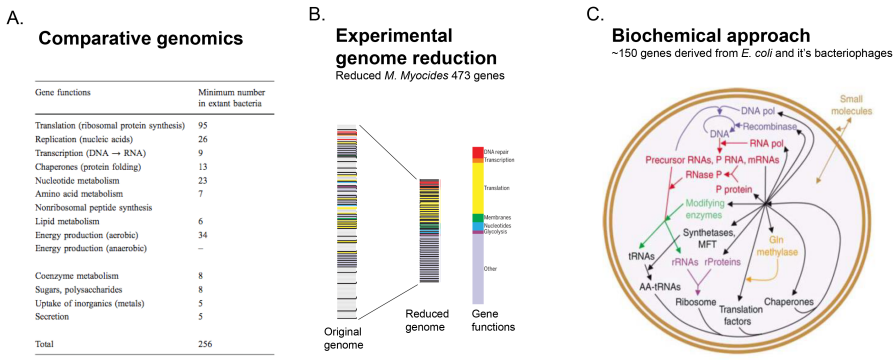


Figure 1.1: Different approaches to create a minimal cell are discussed in literature. A) The outcome of the comparative genomics analysis of *M. genitalium* and *H. influenzae* bacteria, adapted from [6]. B) Whole-genome design and synthesis led to viable cells with a reduced genome (1079-kbp to 531-kbp) of the *M. mycoides*, adapted from [9]. The genomes are displayed on the same scale, together with the gene functions of the reduced genome. C) The biochemical approach identifies gene products necessary for the reconstitution of reactions in a living cell. The processes derived from *E. coli* and its phages (transcription and DNA replication processes) are visually displayed, inside a compartment.

(from 301 proteins) and is mostly composed of metabolic enzymes, transporters and information processing elements. Later Gil et al. reduced the minimal set of protein-coding genes necessary to maintain a functional bacterial cell to 206 genes [2]. This size can hypothetically be further reduced to around 150 genes provided that nucleotides, amino acids and other low molecular weight compounds are abundant in the surrounding medium and able to permeate into the cell, as suggested by Luisi [8]. As proposed therein, one could try to further reduce the list of genes for a 'simple-ribosome[2]' cell or an extremely reduced cell of only 50 genes. The results of comparative genomics studies remain hypothetical, but are helpful towards identifying essential or non-essential genes in the process of reducing genomes of living organisms in the lab.

1.1.2. TOP-DOWN APPROACH TO A MINIMAL CELL

The top-down approach for the realization of a minimal cell aims at reducing the genome of extant 'simple' organisms by identifying and removing non-essential genes. Researchers at the J. Craig Venter Institute applied this strategy successfully using global transposon mutagenesis and identified 100 non-essential coding protein genes in the *M. genitalium*, resulting in a hypothetical viable organism with a genome of only 382 genes [10]. Interestingly, 28% of the proteins had an unknown function. Recently, the design and synthesis of a minimal bacterial genome was taken to the next level by researchers from the same institute. The 1079-kbp genome of *Mycoplasma mycoides* was reduced to 531 kbp (473 genes), a genome smaller than the natural *M. genitalium* [9] (Figure 1.1B). This genome was built synthetically, starting with the chemical synthesis of oligonucleotides followed by PCR-based assembly, error correction, PCR amplification, cloning and yeast assembly. The whole genome was then successfully transplanted into cells of *M. capri-*

[2]The simple ribosome only consists of the RNA molecules that perform the catalytic function

colum, a related *Mycoplasma*, by a method that for the first time allowed to change one species of bacteria into another [11]. The expression and preservation of genome information alone required already 229 genes. Their main criticism of the comparative genomic analysis performed in literature is that redundant genes for essential functions are often scored as nonessential in these studies. It would be interesting to find out if the proposed reduced genome of the smaller *M. genitialium* will also lead to a viable organism, now that the necessary techniques have been developed robustly. The acquired knowledge of this research also paves the way to further engineer bacteria to perform specific reactions such as producing medicines or fuels in the future.

1.1.3. BIOCHEMICAL, BOTTOM-UP, APPROACH

Besides the comparitive genomics and genome reduction approaches, there is a biochemical approach to construct a minimal cell that has its root in biochemistry. This approach aims to identify the gene products essential for the reconstitution of necessary reactions in a living cell. The main advantage of this bottom-up synthetic biology approach is that it enables the combination of genes from different organisms, prokaryotes and eukaryotes, as well as viral genes. It is thus not restricted to the evolution of one species, where redundant genes might still be essential because of interactions with other proteins and the phenomenon of epistasis. This research can benefit from comparative genomics studies investigating the protein functions of LUCA [7]. The first proposal for a minimal genome by the biochemical approach led to a genome consisting of 151 genes, of which 38 RNAs and 113 genes for proteins [12] (Figure 1.1C). In contrast with earlier proposals based on reducing the genome of bacteria, they have chosen to leave out membrane proteins and omitted enzymes for synthesizing lipids and glycolysis substrates. The formation of membrane compartments is proposed to exist due to the different nature of lipids alone and fission can be autocatalytic or triggered by external mechanical forces.

One could also consider a truly bottom-up approach, starting off with only simple molecules as in the origins of life scenario. In the realm of prebiotic chemistry, the chemical reactions that lead to the formation of the building blocks of life are investigated. This all started with the famous 1953-Urey-Miller experiment, wherein by running electric current through a gas (CH_4, NH_3, H_2) and water (H_2O), a mixture organic compounds, including amino acids, were formed. The prebiotic reaction sequences that led to the formation of ribonucleotides were only discovered less than a decade ago [13]. This was an important milestone, as it is widely accepted that there was once an RNA world, a period in the biosphere when both information storage and enzymatic activities were performed by RNA molecules. The polymerization of ribonucleotides and the non-enzymatic replication of RNA molecules [14–16], aided by peptides or inside fatty acid vesicles [17, 18], describes the next step towards more complex systems. Koonin and colleagues constructed a hypothesis that consists of plausible elementary steps towards the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation[3] and subfunctionalization [19]. The starting point for this scenario is a replicating ensemble of selfish cooperators consisting of RNA molecules with various ribozyme activities. The holy grail in protocell and origins of life research is thus

[3] Exaptation means that a trait is evolved as by-product for some other function.

to find a ribozyme polymerase that can copy other ribozymes and itself (a replicase), as this provides a starting point for evolutionairy processes. Of great interest is the recent work of Holliger's group that describes the evolution and engineering of an RNA polymerase ribozyme capable of transcribing another, hammerhead endonuclease, ribozyme [20]. The further engineered ribozyme was capable of catalyzing synthesis of an RNA sequence longer than itself [21], and could assemble from separate oligomers (not longer than 30 nt) under iterative temperature cycling conditions [22]. The latter finding is quite exciting as it bridges for the first time the conceptual gap from the prebiotic chemistry of short RNA oligomers to the enzymatic activity of complex replicating ribozymes.
However, it is probably impossible to mimic the selective pressure that was present at the early Earth giving rise to the first life. The bottom-up approach in the minimal cell community is therefore based upon DNA, RNA and protein molecules, aided by the tools of synthetic biology, rather than starting off with only simple building blocks.

1.2. LAB'S APPROACH TO CONSTRUCTING A MINIMAL CELL

The long-term goal of our lab is to construct a minimal cell using a bottom-up synthetic biology approach. Our aim is to achieve the functions of self-maintenance, reproduction and evolution starting with an extant minimal gene expression system that is compartmentalized inside a container defined by a lipid boundary. The gene expression is based upon the following core elements: DNA (heritable information), RNA (messenger) and protein (enzyme function or structure). Gene expression is defined in this thesis as the combination of the subsequent processes of transcription from DNA that generates RNA molecules, and translation of the RNA molecules by the ribosome and co-factors producing proteins. The flow of information from DNA to RNA to protein is generally known as the central dogma of molecular biology. The machinery of these systems is itself made of proteins and RNA, so if we want the cell to be an autonomous system, we also need to incorporate all the information for the production of these constituents in the cell. Our first -more 'modest'- goals are to expand the gene expression system with the following modules: DNA replication, lipid biosynthesis and division (Figure 1.2A). These processes must be coordinated in time inside cells, which is the motivation to also design and implement genetic networks. To find suitable components and mechanisms that can accomplish these tasks, we search for minimal systems of living cells such as bacteria, but also viral strategies. Once these are defined, DNA templates are designed for gene expression and we study the activity of the modules separately. Finally, all successful models can be combined that will hopefully lead to a minimal cell.
We should never forget that constructing a minimal cell is an extremely daunting task, and the realization should not be seen as the only successful achievement of this line of research. An important result of the reduction of the *Mycoplasma mycoides* genome was that 31% of the protein coding genes had an unknown function [9]. This provides thus a great inspiration and starting point to unravel the functions of these proteins and gain a better understanding of their necessity in living systems. Similarly, one can learn from the bottom-up synthetic biology approach to the minimal cell can learn us for example more about which proteins or conditions improve translation processes, considering we start out from a very minimal gene expression system. It provides a platform of

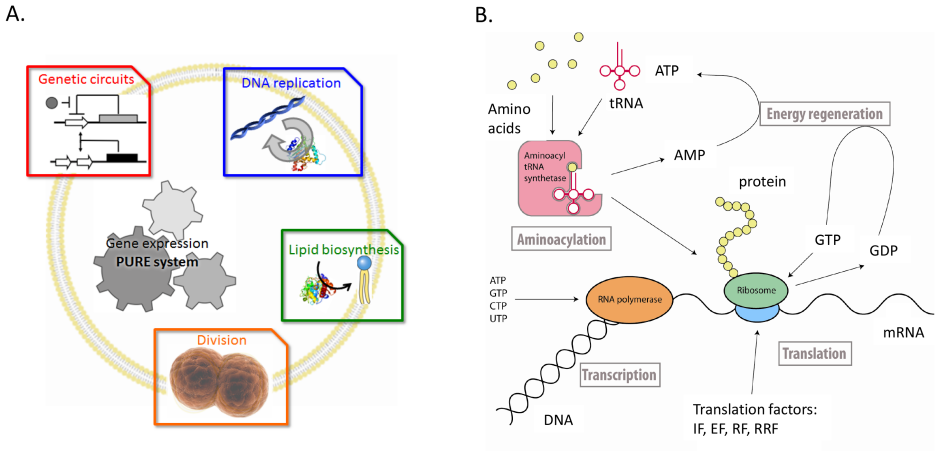


Figure 1.2: Our approach towards constructing a minimal cell is based on expanding a gene expression system with four modules: DNA replication, lipid biosynthesis, liposome division and genetic networks (A). The gene expression system that we use as a platform is the PURE system (B) with its four main processes, adapted from [23].

intermediate complexity in between single-molecule studies and *in vivo* studies, an environment in which protein functions can be studied alongside transcription-translation reactions, or inside cell-sized compartments. Can the reactions that take place in living cells be reproduced by enzymes expressed in the minimal gene expression environment? If yes, we learn about the minimal requirements to achieve a particular function. If not, this opens the door for new questions and their answers can provide insights for a better fundamental understanding of the essential processes of life.

1.2.1. PURE SYSTEM PLATFORM

Our platform for gene expression is a cell-free system that is reconstituted from purified components, termed 'protein synthesis using recombinant elements' (PURE) system [23, 24]. In general, a cell-free system consists of all cell components except the membrane, other large macromolecules and genomic DNA. The problem with conventional cell extracts is that only a fraction of all the components participates in the translation reaction and that some constituents can even inhibit or degrade necessary molecules for the desired processes. The PURE system consists only of these components that are absolutely necessary for transcription and translation to occur. Each component in the PURE system is purified separately from *E. coli* and reconstituted in a solution at known concentrations. The exception is the RNA polymerase (RNAP), which is a single subunit RNAP obtained from the T7 bacteriophage. The advantage of using this system is the complete knowledge of all the components, which assures the controllability of gene expression. Besides transcription and translation machinery, it contains enzymes for energy regeneration and aminoacylation, as schematically depicted in Figure 1.2B . There are two commercially available kits of the PURE system on the market: PURE*frex*

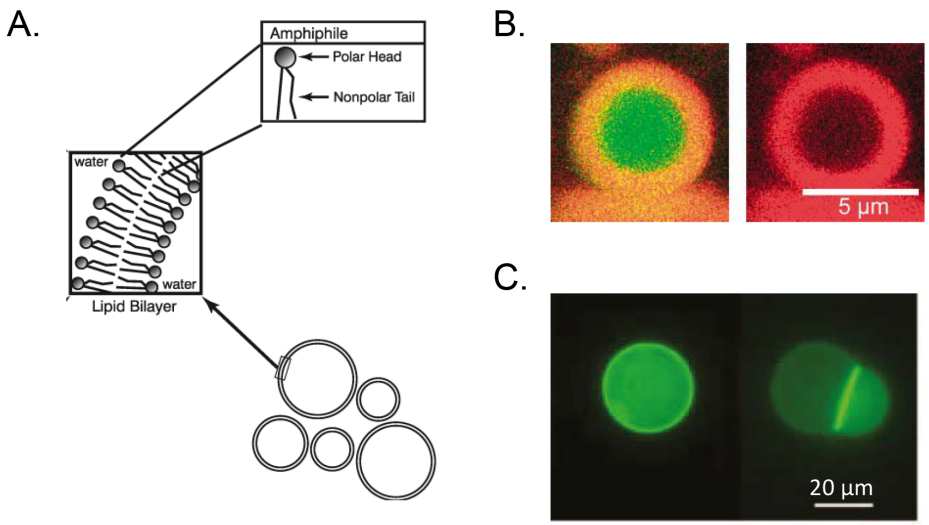


Figure 1.3: Liposomes are utilized as compartments for gene expression reactions. A) A liposome is a spherical lipid bilayer vesicle, formed by amphiphilic lipid molecules consisting of a hydropilic, polar head-group and a hydrophobic, nonpolar tail (image source: Avanti Lipids). B) Cross-sectional image of a vesicle containing a cell-free transcription-translation solution expressing a GFP variant (green color), obtained by confocal laser scanning microscopy. A fluorescent lipophilic tracer was used to stain the membrane (red color). Figure adapted from [25]. C) Cross-section of a single (left) and doublet (right) liposome expressing α -hemolysin (a membrane protein-pore) fused to GFP, adapted from [26].

(GeneFrontier) and PURExpress (New England Biolabs). Both systems originate from the original PURE system developed by Prof. Ueda's group from Japan [24], although the full composition of the latter is not disclosed by the manufacturer. The PURE*frex* system comes in separate solutions for the feeding mix (containing the amino acids, nucleotides and tRNAs), the enzyme mix and the ribosomes. None of these components have a His-tag, thus it is possible to purify a product containing a His-tag after gene expression. Both suppliers also provide modified PURE systems that for example lack initiation factors or amino acids that can be added separately to the reaction solution.

The encapsulation of gene expression systems in a phospholipid vesicle to build cell-like bioreactors (Figure 1.3) has been accomplished by several groups more than a decade ago [25–27]. This has triggered research into gene expression inside lipid vesicles that serves as a core architecture in constructing the minimal cell, which will be discussed further throughout this thesis [28–30].

1.3. THESIS OUTLINE

This thesis contributes to the minimal cell project by first gaining a better quantitative understanding of the transcription and translation processes in the minimal gene expression system and secondly by defining and investigating the DNA replication module. Finally, we give here a brief outline of this thesis:

• **Chapter 2** provides a quantitative understanding of the transcription and translation

reactions in the PURE system, that is obtained by using two orthogonal fluorescence labeling tools that report the amounts of mRNA and protein simultaneously and in real-time. The Spinach RNA aptamer and its fluorogenic probe were used for mRNA detection, and a fluorescent protein was expressed that was detected as well. Applying this dual-reporter assay to the analysis of transcript and protein production inside lipid vesicles revealed that their levels are uncorrelated, most probably a consequence of the low-copy number of some components in liposome-confined reactions.

• **Chapter 3** reports on an improved Spinach reporter with enhanced fluorescence signal, which allows us to monitor transcription with unprecedented sensitivity over longer time periods. We quantified gene expression kinetics in individual vesicles and found that there are large differences between individual liposomes, corroborating our previous results. An interesting finding is that a subset of expressing liposomes produced larger amounts of mRNA and/or proteins than measured in bulk experiments. This research provides a valuable starting point to study factors that can enhance gene expression in bulk and in liposomes.

• **Chapter 4** discusses several DNA replication candidates for the minimal cell. Bacterial DNA replication as well as viral DNA replication systems and *in vitro* DNA amplification systems are concisely reviewed. The aim of this chapter is to inform and provide ideas for anyone interested in the topic of DNA replication in a minimal cell.

• **Chapter 5** describes the investigation of the viral strategy to use RNA as an intermediate information carrier. Herein, we propose how a nucleic acid amplification cycle can be coupled with transcription and translation processes. We report on the properties of an intermediate species of this cycle, the DNA:RNA hybrid template, that can be transcribed by the T7 RNAP, although with much lower efficiency than the double stranded DNA.

• **Chapter 6** presents the viral phi29 DNA replication system as a promising candidate for the minimal cell. The replication activity of the four synthesized replication proteins (DNA polymerase, terminal protein, single-stranded binding protein and double-stranded binding protein) in the PURE system is confirmed. We show that the DNA that encodes for these proteins is replicated, completing one round of the central dogma in biology in a minimal gene expression system.

• **Chapter 7** discusses the future outlook; from gene expression regulation, improving conditions for the DNA replication reactions and studying collisions of enzymes on DNA templates, to amplifying DNA molecules inside liposomes.

REFERENCES

[1] P. L. Luisi, T. Oberholzer, and A. Lazcano, *The notion of a DNA minimal cell: a general discourse and some guidelines for an experimental approach.* Helv Chim Acta **85**, 1759 (2002).

[2] R. Gil, F. J. Silva, J. Peretó, and A. Moya, *Determination of the core of a minimal bacterial gene set.* Microbiology and molecular biology reviews : MMBR **68**, 518 (2004).

[3] E. V. Koonin, *How Many Genes Can Make a Cell: The Minimal-Gene-Set Concept 1,* Annual review of genomics and human genetics **1**, 99 (2000).

[4] H. Huber, M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer, and K. O. Stetter, *A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont.* Nature **417**, 63 (2002).

[5] A. Nakabachi, A. Yamashita, H. Toh, H. Ishikawa, H. E. Dunbar, N. A. Moran, and M. Hattori, *The 160-kilobase genome of the bacterial endosymbiont Carsonella.* Science **314**, 267 (2006).

[6] A. R. Mushegian and E. V. Koonin, *A minimal gene set for cellular life derived by comparison of complete bacterial genomes.* Proceedings of the National Academy of Sciences **93**, 10268 (1996).

[7] N. Kyrpides, R. Overbeek, and C. Ouzounis, *Universal protein families and the functional content of the last universal common ancestor,* Journal of Molecular Evolution **49**, 413 (1999).

[8] P. L. Luisi, *Approaches to the minimal cell,* in *The Emergence of Life* (Cambridge University Press, 2006) pp. 247–254.

[9] C. A. Hutchison, R.-Y. R.-Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z.-Q. Z.-Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y. Y. Suzuki, B. Tsvetanova, KimS.Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, J. C. Venter, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, and J. C. Venter, *Design and synthesis of a minimal bacterial genome,* Science **351**, aad6253 (2016).

[10] J. I. Glass, N. Assad-Garcia, N. Alperovich, S. Yooseph, M. R. Lewis, M. Maruf, C. a. Hutchison, H. O. Smith, and J. C. Venter, *Essential genes of a minimal bacterium.* Proceedings of the National Academy of Sciences of the United States of America **103**, 425 (2006), arXiv:/www.nature.com/nature/journal/v473/n7346/abs/10.1038-nature10011-unlocked.html#supplementary-information [http:] .

[11] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, H. O. Smith, and J. C. Venter, *Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome,* Science **329**, 52 (2010), arXiv:z0024 .

[12] A. C. Forster and G. M. Church, *Towards synthesis of a minimal cell.* Molecular systems biology **2**, 45 (2006).

[13] M. W. Powner, B. Gerland, and J. D. Sutherland, *Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions.* Nature **459**, 239 (2009).

[14] J. P. Ferris and G. Ertem, *Montmorillonite catalysis of RNA oligomer formation in aqueous solution. A model for the prebiotic formation of RNA.* Journal of the American Chemical Society **115**, 12270 (1993).

[15] S. Rajamani, A. Vlassov, S. Benner, A. Coombs, F. Olasagasti, and D. Deamer, *Lipid-assisted synthesis of RNA-like polymers from mononucleotides.* Origins of life and evolution of the biosphere the journal of the International Society for the Study of the Origin of Life **38**, 57 (2008).

[16] J. W. Szostak, *The eightfold path to non-enzymatic RNA replication,* Journal of Systems Chemistry **3**, 2 (2012).

[17] T. Z. Jia, A. C. Fahrenbach, N. P. Kamat, K. Adamala, and J. W. Szostak, *Oligoarginine Peptides Slow Strand Annealing and Assist Nonenzymatic RNA Replication,* Nature Chemistry **8**, In Press (2016).

[18] S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. a. Treco, and J. W. Szostak, *Template-directed synthesis of a genetic polymer in a model protocell.* Nature **454**, 122 (2008).

[19] Y. I. Wolf and E. V. Koonin, *On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization.* Biology direct **2**, 14 (2007).

[20] A. Wochner, J. Attwater, A. Coulson, and P. Holliger, *Ribozyme-catalyzed transcription of an active ribozyme.* Science (New York, N.Y.) **332**, 209 (2011).

[21] J. Attwater, A. Wochner, and P. Holliger, *In-ice evolution of RNA polymerase ribozyme activity.* Nature chemistry **5**, 1011 (2013).

[22] H. Mutschler, A. Wochner, and P. Holliger, *Freeze–thaw cycles as drivers of complex ribozyme assembly,* Nature Chemistry **7**, 502 (2015).

[23] Y. Shimizu, T. Kanamori, and T. Ueda, *Protein synthesis by pure translation systems.* Methods (San Diego, Calif.) **36**, 299 (2005).

[24] Y. Shimizu, a. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda, *Cell-free translation reconstituted with purified components.* Nature biotechnology **19**, 751 (2001).

[25] S.-i. M. Nomura, K. Tsumoto, T. Hamada, K. Akiyoshi, Y. Nakatani, and K. Yoshikawa, *Gene expression within cell-sized lipid vesicles.* Chembiochem **4**, 1172 (2003).

[26] V. Noireaux and A. Libchaber, *A vesicle bioreactor as a step toward an artificial cell assembly.* Proceedings of the National Academy of Sciences of the United States of America **101**, 17669 (2004).

[27] K. Ishikawa, K. Sato, Y. Shima, I. Urabe, and T. Yomo, *Expression of a cascading genetic network within liposomes.* FEBS letters **576**, 387 (2004).

[28] Z. Nourian, W. Roelofsen, and C. Danelon, *Triggered Gene Expression in Fed-Vesicle Microreactors with a Multifunctional Membrane,* Angew. Chem. Int. Ed **51**, 3114 (2012).

[29] T. P. de Souza, A. Fahr, P. L. Luisi, and P. Stano, *Spontaneous Encapsulation and Concentration of Biological Macromolecules in Liposomes: An Intriguing Phenomenon and Its Relevance in Origins of Life.* Journal of molecular evolution , 179 (2014).

[30] P. Stano, E. D'Aguanno, J. Bolz, A. Fahr, and P. L. Luisi, *A Remarkable Self-Organization Process as the Origin of Primitive Functional Cells.* Angewandte Chemie (International ed. in English) , 1 (2013).

2

TRACKING OF THE PROGRESSION OF mRNA AND PROTEIN SYNTHESIS IN BULK AND INSIDE LIPID VESICLES

Compartmentalization of a cell-free gene expression system inside a self-assembled lipid vesicle is envisioned as the simplest chassis for the construction of a minimal cell. Although crucial for its realization, quantitative understanding of the dynamics of gene expression in bulk and liposome-confined reactions is scarce. Here, we used two orthogonal fluorescence labeling tools capable of reporting the amounts of mRNA and protein produced in a reconstituted biosynthesis system simultaneously and in real-time. The Spinach RNA aptamer and its fluorogenic probe were used for mRNA detection. Applying this dual-reporter assay to the analysis of transcript and protein production inside lipid vesicles revealed that their levels are uncorrelated, most probably a consequence of the low-copy number of some components in liposome-confined reactions. We believe that the unveiled stochastic nature of gene expression should be apprehended as a design principle for the assembly of a minimal cell.

This chapter has been published as *Unbiased Tracking of the Progression of mRNA and Protein Synthesis in Bulk and inside Lipid Vesicles*, Pauline van Nies, Zohreh Nourian, Maurits Kok, Roeland van Wijk, Jonne Moeskops, Ilja Westerlaken, Jos M. Poolman, Rienk Eelkema, Jan H. van Esch, Yutetsu Kuruma, Takuya Ueda, and Christophe Danelon, *ChemBioChem* **2013**, *14*, 1963-1966.

2.1. INTRODUCTION

The assembly of a genetically controlled artificial cell model is arguably one of the most challenging goals within the synthetic biology field. In recent developments, a minimal chassis based on cell-free protein biosynthesis compartmentalized inside a self-assembled lipid vesicle could be built up[1–6]. Yet, the dynamics of DNA transcription and mRNA translation in liposomes is not understood, which impedes the realization of a programmable elementary cell operating the expression of multiple genes. For instance, the relationship between the amounts of synthesized mRNA and protein in vesicle-confined reactions is not established.

By eliminating cellular support for gene expression, in vitro transcription and translation systems enable higher levels of control and manipulation for tailored protein production [7–10]. Of particular relevance for the assembly of an elementary cell is the PURE system, a minimal gene expression system reconstituted solely from purified constituents for transcription, translation, aminoacylation and energy regeneration [11, 12]. The PURE system is made of well-defined elements with known concentrations and intended functions, which provides a unique cell-free platform for reliable modeling and mathematical description of the reaction dynamics.

However, many quantitative aspects underlying the dynamics of protein synthesis from a DNA template remain poorly understood in cell-free bulk reactions and to a larger extent in liposome-confined reactions. More insights would be gained if the progression of the synthesis of mRNA, the key intermediate reactional product, and the output protein kinetics could be monitored simultaneously. Indeed, only if both mRNA and protein production is observed in real-time can the precise individual steps of gene expression be revealed [13]. While the newly synthesized proteins can be readily visualized by using autofluorescent proteins, time series measurements of mRNA copy numbers ask for indirect and orthogonal labelling tools.

A number of fluorescent probes have been developed to detect mRNA in live cells and in in vitro systems: GFP (green fluorescent protein)-tagged proteins that recognize specific sites on the target RNA [14], hybridization probes including molecular beacon [15, 16], and side-by-side probes [17, 18], and fluorogenic dyes that bind to an artificial RNA motif, so-called aptamer [19–25]. Of particular relevance is a recently described RNA aptamer that can recognize and activate a derivative of the GFP chromophore 4-hydroxybenzylidene imidazolinone, the 3,5-difluoro-4-hydroxybenzylidene (DFHBI) [25]. Thanks to remarkable brightness and photostability of the corresponding RNA-dye complex, named Spinach, endogenous RNA that was genetically tagged with the Spinach aptamer sequence could be visualized in live cells[25–27]. Molecular beacon and side-by-side oligonucleotide probes have recently been employed to measure the kinetics of synthesized transcripts and proteins in the commercially available PURE system (PURExpress) and some important parameters for transcription-only [18] and coupled transcription-translation reactions were determined [16]. However, the delay time introduced by probe hybridization was not quantified, which may lead to erroneous estimations of the effective transcription rate constant. In particular molecular beacons are recognized to have

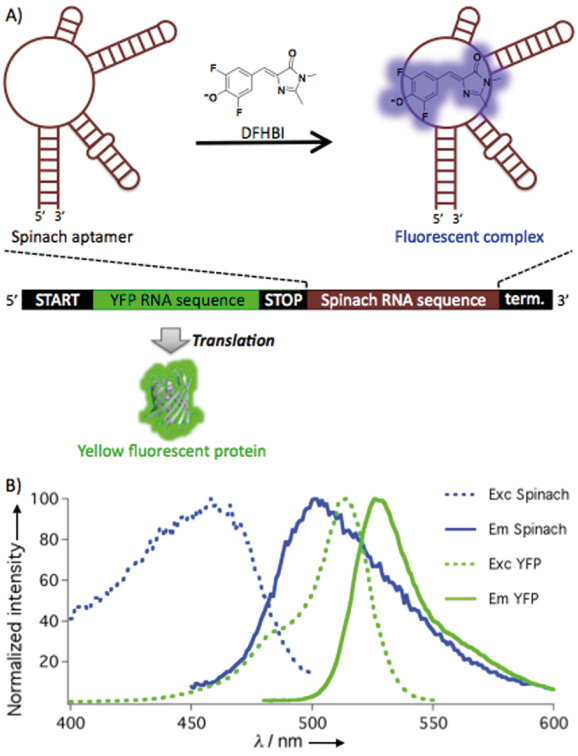


Figure 2.1: Spinach-based fluorescence detection of mRNA in a cell-free transcription-translation system. A) Schematic of the YFP-Spinach construct and aptamer-dye binding. B) Excitation and emission spectra of Spinach complex and YFP obtained 4 h post-expression in bulk at 37 °C. Spinach spectra were measured without ribosome in the presence of 20 μM DFHBI, while YFP signal was collected in the absence of DFHBI.

slow binding rates [28]. Besides, the absence of adverse effects on the efficiency of the transcription and translation reactions has yet to be demonstrated. Lastly, these RNA probes are not membrane permeable limiting their utilization as a visualization method in lipid vesicle-based systems.

Herein, we report on the use of Spinach as a superior mRNA labeling tool to dynamically monitor the transcription and translation processes in the PURE*frex* (GeneFrontier) for bulk reactions and liposome-based artificial cell models. The DNA template is under the control of a T7 promoter and encodes for the Spinach sequence fused to the 3'-untranslated region, downstream of the coding sequence of the yellow fluorescent protein (YFP) (Figure 2.1A). Hence, the levels of transcripts and output proteins can be monitored through the fluorescence intensity of Spinach and YFP, respectively (Figure 2.1B).

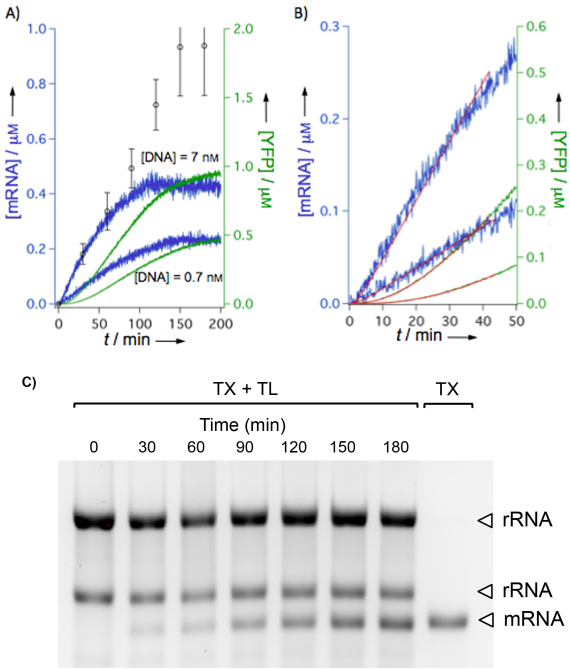


Figure 2.2: Simultaneous detection of mRNA and protein synthesis kinetics in bulk reactions. A, B) Simultaneous recording of the apparent kinetics of mRNA and protein production using real-time Spinach-based fluorescence with 0.7 nM (bottom blue and green traces) or 7 nM (top blue and green traces) of the YFP-Spinach construct. A) Appended symbols correspond to mRNA concentrations measured at different time points by RNA gel analysis as shown in C). Error bars indicate SEM, n=3. B) Expanded view of the early phase of the kinetics shown in A), which were fitted to the reaction model described in Section 2.2.2 (red lines). C) The actual kinetics of synthesized mRNA was measured by collecting a small fraction of the gene expressing solution at different time points. The RNA gel shows stable amounts of 1.5 and 2.9-kb ribosomal RNA (rRNA) and increasing levels of mRNA produced in the transcription (TX)-translation (TL) system PURE*frex*. The right-most lane corresponds to purified mRNA synthesized in a TX-only system, which was then used for quantification by absorbance measurement and served as a reference for calibrating the concentration of mRNA produced in the PURE*frex*.

2.2. RESULTS AND DISCUSSION

2.2.1. DETERMINISTIC GENE EXPRESSION IN BULK REACTIONS

The profiles of synthesized mRNA and protein were simultaneously detected in bulk reactions in the presence of DFHBI (Figure 2.2A,B) at 9 s time resolution. We evaluated a number of criteria that the Spinach-based detection method should satisfy to reliably report on the dynamics of synthesized mRNA molecules.

First, we verified that DFHBI does not interfere with the transcription and translation processes (Figure 2.3A,B). The amount of mRNA produced after 100 min, the kinetics and final concentration of synthesized YFP, are identical in the absence or presence of 20 μM DFHBI, demonstrating that DFHBI has no adverse effects on the transcription and translation reactions. In contrast to the non-interfering DFHBI molecule, the side-by-

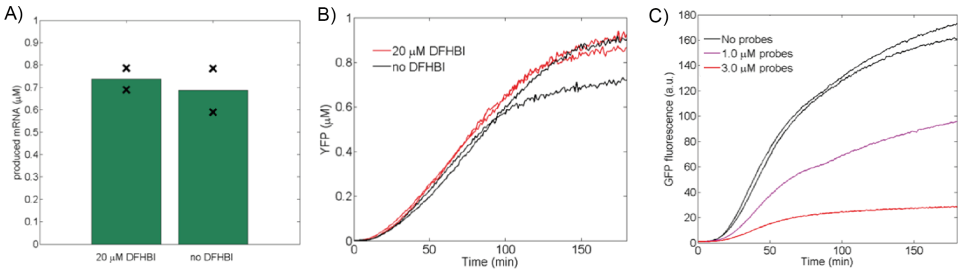


Figure 2.3: DFHBI does not interfere with the gene expression reactions in contrast to side-by-side probes that caused detrimental effects. a) Concentration of mRNA in the absence or presence of 20 μM DFHBI. The mRNA was produced in the PURE*frex* for 100 min using 7.0 nM YFP-Spinach and its concentration was quantified as described above. The bars represent the average values of two independent experiments (symbols). b) Kinetics of YFP synthesis in the absence (black traces) or presence (red traces) of 20 μM DFHBI. DNA template was 7.0 nM YFP-Spinach. Two independent experiments are displayed for each condition. c) Apparent kinetics of deGFP produced in the PURExpress in the presence of different side-by-side probe concentrations. DNA template was 9.0 nM deGFP.

side probes showed a detrimental effect on protein expression in the PURExpress (Figure 2.3C). As we believe that this deleterious effect on translation and/or transcription is independent of the probe-sequence (as they bind in the untranslated region), we did not further investigate the side-by-side probes and favored the Spinach method for real-time detection of synthesized mRNAs.

Second, we quantified the time delay between mRNA synthesis and Spinach fluorescence detection (Figure 2.4A). During the course of gene expression, DNaseI was added to degrade the DNA templates. The observed residual increase of Spinach fluorescence is due to combined effects of Spinach RNA folding, DFHBI binding and activation (Figure 2.4b). The associated Spinach maturation rate constant is 0.38 min^{-1}, corresponding to a delay time of 2.6 min. The delay time in protein dynamics due to YFP maturation was previously determined in the PURExpress and a maturation rate constant of 0.046 min^{-1} was found [29].

Next, we demonstrated that the apparent kinetics of mRNA production monitored by fluorescence is similar to that of the actual time-dependent increase of transcripts (Figure 2.2A,C). Only after 90 min a deviation is observed with higher transcript amounts produced in test-tube versus cuvette (see Supporting Information for a discussion). Besides, we found that mRNA is not significantly degraded (Figure 2.14).

Fluorescence intensity values were converted into concentrations of mRNAs (Figures 2.12,2.13) and YFPs (Figures 2.10,2.11) as described in detail in the Supporting Information. Starting with 7 nM DNA, the mRNA and YFP accumulated within 1 h to concentrations of 0.35 μM and 0.55 μM, respectively, which corresponds to a stoichiometry DNA:mRNA:protein = 1:50:79 (Figure 2.4C, Figure 2.6). A quantitative model of gene expression was developed and used to extract the relevant reaction parameters of the kinetics of synthesized transcripts and proteins.

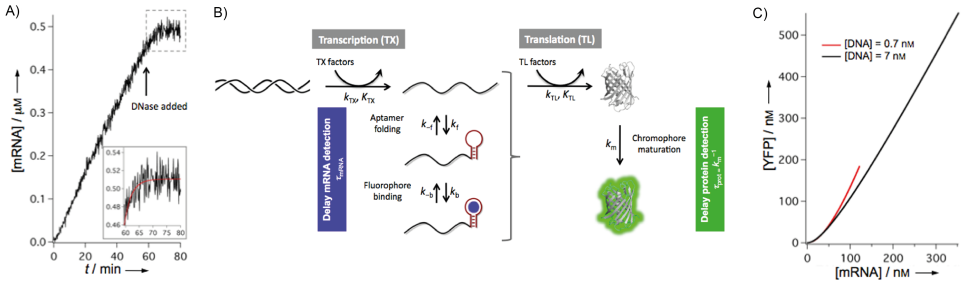


Figure 2.4: A Coarse-grained kinetic model of the transcription and translation reactions in the PURE system depends on several parameters. A) The delay time between mRNA production and its detection was quantified by measuring the residual increase of Spinach fluorescence upon YFP-spinach DNA degradation in order to ascertain a stable amount of mRNA. The red line is a monoexponential fit to the fluorescence signal shown in the dashed box, which gives an apparent maturation rate constant of 0.38 min^{-1} ± 0.15 min^{-1} (n=4). [DNA] = 14 nM, [DFHBI] = 20 μM. B) The main steps introducing a delay time between product synthesis and detection are indicated. The Spinach RNA aptamer is shown as a brown hairpin and the complexed DFHBI fluorophore as a blue circle. C) Actual concentrations of YFP versus mRNA as calculated by fitting the first 60 min of the traces in 2.2A) and B) to the reaction kinetics model presented.

2.2.2. MODELING GENE EXPRESSION

Our model relies on Michaelis-Menten dynamics of the transcription and translation reactions. This formalism has recently been used in three studies to quantify the progression of mRNA or/and protein production with *E. coli* cell extracts [30] or with the PURExpress[16, 18]. Thus, it provides a good analysis framework for the reaction kinetics measured in the PURE*frex* using Spinach-based fluorescence detection of mRNA (Figure 2.4). In spite of the recent development of a model that encompasses the late phase of gene expression where synthesis expires [16], we only focused our quantitative analysis to the early phase of the production since it suffices to extract the main reaction constants with only four unknown parameters. This means that the cessation of expression due to resource consumption or enzyme inactivation was not captured in our model. Besides, Spinach-based labeling reliably reports on the dynamics of mRNA synthesis during the first 90 min of production (Figure 2.2a). The delay times for mRNA and protein maturation were implemented in the model for unbiased characterization of the catalytic rate constants of the transcription and translation reactions. The following set of differential equations was used:

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{mRNA}] = k_{TX} \cdot \frac{[\mathrm{RNAP}] \cdot [\mathrm{DNA}]}{K_{TX} + [\mathrm{DNA}]} - k_{\mathrm{mat}}^{\mathrm{mRNA}} \cdot [\mathrm{mRNA}] \tag{2.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathrm{mRNA}^*\right] = k_{\mathrm{mat}}^{\mathrm{mRNA}} \cdot [\mathrm{mRNA}] \tag{2.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{YFP}] = k_{\mathrm{TL}} \cdot \frac{[\mathrm{E_{TL}}] \cdot ([\mathrm{mRNA}] + [\mathrm{mRNA}^*])}{\mathrm{K_{TL}} + ([\mathrm{mRNA}] + [\mathrm{mRNA}^*])} - k_{\mathrm{mat}}^{\mathrm{YFP}} \cdot [\mathrm{YFP}] \tag{2.3}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathrm{YFP}^*\right] = k_{\mathrm{mat}}^{\mathrm{YFP}} \cdot [\mathrm{YFP}] \tag{2.4}$$

where [RNAP] is the total concentration of the T7 RNA polymerase, $[\mathrm{E_{TL}}]$ is the total concentration of the translation machinery, k_{TX} (k_{TL}) and K_{TX} (K_{TL}) are the catalytic rate

and Michaelis constants for transcription (translation), k_{mat}^{mRNA} and k_{mat}^{YFP} are the maturation rate constants for mRNA and YFP, respectively. The star superscript indicates the detected mature products.

Our model is built upon the following considerations:
• Transcription and translation reactions are uncoupled.
• No co-transcriptional folding of Spinach as the aptamer sequence is located in the 3'-UTR.
• Only the full-length transcript serves as a template for YFP translation.
• The full-length mRNA can be translated regardless of the Spinach folding and DFHBI occupancy states.
• All synthesized proteins are active.
• mRNA (Figure 2.14) and proteins are not degraded.
• Decrease of transcription or translation activity through depletion of resources or inactivation is negligible within the first hour.
• Note that in the classical reaction scheme, where the complex is formed with a 1:1 stoichiometry of enzyme and substrate, Eq. 2.1 is only valid if [RNAP] + K_{TX} << [DNA]. However, this approximation is erroneous when using the PURE*frex* as [RNAP] = 30 nM and the DNA concentrations range from subnanomolar to 10 nM. Nonetheless, Eq. 2.1 is suitable if one considers that the T7 promoter is always accessible, i.e. the mean time between two consecutive binding events is higher than the mean residence time of the T7 RNA polymerase onto the promoter sequence. In other words, we assume that $[DNA]_{total}$ = $[DNA]_{free}$. This is supported by a recent fluorescence-based analysis of the binding events and transcription elongation of single T7 RNA polymerases with DNA molecules [31]. Note that this reaction scheme is compatible with the fact that multiple T7 RNA polymerases can bind a single DNA molecule in a noncooperative manner.
• In analogy to the above assumption for transcription, we consider that a single mRNA molecule can be simultaneously processed by multiple ribosomes and that [mRNA]total = $[mRNA]_{free}$ at any time. Here, we used $[E_{TL}]$ = 0.1 μM, which was motivated by the fact that the amount of translation complex is presumably limited by the lowest concentration of one of the necessary proteins, which is the concentration of the initiation factor IF-1 in the PURE*frex* [11].
• Contrary to previous studies, we explicitly introduced the delay time for mRNA maturation that results from Spinach folding and subsequent DFHBI binding (Eq. 2.2). This time has been experimentally measured (Figure 2.4a, main text) and was used as a fixed parameter in our model (Table 2.1).
• The delay time for the detection of YFP was also implemented (Eq. 2.4). We assumed that chromophore maturation was the rate-limiting step for YFP detection and a value of 0.046 min^{-1}, previously measured in the PURExpress [29], was used (Table 2.1).

We monitored apparent mRNA and YFP synthesis kinetics at 0.7 nM and 7 nM DNA concentrations. For every experiment, i.e. every pair of mRNA and YFP curves, the mRNA synthesis kinetics was first fitted separately using Eq. 2.1 and 2.2. The extracted values of k_{TX} and K_{TX} were then used as fixed parameters to fit YFP synthesis kinetics using Eq. 2.3 and 2.4, upon which values for kTL and KTL were determined. The results are summarized in Table 2.1.

Fixed Parameters		Free parameters SD	
DNA	0.7 or 7.0 nM	k_{TX}	0.29 ± 0.01 min-1, 5.6 NTP s^{-1}
RNAP	30 nM	K_{TX}	2.4 ± 0.7 nM
E_{TL}	100 nM	k_{TL}	0.086 ± 0.026 min-2, 0.44 aa s^{-1}
k_{mat}^{mRNA}	0.38 min^{-1}	K_{TL}	46 ± 33 nM
k_{mat}^{YFP}	0.046 min^{-1}		

Table 2.1: Analysis of mRNA and protein synthesis dynamics by model fitting. The values of the free parameters correspond to averaged values obtained with 0.7 and 7 nM DNA for a total of five independent experiments. Abbreviations: NTP, nucleotide triphosphate; aa, amino acids; SD, standard deviation.

The transcription rate constant is 0.29 min^{-1}, equivalent to 5.6 NTP s^{-1}. The translation rate constant is 0.086 min^{-1}, equivalent to 0.44 amino acid s^{-1}. The Michaelis constants for transcription and translation are 2.4 nM and 46 nM, respectively.

Note that the k_{TL} value based on the YFP concentration obtained by FCS is 0.032 ± 0.01 min^{-1}, corresponding to 0.16 aa s^{-1}, with all other reaction parameters remaining the same. All values fall into a reasonable range upon comparison with those reported in the PURExpress [16, 18]. For direct comparison transcription and translation rate constants were converted into rates by multiplying by [RNAP] and [E_{TL}], respectively, giving a maximal transcription rate of 8.7 ± 0.3 nM/min and an estimated maximal translation rate of 8.6 ± 2.6 nM/min. We also verified using a broad range of DNA concentrations that, in the linear regime of mRNA synthesis (i.e. subsequent to the few minutes of time lag), the profile of measured transcription rates was consistent with the Michaelis-Menten kinetics (Figure 2.5). The data were fitted to the first term in the right-hand side of Eq. 2.1, which is justified since the production rates of total and mature mRNAs are the same once the time lag is over (Figure 2.6a, c). Values of K_{TX} = 1.8 ± 0.2 nM and k_{TX} = 0.27 ± 0.01 min^{-1} (corresponding to a maximal transcription rate of 8.0 ± 0.3 nM/min) were obtained. These values are in good agreement with those determined by fitting the entire initial phase using Eq.2.1 and 2.2.

Although the amount of transcripts increases in the first 1.5 h (Figure 2.2a), the YFP production rate does not increase but reaches a steady slope after a few minutes starting with 7.0 nM DNA (Figure 2.6b) and after around 40 min starting with 0.7 nM DNA (Figure 2.6d). At these time points, the mRNA concentration reached may already saturate the translation machinery. The new transcripts are not readily translated into proteins, which explains the differences in the DNA:mRNA:protein stoichiometry between the two DNA concentrations, 1:189:287 at 0.7 nM DNA and 1:50:79 at 7 nM DNA after 1 h synthesis. In Figure 2.4 we note that the protein to mRNA ratio is higher for 0.7 nM DNA than for 7 nM DNA, which points towards a more efficient translation at low DNA concentrations.

2.2.3. STOCHASTIC GENE EXPRESSION IN LIPOSOME-CONFINED REACTIONS

Next we sought to exploit the Spinach technology to visualize mRNA in protein synthesizing liposomes. The PURE system enzymes and cofactors along with the YFP-Spinach DNA template were compartmentalized inside surface-tethered vesicles as previously described [6]. Gene expression was triggered by external supply of the feeding solution,

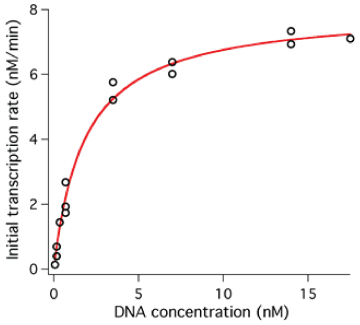


Figure 2.5: Michaelis-Menten kinetics of transcription. Transcription rates were calculated by linearly fitting the 3-20 min period of the apparent mRNA synthesis kinetics. The red curve is a fit to Eq. 2.1, where the term for mRNA maturation is omitted.

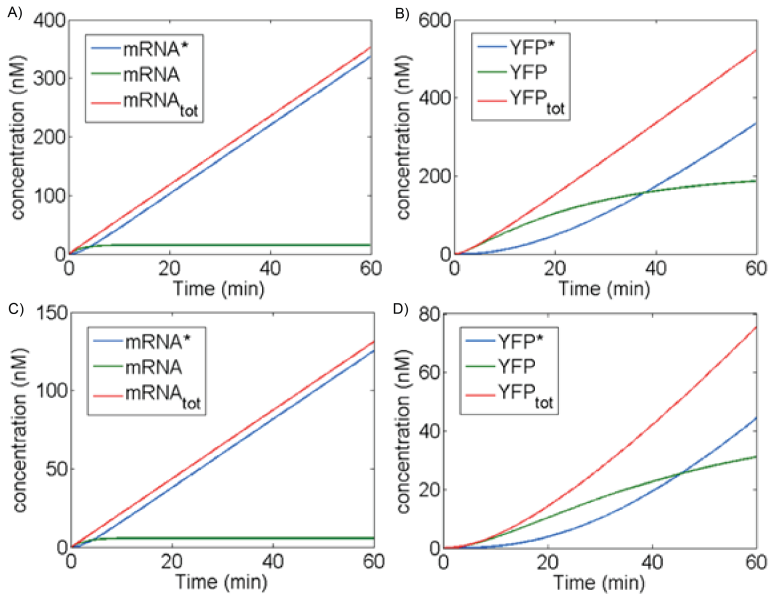


Figure 2.6: Simulated traces of the different mRNA and YFP species based on the reaction parameters shown in Table 2.1 and using 7 nM a), b) or 0.7 nM c), d) of DNA

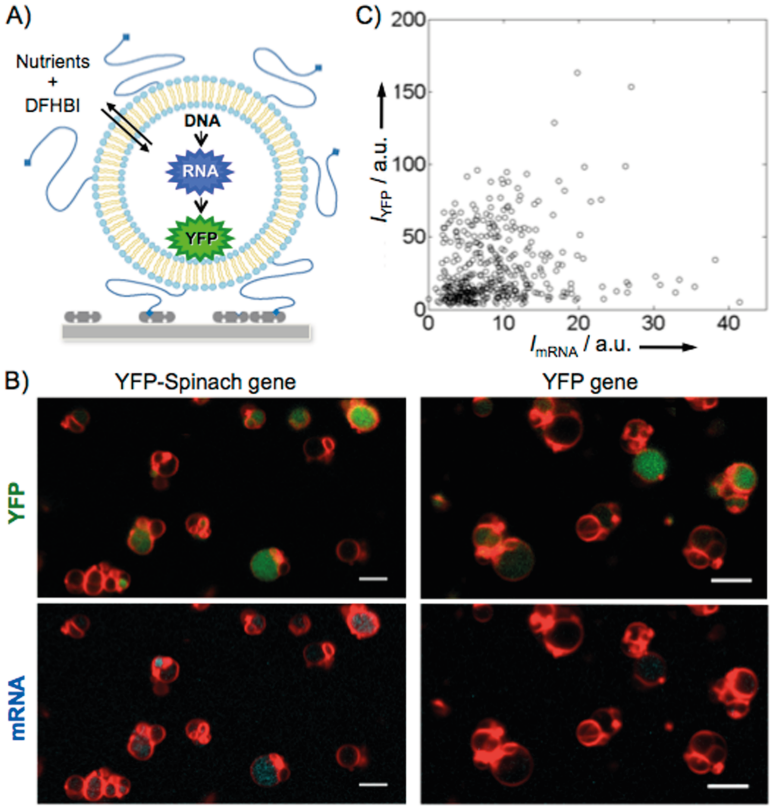


Figure 2.7: Vesicle-confined gene expression. A) Protein biosynthesis is compartmentalized inside surface-immobilized liposomes and internal reactions are fueled by inwards diffusion of the necessary nutrients. The fluorophore DFHBI was supplemented to the outside solution and lights up mRNA upon binding to the Spinach aptamer. B) Fluorescence confocal images of liposomes (membrane dye, red), synthesized YFP (green) and mRNA (Spinach, cyan). The YFP construct missing the Spinach sequence was used as a control. Bulk DNA concentration was 30 nM. Scale bars are 5 μm. C) 2D-intensity plot of YFP and Spinach signals inside individual vesicles measured using the YFP-Spinach gene (intensity was corrected using the YFP construct as a control, Figure 2.15).

which contained the amino acids, nucleotide triphosphates and transfer RNAs along with DFHBI 20 μM. We used fluorescence confocal microscopy to analyze the amounts of transcripts and YFPs generated in individual liposomes (Figure 2.7). Contrary to the relationship observed in bulk reactions (Figure 2.2E) stemming from the deterministic nature of gene expression, the level of output proteins does not scale with the amount of mRNA in vesicles (Figure 2.7C). In analogy to our recent finding that the number of encapsulated genes does not correlate with the yield of synthesized proteins [32], we hypothesize that the low-copy number of some components of the translation (or associated reactions) introduces nonlinearity and vesicle-to-vesicle variability. Another possible source of variability is the solute exchange efficiency due to differences in liposome size and lamellarity. The stochastic nature of mRNA and protein synthesis in micrometer-sized reaction vessels is a feature of our system that resembles that observed in living cells [33]. We postulate that such a property has to be apprehended as a design principle for the construction of a minimal cell.

2.3. CONCLUSION

In conclusion, we have fused the recently described Spinach aptamer to the DNA sequence of the YFP and demonstrated simultaneous tracking of mRNA and protein synthesis dynamics in a minimal gene expression system. We obtained quantitative insights into the production rates in bulk reactions and validated the method for imaging the levels of synthesized mRNA and proteins in vesicle-based artificial cell models. For the first time, we showed that compartmentalization of gene expression inside cell-sized vesicles leads to loss of correlation between the levels of synthesized mRNAs and proteins. This stochastic effect can be attributed to the low-copy number of some constituents of the translation process and is of primary importance toward the assembly of a gene-based synthetic cell. Ongoing work includes sequence optimization of the linkers flanking the Spinach sequence, the development of a complete model for gene expression dynamics – i.e. comprising the whole synthesis kinetics – in bulk and liposome-compartmentalized reactions, and mRNA multi-color imaging of reconstituted gene regulatory networks [34–36]. We believe that implementing Spinach-based mRNA readout can provide an additional level of control to tailor the rate and yield of protein production in cell-free expression systems[37].

2.4. MATERIALS AND METHODS

DFHBI SYNTHESIS

General procedure. All solvents and reagents were obtained commercially from Acros, Alfa Aesar, Fluka and Fluorochem. ^{1}H-NMR measurements were performed on a Bruker Avance-400 spectrometer (400 MHz). Chemical shifts are given in ppm (δ) relative to residual solvent peaks (CD_3OD, ^{1}H δ = 3.31). Coupling constants (J) are given in Hz.

DFHBI was synthesized using a modified literature procedure.

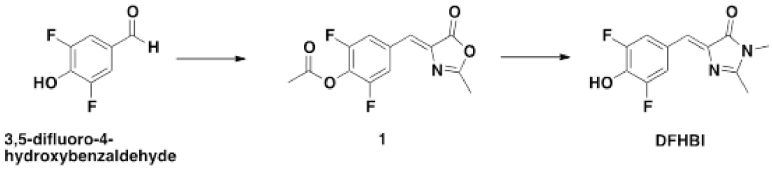


Figure 2.8: DFHBI synthesis.

(Z)-2,6-difluoro-((2-methyl-5-oxooxazol-4(5H)-ylidene)methyl)phenylacetate (1). 3,5-difluoro-4-hydroxybenzaldehyde (2.00 g, 12.65 mmol) was dissolved in acetic anhydride (15 mL) and heated to 140°C. After 2 h, sodium acetate (1.04 g, 12.65 mmol) and N-acetylglycine (1.48 g, 12.65 mmol) were added. ^{1}H-NMR showed complete consumption of the starting materials upon which the reaction mixture was poured out in cold ethanol (30 mL) and was stirred overnight. The solution was partially concentrated under vacuum (ca. 15 mL) leading to the formation of a yellow precipitate. The precipitate was filtered off and washed with cold ethanol (10 mL), hot water (10 mL) and heptane (10 mL), and subsequently dried under vacuum at 40°C yielding compound **1** as a yellow solid (1.62 g, 5.75 mmol, 45%). Spectroscopic data are in accordance with reported values [25].
^{1}H-NMR (400 MHz, CD3OD) δ 7.92 (d, J = 4.0, 2H), 7.07 (s, 1H), 2.42 (s, 3H), 2.37 (s, 3H)

(Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one (DFHBI). Compound **1** (800 mg, 2.84 mmol) was dissolved in ethanol (16 mL). Potassium carbonate (57 mg, 0.41 mmol) and 4-dimethylaminopyridine (DMAP, 17 mg, 0.14 mmol) were added while stirring. Note: we found the use of DMAP catalyst essential for achieving high yields and a selective conversion. Methylamine (33% in EtOH, 0.82 mL) was added and the reaction was heated to 100°C in a closed reflux setup. After 4 h the reaction mixture was allowed to cool down and was poured out in water (25 mL). The pH was adjusted to 3 using conc. HCl (aq.). A yellow precipitate was formed and filtered off. The solids were washed with cold ethanol and dried under vacuum at 50°C yielding yellow DFHBI (463 mg, 1.84 mmol, 76%). Spectroscopic data are in accordance with reported values [25].
^{1}H-NMR (400 MHz, CD3OD) δ 7.77 (d, J = 4.0, 2H), 6.89 (s, 1H), 3.18 (s, 3H), 2.40 (s, 3H).

DNA CONSTRUCT PREPARATION

The YFP-Spinach sequence, including the T7 promoter, RBS and leader sequences derived from the pRSET_B vector containing the EYFP gene (Invitrogen), was synthesized and cloned into the pMK_RQ vector using SFiI sites by Life Technologies. The linear DNA template was then prepared by polymerase chain reaction (PCR). The DNA template encoding for the EYFP protein only, i.e. devoid of the Spinach sequence, was prepared from the pRSET_B vector. The following forward and reverse primers (Biolegio) were used:
Forward primer YFP-Spinach/YFP:
5'- GCGAAATTAATACGACTCACTATAGGGAGACC-3'
Reverse primer YFP-Spinach:
5'- AAAAAACCCCTCAAGACCCGTTTAGAGG-3'

Figure 2.9: YFP-Spinach DNA template. The control gene devoid of the Spinach sequence corresponds to the same construct, where the sequence element into the dashed box is omitted.

Reverse primer YFP:
5'- TTCGCTATTACGCCAGATCCGGATATAGTTCC-3'

The sequence of the spinach with its two flanking linkers (underlined) is:
5'-CTAGCAGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGGTGTGGCTGCTTCG
GCAGTGCAGCTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTCGCGTCCTAGCA-3'

PCR reactions were performed with 10 ng of plasmid DNA as template, 1 Unit of Phusion polymerase (Finnzymes) in HF buffer containing 0.2 mM dNTPs, 0.2 μM forward and 0.2 μM reverse primers in a final volume of 50 μL. The generated linear DNA fragment was purified with the PCR Clean-up kit from Promega according to the manufacturer's protocol. The purity of the DNA product was checked on a 0.9% agarose gel using 50 ng of DNA and ethidium bromide staining. The 1-Kb Plus ladder from Invitrogen was used.

PURE SYSTEMS USED

The cell-free reaction solutions of PURE*frex* were kindly provided by GeneFrontier Corporation (Japan). The reaction solutions were initially separated into PURE*frex* enzyme mixture (T7 RNA polymerase, translation factors, energy recycling system, etc.), buffer (feeding) mixture (amino acids, NTPs, tRNAs, etc.), and ribosome. In addition to PURE*frex*, the commercially available PURExpress (New England Biolabs) was also used. The constituting enzymes are similar in both the PURE*frex* and PURExpress, with the exception that the enzymes in PURE*frex* are devoid of the histidine tag that was utilized for purification. Also note that the native composition of the PURE system has been modified by the supplier, such that the concentrations of the constituents of the PURExpress are not exactly known.

BULK REACTIONS AND KINETICS MEASUREMENTS

The following solution was prepared on ice and transferred in a 15-μL cuvette (Hellma): 1 μL of PURE*frex* enzyme mix, 1 μL of PURE*frex* ribosome, 10 μL of PURE*frex* feeding mix, 10 or 100 ng of DNA template (unless otherwise indicated), 0.5 μL of Superase inhibitor (10 units final; SUPERase.In, Ambion), 1 μL of DFHBI (20 μM final concentration) and the volume was adjusted to 20 μL with nuclease-free water. The cuvette was mounted in the temperature-controlled holder of a fluorescence spectrophotometer (Cary Eclipse from Varian) held at 37°C and the fluorescence was recorded every 9 s using the following excitation/emission wavelengths: Spinach, 460/502 nm; YFP, 515/528 nm. The cuvettes need to be thoroughly cleaned immediately after each experiment to maximize data reproducibility. Washing was done by successively filling the cuvette with Hellmanex 0.2%, KOH 1 M, nuclease-free water, EtOH 100% for 1 min in a bath sonicator, applying three washes with nuclease-free water in between the different reagent treatments.

MRNA PURIFICATION AND GEL ANALYSIS

Conversion of the measured fluorescence intensity of the Spinach signal into a mRNA concentration was done by calibrating the intensity of gel bands of mRNA produced in the PURE*frex* with reference mRNA of known concentration. The reference mRNA was produced from the YFP-Spinach DNA (14 nM) in the RiboMAX^{TM} Large Scale RNA production kit (Promega) congruous with the recommended protocol. All RNA purification was done with RNeasy MinElute Cleanup kit (Qiagen) following the manufacturer's protocol. RNA concentrations were determined using a Nanodrop (Thermo Scientific) with absorbance measurements performed at 280 nm. After 1-h expression of the YFP-Spinach template at 37 °C in the PURE*frex*, DNaseI (0.5 μL, 0.5 unit final; Promega) was added, while monitoring the fluorescence intensity of Spinach in the spectrofluorometer. After reaching a stable fluorescence signal the solution was collected for RNA purification. RNA was imaged on a 1.2% agarose gel containing EtBr and the band intensities were analyzed with ImageLab. mRNA production was alternatively studied by collecting 2-μL samples at 30 min intervals from a tube containing the PURE*frex* reaction mixture and 7.0 nM DNA of the YFP-Spinach construct (initial volume 50 μL) at 37 °C. The samples were diluted 50-fold in RNase-free water and stored at -20 °C until further use. Following RNA purification, 6 μL of the samples were imaged on a 1.2% agarose gel containing EtBr and the band intensities were subsequently analyzed with ImageLab. To correct for variability between different samples stemming from the RNA purification procedure and gel loading, the mRNA band intensities were normalized to the rRNA bands (1.5- or 2.9-kb bands).

LIPOSOME FORMATION, IMMOBILIZATION AND NUTRIENT SUPPLY

Surface-tethered lipid vesicles were prepared as previously described [6]. Briefly, 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC, Avanti Polar Lipids), 1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt) (DOPG, Avanti Polar Lipids), N-(6-tetramethylrhodaminethiocarbamoyl)-1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine (TRITC-DHPE, Invitrogen), and 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[biotinyl(polyethylene glycol)-2000] (DSPE-PEG-biotin, Avanti Polar Lipids), all dissolved in chloroform, were mixed at a molar ratio DOPC:DOPG:TRITC-DHPE:DSPE-PEG-biotin = 80:20:0.5:0.1 and added to 212-300-μm glass beads (Sigma-Aldrich) in a round-bottom glass flask at a mass ratio lipids:beads of 0.002:1. The solvent was rotary-evaporated at about 400 mbar overnight. Approximately 10 μL of newly prepared lipid-coated beads were poured in a reaction tube and immersed in 11.5 μL swelling solution consisting of 11.5 μL PURE*frex* enzyme mix, 1 μL PURE*frex* ribosome, 8.5 μL nuclease-free water, 0.5 μL Superase inhibitor (10 units final) and 0.5 μL linear DNA template to reach a final bulk DNA concentration of 30 nM. Note that in our previous study the PURExpress system (New England Biolabs) was used [6]. Liposomes were formed by swelling at 30°C for 2 h and were then subjected to four freeze-thaw cycles. Next, 1 μL of the liposome-containing solution was carefully harvested and immobilized at the bottom of a poly(dimethylsiloxane) chamber bound onto a microscope coverslip functionalized with BSA-biotin and neutravidin. Lastly, the surface-immobilized vesicles were diluted with 5 μL of the PURE*frex* feeding mix and DFHBI (20 μM final concentration) and incubated at 37°C for about 3 h.

FLUORESCENCE MICROSCOPY AND IMAGE ANALYSIS

A laser scanning confocal microscope (LSM710, Zeiss) equipped with a x40 oil immersion objective was used with the following fluorescence excitation/emission wavelengths: TRITC 543/557-797 nm, YFP 514/520-531 nm, and Spinach 458/461-520 nm. All measurements were performed at room temperature (20 °C). Fluorescence images were analyzed with the software ImageJ [38]. Single liposomes were localized and the Spinach and YFP fluorescence intensities inside the lumen were measured. The contribution of the YFP signal in the Spinach channel was determined with a control experiment using a DNA construct coding for the YFP but lacking the Spinach sequence and intensity values were corrected accordingly (Figure 2.15).

FLUORESCENCE CORRELATION SPECTROSCOPY

Calibration. Fluorescence correlation spectroscopy (FCS) measurements were performed using the MicroTime200 laser scanning confocal microscope from PicoQuant equipped with a x60 water-immersion objective (numerical aperture 1.2, Olympus UPlanSAPO). The 485-nm laser line (beam width 4.2 mm), a 50 μm pinhole and a band-pass emission filter 542 ± 25 nm (Semrock) were used. The laser was focused 20 μm into the solution. Calibration of the detection volume was performed with a solution of Alexa-488 (Invitrogen) of known concentration. Typically, a volume of 0.95 fL with waist w_{xy} = 270 nm and structure parameter k = 8.7 is obtained when fitting the measured autocorrelation function with the following equation for free 3-D diffusion of Alexa-488:

$$G(\tau) = \frac{1}{N} \cdot \left(1 + \frac{p}{1-p} exp\left(-\frac{\tau}{\tau_t}\right)\right) \cdot \left(1 + \frac{4D\tau}{w_{xy}^2}\right)^{-1} \cdot \left(1 + \frac{4D\tau}{k^2} w_{xy}^2\right)^{1/2} \tag{2.5}$$

where N is the average number of fluorescent molecules in the detection volume, D is the fluorophore diffusion coefficient, p is the fraction of molecules that occupy the triplet state and τ_t is the triplet lifetime. Actually, the effective volume, V_{eff}, was determined without prior knowledge of its shape at G(τ=0) as $V_{eff} = N(N_A C_{Alexa488})$, with N_A the Avogadro's number and $C_{Alexa488}$ a known concentration of Alexa-488 ranging from 0.5 to 10 nM. Note that afterpulsing effects were filtered out prior to calculating $G(0)$ through fluorescence lifetime correlation spectroscopy (FLCS) analysis. The FLCS algorithm was systematically applied on all collected FCS data. The autocorrelation curve of Alexa-488 was fitted to Eq. 2.5 using τ_t = 3 μs and D = 415 $\mu m^2 s^{-1}$ as fixed parameters [39] and values for w_{xy} and k were obtained.

YFP measurements. The YFP protein was produced from the YFP-Spinach DNA construct with the PURE*frex* or the PURExpress in bulk reactions for 4 h as described in the *bulk reactions* section. The solution was diluted in the PURE buffer (50 mM HEPES, 100 mM potassium glutamate, 13 mM magnesium acetate, pH 7.6), supplemented with 0.05% Tween-20 to minimize depletion effects due to adsorption of YFP at the air-aqueous solution interface. A 50μL droplet was deposited onto a clean glass coverslip. The autocorrelation curves were fitted to Eq. 2.5 with the parameters $w - xy$ and k fixed to their values obtained by calibration. The concentration of YFP was determined as:

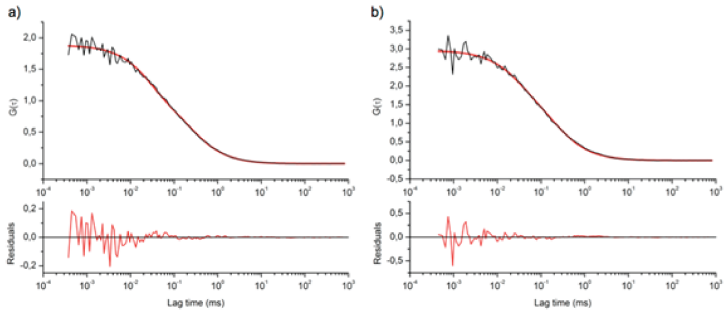


Figure 2.10: Fluorescence autocorrelation curves of a 300- or 1,800-fold diluted solution of YFP synthesized in the PURE*frex* a) or PURExpress b), respectively. The fits to Eq. 2.5 (red lines) and the fit residuals (bottom graphs) are also shown.

$$C_{YFP} = \frac{N}{N_A \pi^{3/2} w_{xy}^3 k} \tag{2.6}$$

The actual concentration of synthesized YFP was then obtained after correcting for the dilution factor. All measurements were performed at room temperature (20°C).

ABSORBANCE MEASUREMENTS

Absorbance of PURE samples was measured at an excitation wavelength of 515 nm using a molar extinction coefficient of 83,400 $M^{-1}cm^{-1}$ for YFP [40] (NanoDrop 2000, Thermo Scientific). The PURE system devoid of DNA (background) was used as the blank.

SIDE-BY-SIDE PROBES

We also investigated the side-by-side probes RNA labeling method to monitor mRNA production in the PURE system. Adjacent hybridization sites for short 14-nt oligoprobes with either a donor dye (Cy3) or an acceptor dye (Cy5) were designed in the 3'-UTR of a deGFP construct (plasmid from V. Noireaux[41]), 25 nt after the stop codon and 27 nt before the T7-terminator. The donor probe, 5'-CAGCCAACUCAGCU-Cy3, and the acceptor probe, Cy5-UCGGGCUUGUUAG-3' (both 2'-O-methylated, Biolego) bind with a spacing of 5 nt on the target sequence of the newly synthesized mRNA, bringing the fluorescence pair in close proximity, which leads to a specific fluorescence resonance energy transfer (FRET) signal.

2.4.1. SUPPLEMENTARY INFORMATION

DETERMINING THE CONCENTRATION OF SYNTHESIZED YFP

We first sought to measure the concentration of YFP produced with the PURE*frex* using absorbance measurements. However, the amount of synthesized YFP was not high enough to obtain reliable values. Therefore, we performed FCS measurements as sub-nanomolar concentrations of fluorescent dyes can be detected. The autocorrelation curve was fitted to Eq. 2.5 (Figure 2.10a). A YFP concentration of 0.41 μM was found (Table 2.2). Diffusion coefficient values are similar to that previously reported for the

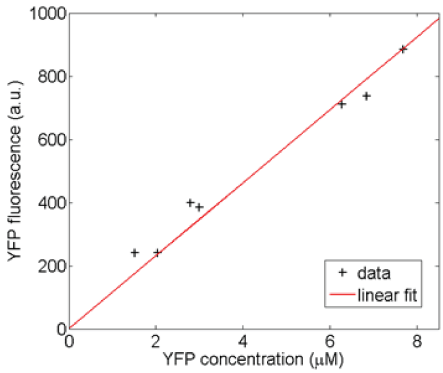


Figure 2.11: Calibration of YFP concentration generated with the PURExpress. Data from three independent experiments are reported.

	PURExpress	PURE*frex*
C_{YFP}, diluted	0.79 ± 0.05 nM	1.38 ± 0.12 nM
C_{YFP}, dilution-corrected	1.42 μM	0.41 μM
Diffusion coefficient	96 ± 16 $\mu m^2/s$	93 ± 20 $\mu m^2/s$
C_{YFP} Absorbance	6.8 μM	N.D.
Triplet fraction	0.22 ± 0.06	0.33 ± 0.06

Table 2.2: The diffusion coefficients and triplet fraction parameters of YFP extracted from the FCS measurements are also indicated. Triplet time was fixed to 25 μs.

eGFP [39]. Because YFP is produced at larger amounts using the PURExpress, its concentration could also be measured by absorbance. The results of FCS and absorbance measurements are summarized in Table 2.2. The YFP concentration value measured by absorbance is 4.7-fold higher than the one determined by FCS (Table2.2). Although FCS is a highly sensitive method, the possibility that YFPs are partly depleted to the water-air interface cannot be excluded, which would lead to an underestimation of the concentration. Therefore, we opted for the absorbance measurements as a more reliable method to perform the calibration curve for YFP concentrations.

Converting YFP fluorescence intensities into concentrations was realized by a linear fit of the calibration curve shown in Figure 2.11. We found that 1 A.U. (fluorescence) corresponds to 8.5 nM and 0.64 nM YFP in the medium and high voltage settings, respectively. As we want to determine the concentration of YFP synthesized in the PURE*frex* using the conversion coefficient obtained in the PURExpress, we had to verify that the fluorescence properties of the YFP produced in either expression system are similar. To this aim we ensured that the excitation and emission spectra are identical and remain unchanged upon dilution (data not shown).

QUANTIFYING THE YIELD OF SYNTHESIZED MRNA

The fluorescence intensity of the Spinach signal was first corrected for non-specific signal arising from two contributions: the initial background from the free DFHBI molecules

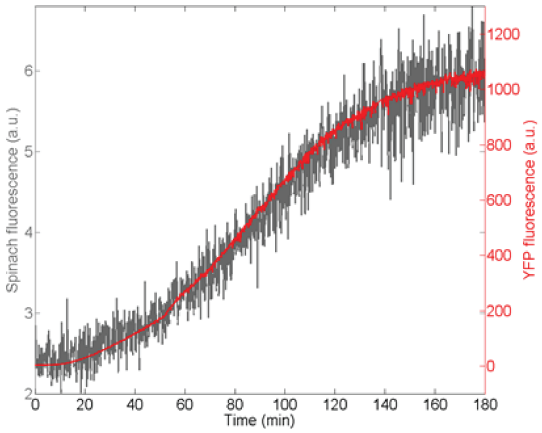


Figure 2.12: Time course of the YFP fluorescence (red trace) and corresponding cross-talk in the Spinach channel (gray trace). A concentration of 0.7 nM of the control YFP construct was used along with 20 μM DFHBI.

and from the PURE system, and secondly the cross-talk of the produced YFP into the Spinach channel. The first correction is done by subtracting the initial value of the Spinach signal from the total signal. The leakage of the YFP signal in the Spinach channel was determined by a control experiment, where the YFP construct lacking the Spinach sequence was expressed in the presence of 20 μM DFHBI as shown in Figure 2.12. We find that 0.33% of the YFP signal contributes to the Spinach channel and therefore apply this subtraction at each individual time point as a second correction.

The corrected fluorescence intensity units were converted into mRNA concentrations by using a calibration curve with mRNA of known concentrations produced in an in vitro transcription kit (Figure 2.13). The kinetics of mRNA synthesis in the PURE*frex* was monitored by spectrofluorometry. Transcription was stopped with the addition of DNaseI and the stable fluorescence intensity was measured. The solution was harvested and RNA was then purified and analyzed on gel (Figure 2.13a). The band intensity was appended on the calibration curve and the mRNA concentration in the initial PURE*frex* solution was calculated Figure 2.13b). We found that 1 A.U. (fluorescence) corresponds to 0.33 μM and 0.025 μM mRNA in the medium and high voltage settings, respectively (averaged values over three independent experiments, standard deviation is 20% of the mean).

A representative time series of mRNA amounts synthesized in the PURE*frex* is shown in Figure 2.2c in the main text and the corresponding mRNA concentrations were reported in Figure 2.2a. The discrepancy between fluorescence-based and gel-based concentrations observed after 90 min makes it impossible to reliably assess the dynamics of mRNA synthesis on longer time scales. We could exclude photobleaching of spinach as the source of the kinetics flattening by measuring constant fluorescence signal over several hours at fixed mRNA concentration (data not shown). One can invoke different

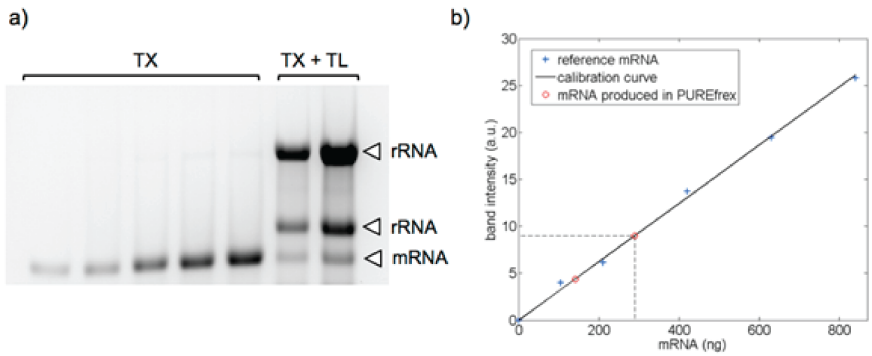


Figure 2.13: Calibration of mRNA concentration. a) Reference mRNA was produced in a transcription (TX)-only reaction, purified and gel analyzed as described in Section 1.5 (first four lanes). The last two lanes correspond to the mRNA produced in the PURE*frex* from the same DNA template (14 nM) and diluted at two different dilution factors. The corresponding ribosomal RNA (rRNA) bands are also indicated. b) Calibration curve plotted using reference mRNAs of known concentrations. The amount of mRNA synthesized in thePURE*frex* was determined by reporting its gel band intensity onto the calibration curve.

reaction environments in cuvette versus test tube experiments as a plausible explanation. Another is the instability of the Spinach conformation in the PURE system. Greater stability may be obtained by optimizing the sequence of the Spinach-flanking regions to avoid the possible formation of interfering secondary structures.

mRNA IS NOT SIGNIFICANTLY DEGRADED

Two studies reported contradicting results regarding mRNA degradation in the commercial PURExpress system. In the first one, a negligible degradation rate was indirectly obtained by model fitting the dynamics of transcript and protein synthesis [16]. In the second, qRT-PCR and FRET signal derived from the hybridization of side-by-side oligonucleotide probes onto target mRNA were jointly employed to show that mRNA degradation occurred at a rate 10-fold higher than was found in the previous study [18]. We examined mRNA stability in the PURE*frex* reaction mixture by using pre-synthesized mRNA of known concentration. The solution was incubated at 37°C and 2-μL samples were harvested every 30 min for RNA purification and gel analysis. The mRNA band intensities, normalized at every time point to that of the corresponding 1.5-kb rRNA for correcting differences in gel loading, demonstrate that the transcript is not significantly degraded within the first 3 h of expression (Figure 2.14). This result is in agreement with the study of Stögbauer et al. [16] where a degradation rate constant of 7.8 x 10^{-4} min^{-1} was determined by model fitting. However, this contrasts with the work of Niederholtmeyer et al.[18], in which a value of 8.5 x 10^{-3} min^{-1} was measured. The profiles of mRNA levels expected from these two studies are reported in Figure 2.14b for comparison.

LIPOSOME-CONFINED REACTIONS

Molecular diffusion across the lipid bilayer. The possible mechanisms for membrane permeability, including nutrient uptake and leakage of entrapped components, have been experimentally addressed and extensively discussed in our previous studies.

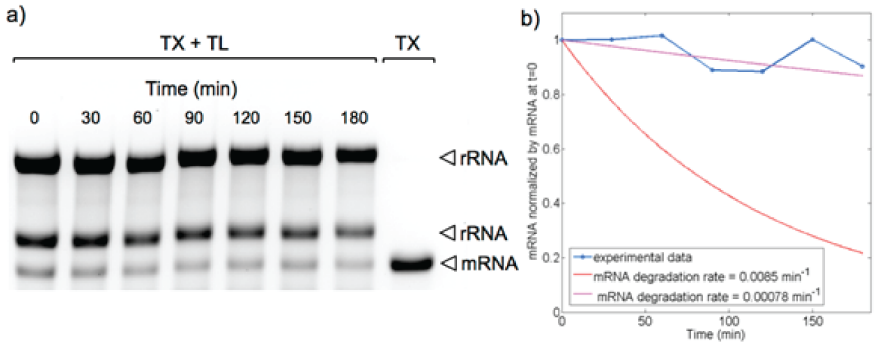


Figure 2.14: a) RNA gel showing presynthesized mRNA (from the YFP-Spinach DNA) used as a template in a translation (TL) reaction performed with the PURE*frex* for different time periods. The ribosomal RNA (rRNA) bands were used for normalization. The right-most lane shows the purified mRNA presynthesized in a transcription (TX)-only reaction. b) Quantification of the amount of presynthesized mRNA normalized to that of time zero and to the band intensity of the 1.5-kb rRNA at the corresponding time point. The purple and red traces correspond to the profiles of mRNA levels as found by Stögbauer et al.[16] and Niederholtmeyer et al. [18], respectively.

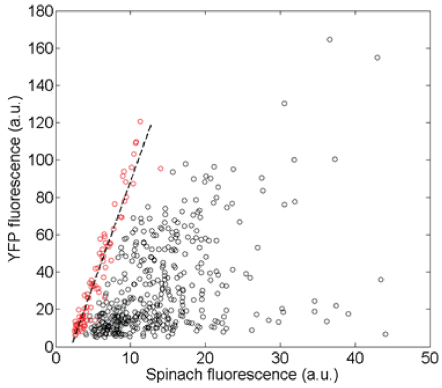


Figure 2.15: Uncorrected two-dimensional intensity plot of YFP and Spinach fluorescence inside individual liposomes. Each point represents the mean fluorescence intensities from a single vesicle. The YFP-Spinach (black circles) or the YFP (red circles) construct was used at a concentration of 30 nM. The dashed line is a linear fit to the data obtained with the YFP gene.

Cross-talk correction. The percentage of YFP signal leaking to the Spinach channel was quantified using the YFP gene by linearly fitting the data (dashed line, Figure 2.15) and a value of 8.7% was obtained. The Spinach signal measured in each liposome using the YFP-Spinach construct was then corrected for cross-talk of the YFP fluorescence (Figure 2.7c, main text) as follows:

$$I^{\text{Spinach}}_{\text{corrected}} = I^{\text{Spinach}}_{\text{measured}} - I^{\text{YFP}}_{\text{measured}} \times 0.087 \tag{2.7}$$

2.5. ACKNOWLEDGEMENTS

We thank Yutetsu Kuruma and Takuya Ueda for collaboration on the PURE system. We thank Jos M. Poolman, Rienk Eelkema and Jan van Esch for synthesizing DFHBI. We thank Andrew Scott and David Dulin for useful discussions. I would like to thank Christophe for writing most of this paper presented in this chapter that was published as [42]. Thanks to Zohreh who taught me patiently how to perform liposome experiments, to Maurits Kok for the FCS measurements and editing of the paper, to Roeland van Wijk for his contribution to fitting procedures and Figure 2.5 and Jonne Moeskops for preliminary investigations into the Spinach constructs, and to Ilja for help in designing and producing the templates used in this study and for always being available for kind advice and introductions on lab-related issues.

REFERENCES

[1] W. Yu, K. Sato, M. Wakabayashi, T. Nakaishi, E. P. Ko-Mitamura, Y. Shima, I. Urabe, and T. Yomo, *Synthesis of functional protein in liposome.* Journal of bioscience and bioengineering **92**, 590 (2001).

[2] S.-i. M. Nomura, K. Tsumoto, T. Hamada, K. Akiyoshi, Y. Nakatani, and K. Yoshikawa, *Gene expression within cell-sized lipid vesicles.* Chembiochem **4**, 1172 (2003).

[3] V. Noireaux and A. Libchaber, *A vesicle bioreactor as a step toward an artificial cell assembly.* Proceedings of the National Academy of Sciences of the United States of America **101**, 17669 (2004).

[4] G. Murtas, Y. Kuruma, P. Bianchini, A. Diaspro, and P. L. Luisi, *Protein synthesis in liposomes with a minimal set of enzymes.* Biochemical and biophysical research communications **363**, 12 (2007).

[5] H. Saito, Y. Kato, M. Le Berre, A. Yamada, T. Inoue, K. Yosikawa, and D. Baigl, *Time-Resolved Tracking of a Minimum Gene Expression System Reconstituted in Giant Liposomes,* ChemBioChem **10**, 1640 (2009).

[6] Z. Nourian, W. Roelofsen, and C. Danelon, *Triggered Gene Expression in Fed-Vesicle Microreactors with a Multifunctional Membrane,* Angew. Chem. Int. Ed **51**, 3114 (2012).

[7] C. E. Hodgman and M. C. Jewett, *Cell-free synthetic biology: Thinking outside the cell,* Metabolic Engineering **14**, 261 (2012).

[8] M. C. Jewett, K. A. Calhoun, A. Voloshin, J. J. Wuu, and J. R. Swartz, *An integrated cell-free metabolic platform for protein production and synthetic biology.* Molecular systems biology **4**, 220 (2008).

[9] J. W. Whittaker, *Cell-free protein synthesis: The state of the art,* (2013).

[10] E. D. Carlson, R. Gan, C. E. Hodgman, and M. C. Jewett, *Cell-free protein synthesis: Applications come of age,* (2012).

[11] Y. Shimizu, a. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda, *Cell-free translation reconstituted with purified components.* Nature biotechnology **19**, 751 (2001).

[12] Y. Shimizu, T. Kanamori, and T. Ueda, *Protein synthesis by pure translation systems.* Methods (San Diego, Calif.) **36**, 299 (2005).

[13] J. M. Pedraza and J. Paulsson, *Effects of molecular memory and bursting on fluctuations in gene expression.* Science (New York, N.Y.) **319**, 339 (2008).

[14] E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long, *Localization of ASH1 mRNA Particles in Living Yeast,* Molecular Cell **2**, 437 (1998).

[15] S. Tyagi and F. R. Kramer, *Molecular Beacons: Probes that Fluoresce upon Hybridization,* Nature biotechnology **14**, 303 (1996).

[16] T. Stögbauer, L. Windhager, R. Zimmer, and J. O. Rädler, *Experiment and mathematical modeling of gene expression dynamics in a cell-free system.* Integrative biology : quantitative biosciences from nano to macro **4**, 494 (2012).

[17] Y. Sei-Iida, H. Koshimoto, S. Kondo, and A. Tsuji, *Real-time monitoring of in vitro transcriptional RNA synthesis using fluorescence resonance energy transfer,* Nucleic acids research **28**, e59 (2000).

[18] H. Niederholtmeyer, L. Xu, and S. J. Maerkl, *Real-Time mRNA Measurement during an in Vitro Transcription and Translation Reaction Using Binary Probes,* ACS synthetic biology **2**, 411 (2013).

[19] J. R. Babendure, S. R. Adams, and R. Y. Tsien, *Aptamers Switch on Fluorescence of Triphenylmethane Dyes,* Journal of the American Chemical Society **125**, 14716 (2003).

[20] B. A. Sparano and K. Koide, *A strategy for the development of small-molecule-based sensors that strongly fluoresce when bound to a specific RNA,* Journal of the American Chemical Society **127**, 14954 (2005).

[21] B. A. Sparano and K. Koide, *Fluorescent sensors for specific RNA: A general paradigm using chemistry and combinatorial biology,* Journal of the American Chemical Society **129**, 4785 (2007).

[22] S. Sando, A. Narita, M. Hayami, and Y. Aoyama, *Transcription monitoring using fused RNA with a dye-binding light-up aptamer as a tag: a blue fluorescent RNA,* Chemical Communications **23**, 3858 (2008).

[23] J. Lee, K. H. Lee, J. Jeon, A. Dragulescu-Andrasi, F. Xiao, and J. Rao, *Combining SELEX screening and rational design to develop light-up fluorophore-RNA aptamer pairs for RNA tagging,* ACS Chemical Biology **5**, 1065 (2010).

[24] A. Murata, S.-i. Sato, Y. Kawazoe, and M. Uesugi, *Small-molecule fluorescent probes for specific RNA targets.* Chemical communications (Cambridge, England) **47**, 4712 (2011).

[25] J. S. Paige, K. Y. Wu, and S. R. Jaffrey, *RNA mimics of green fluorescent protein.* Science (New York, N.Y.) **333**, 642 (2011).

[26] J. S. Paige, T. Nguyen-Duc, W. Song, and S. R. Jaffrey, *Fluorescence Imaging of Cellular Metabolites with RNA,* Science **335**, 1194 (2012), arXiv:NIHMS150003 .

[27] C. A. Kellenberger, S. C. Wilson, J. Sales-Lee, and M. C. Hammond, *RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP,* Journal of the American Chemical Society **135**, 4906 (2013), arXiv:NIHMS150003 .

[28] S. Tyagi, *Imaging intracellular RNA distribution and dynamics in living cells.* Nature methods **6**, 331 (2009), arXiv:NIHMS150003 .

[29] R. Iizuka, M. Yamagishi-Shirasaki, and T. Funatsu, *Kinetic study of de novo chromophore maturation of fluorescent proteins,* Analytical Biochemistry **414**, 173 (2011).

[30] E. Karzbrun, J. Shin, R. H. Bar-Ziv, and V. Noireaux, *Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System,* Physical Review Letters **106**, 048104 (2011).

[31] J. H. Kim and R. G. Larson, *Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules.* Nucleic acids research **35**, 3848 (2007).

[32] Z. Nourian and C. Danelon, *Linking genotype and phenotype in protein synthesizing liposomes with external supply of resources.* ACS synthetic biology **2**, 186 (2013).

[33] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic gene expression in a single cell.* Science (New York, N.Y.) **297**, 1183 (2002).

[34] V. Noireaux, R. Bar-Ziv, and A. Libchaber, *Principles of cell-free genetic circuit assembly.* Proceedings of the National Academy of Sciences of the United States of America **100**, 12672 (2003).

[35] J. Shin and V. Noireaux, *An E. coli cell-free expression toolbox: Application to synthetic gene circuits and artificial cells,* ACS Synthetic Biology **1**, 29 (2012).

[36] D. K. Karig, S. Iyer, M. L. Simpson, and M. J. Doktycz, *Expression optimization and synthetic gene networks in cell-free systems,* Nucleic Acids Research **40**, 3763 (2012).

[37] G. Rosenblum, C. Chen, J. Kaur, X. Cui, Y. E. Goldman, and B. S. Cooperman, *Real-time assay for testing components of protein synthesis,* Nucleic Acids Research **40** (2012), 10.1093/nar/gks232.

[38] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, *Fiji: an open-source platform for biological-image analysis,* Nature Methods **9**, 676 (2012), arXiv:1081-8693 .

[39] Z. Petrasek and P. Schwille, *Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy,* Biophysical journal **94**, 1437 (2008).

[40] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien, *A guide to choosing fluorescent proteins,* Nature Methods **2**, 905 (2005).

[41] J. Shin and V. Noireaux, *Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70.* Journal of biological engineering **4**, 8 (2010).

[42] P. van Nies, Z. Nourian, M. Kok, R. van Wijk, J. Moeskops, I. Westerlaken, J. M. Poolman, R. Eelkema, J. H. van Esch, Y. Kuruma, T. Ueda, and C. Danelon, *Unbiased tracking of the progression of mRNA and protein synthesis in bulk and in liposome-confined reactions.* Chembiochem : a European journal of chemical biology **14**, 1963 (2013).

3

IMPROVED SPINACH-BASED TRANSCRIPTION-TRANSLATION DUAL REPORTER

With rising interest in utilizing cell-free gene expression systems in bottom-up synthetic biology projects, novel labeling tools need to be developed to accurately report the dynamics and performance of the biosynthesis machinery operating in various reaction conditions. Monitoring the transcription activity has been simplified by the Spinach technology, an RNA aptamer that emits fluorescence upon binding to a small organic dye. Recently, we tracked the fluorescence of Spinach-tagged messenger RNA (mRNA) and its translation product the yellow fluorescent protein (YFP), both synthesized in the PURE (Protein synthesis Using Recombinant Elements) system from a DNA template. Building on our previous study, we describe in the first part of this chapter an improved Spinach reporter with modified flanking sequences that confer higher propensity for aptamer folding and thus enhanced fluorescence brightness. Hence, the kinetics of mRNA and YFP production could be simultaneously monitored with unprecedented sensitivity. In the second part of this chapter, we demonstrate that the new Spinach construct greatly enhances mRNA detection when gene expression is confined inside self-assembled lipid vesicles. This allowed us to quantify gene expression in individual liposomes for the first five hours from time lapse images obtained with confocal microscopy. We find that there are large differences in gene expression kinetics between individual liposomes, corroborating the conclusion from our previous study. Additionally, we find that a subset of expressing liposomes produces more mRNA and/or YFP than measured in bulk experiments. Further improvements to enhance gene expression in liposomes are discussed.

Parts of this chapter have been published as *Monitoring mRNA and Protein Levels in Bulk and in Model Vesicle-Based Artificial Cells*, Pauline van Nies, Alicia Soler Canton, Zohreh Nourian, Christophe Danelon, *Methods in Enzymology*, Volume 550:187-214, 2015

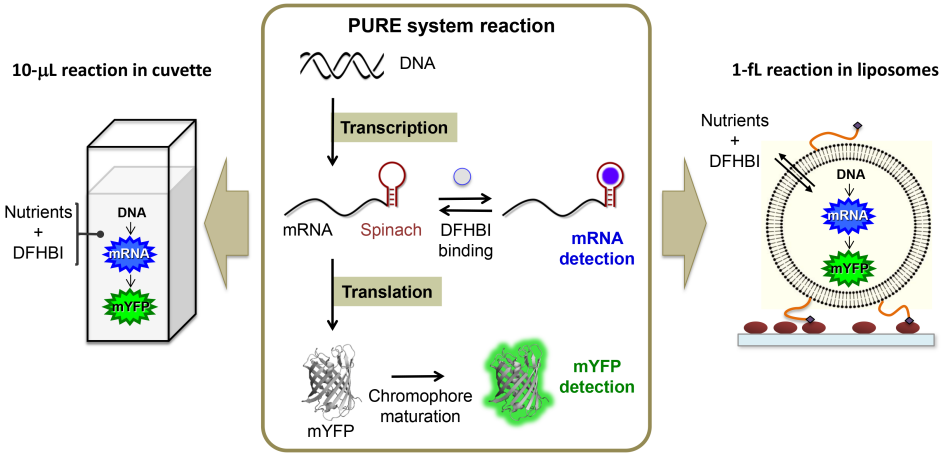


Figure 3.1: Schematic overview of our two-reporter assay for mRNA and protein levels in bulk (microtubes or cuvettes) and in liposome-confined gene expression reactions. The Spinach technology consisting of an RNA aptamer sequence that binds and turn the chromophore DFHBI into a fluorescent state was used to monitor transcription activity.

3.1. INTRODUCTION

In our previous study, we expressed Spinach [1], an RNA aptamer that forms a fluorescent complex upon binding with DFHBI, to monitor transcription activity in a cell-free expression system [2]. We noticed that the Spinach fluorescence signal consistently reports the kinetics of RNA production only during the first 90 min and then it deviates to give underestimated intensity values, a result we attributed to suboptimal folding stability of the aptamer at 37 °C. Here, we sought to design a new Spinach reporter with improved folding in cell-free gene expression systems in order to reliably monitor the complete transcription dynamics in real-time.

Recently, a second generation of Spinach with improved brightness [3] and of GFP-like fluorophores with tailored spectral properties [4] have been developed, further expanding the toolbox of the Spinach technology. Additionally, Spinach can be fused to aptamers that bind to specific metabolites, which in turn modulates the Spinach fluorescence [5, 6].

Herein, we report on a variant of Spinach with modified flanking sequences that confer improved stability resulting in enhanced brightness in bulk and liposome-confined reactions. We characterize the Spinach fluorescence and validate its use in combination with a fluorescent protein reporter for quantitative monitoring of transcription and translation activities.

3.2. CONSTRUCT DESIGN AND CHARACTERIZATION OF IMPROVED SPINACH FLUORESCENCE IN THE PURE SYSTEM

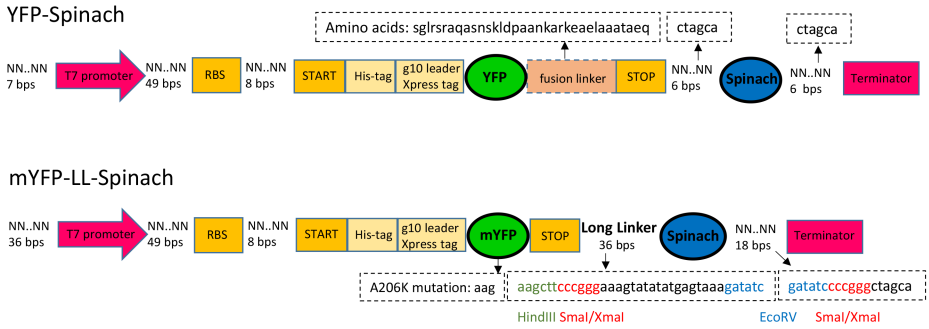


Figure 3.2: Schematic representation of the two DNA constructs primarily used in this study. Their main features and regulatory elements are depicted. The new construct mYFP-LL-Spinach has been designed based on our previously described YFP-Spinach gene [2]. The abbreviation RBS stands for ribosome binding site.

DESIGN OF NEW SPINACH CONSTRUCT

In our previous construct, denoted YFP-Spinach, the Spinach was flanked with 6-nucleotide sequences, i.e. six nucleotides after the translation stop codon and six nucleotides before the transcription termination sequence (Figure 3.2). In the new construct, named mYFP-LL-Spinach, the 6-nucleotide linker upstream of Spinach was extended to 36 nucleotides with the intention to minimize possible interference between aptamer folding and ribosome activity. Three cloning sites were introduced into this linker, of which the EcoRV and the SmaXI are also present in the 18-nucleotide linker downstream of Spinach (Figure 3.2). The palindromic nature of the EcoRV restriction site enables a partial hybridization of the new flanking regions as shown using the Mfold Web Server [7], thus extending and stabilizing the stem-I structure of the Spinach aptamer. The mYFP sequence was derived from the pRSETB-YFP plasmid with two modifications: the A206K mutation was introduced to generate a monomeric variant of YFP and the fusion linker was deleted (Figure 3.2). Both YFP-Spinach and mYFP-LL-Spinach genes were cloned into the pmK_RQ vector using SfiI sites by Life Technologies. The linear DNA templates used for gene expression were prepared by polymerase chain reaction (PCR) (Phusion polymerase, Finnzymes) with primers 5'-GCGAAATTAATACGACTCACTATAGGGAGACC-3' (forward primer YFP-Spinach) or 5'-GAATTGAAGGAAGGCCGTCAAG-3' (forward primer mYFP-LL-Spinach), and 5'-AAAAAACCCCTCAAGACCCGTTTAGAGG-3' (reverse primer for both constructs). The construct mYFP lacking the LL-Spinach domain was produced by removing the Spinach sequence with SmaI. The corresponding linear DNA template was generated by PCR using the same forward and reverse primers used for the mYFP-LL-Spinach.

CHARACTERIZATION OF NEW SPINACH FLUORESCENCE

The characterization of the mYFP-LL-Spinach was carried out by the same methods as described in the previous chapter. Before performing kinetics measurements, we confirmed that fluorescence excitation and emission spectra of Spinach were similar for the YFP-Spinach and mYFP-LL-Spinach transcripts (Figure 3.3A and [2]). Next, we examined the time courses of Spinach fluorescence arising from transcription of the YFP-Spinach

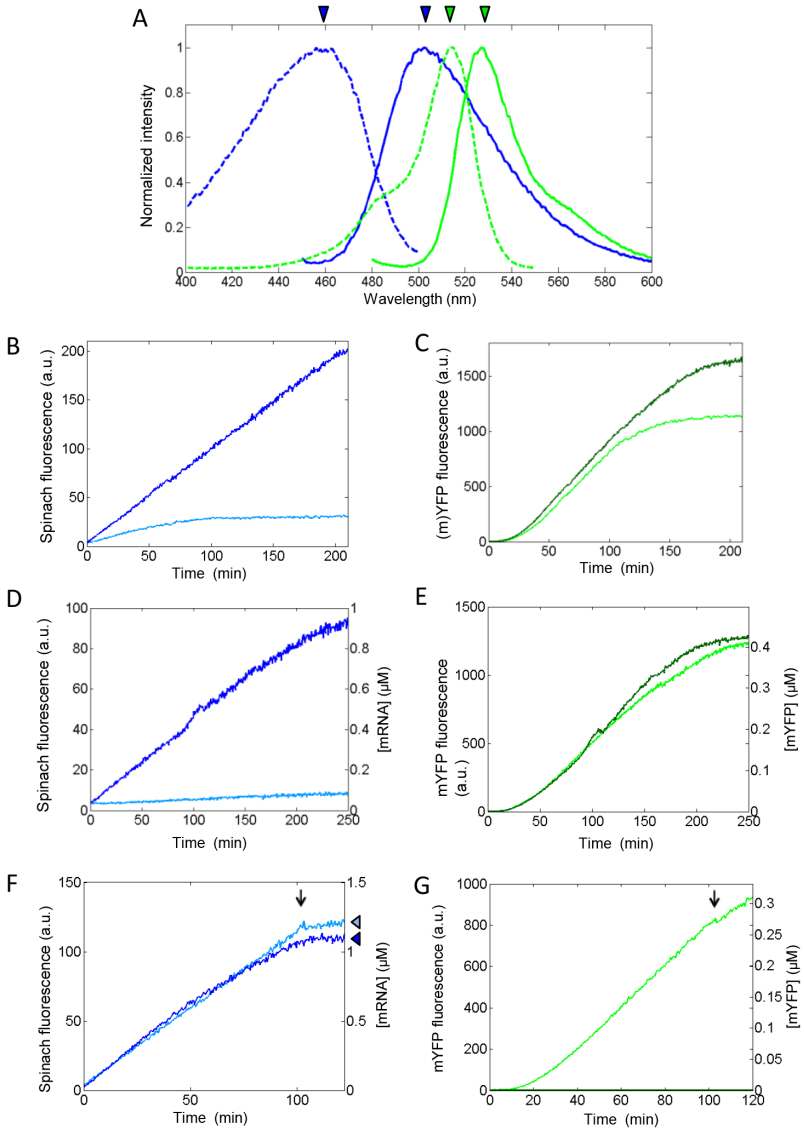


Figure 3.3: (A) Fluorescence excitation (dashed lines) and emission (solid lines) spectra of LL-Spinach (blue) and mYFP (green) measured in the PURE*frex* expressing the mYFP-LL-Spinach gene. The LL-Spinach spectra were measured in the PURE*frex* ΔR, that is devoid of ribosome, in the presence of 20 μM DFHBI. The mYFP spectra were collected in the PURE*frex* without DFHBI. The arrowheads depict the excitation and emission wavelengths used for kinetics measurements. (B) Fluorescence intensity profiles of Spinach produced from the mYFP-LL-Spinach (dark blue) or YFP-Spinach (light blue) construct. (C) Apparent kinetics of mYFP (dark green) and YFP (light green) synthesis monitored simultaneously as in (B). (B, C) DNA concentration for both genes was 7.4 nM. (D) Plots of LL-Spinach fluorescence versus time using the mYFP-LL-Spinach (dark blue) or mYFP (light blue) construct. (E) Apparent kinetics of mYFP produced from the mYFP-LL-Spinach (dark green) or mYFP (light green) construct and monitored simultaneously as in (D). (D, E) DNA concentration for both genes was 0.74 nM. (F) Progression of Spinach fluorescence versus time in a PURE*frex* ΔR (dark blue) or PUREfrex (light blue) reaction starting from 11.7 nM of the mYFP-LL-Spinach DNA. The arrowheads on the right axis point to the final intensity values used for calculating the conversion factor between fluorescence a.u. and mRNA concentration. (G) Apparent kinetics of mYFP synthesis in a PURE*frex* ΔR (dark green) or PUREfrex (light green) reaction monitored simultaneously as in (F). (F, G) The arrow at around 100 min indicates the addition of DNaseI to stop transcription. (D, F) Concentrations of mRNA were calculated using a conversion factor of 10 nM / a.u. (E, G) Concentrations of mYFP were calculated using a conversion factor of 0.33 nM / a.u.

and mYFP-LL-Spinach constructs (Figure 3.3B). Strikingly, transcription of the mYFP-LL-Spinach DNA leads to 8-fold higher Spinach fluorescence intensity (at 7.4 nM DNA after 3 h) compared to that with the YFP-Spinach construct. Moreover, the signal from the mYFP-LL-Spinach template increases for 7 h (not shown). In contrast, the fluorescence arising from transcription of the YFP-Spinach gene levels off after about 90 min Figure 3.3B), which is premature given the linear increase of mRNA amount observed on gel (Figure 2.2 [2]). Compared to tRNA scaffold-stabilized Spinach [8], our LL-Spinach is shorter, reducing consumption of NTPs.

Factors that could affect the Spinach fluorescence intensity and dynamics in transcription-translation reactions include aptamer misfolding due to thermal instability at 37°C in the PURE system and interference of the translation process with folding. We consider unlikely a change in the photochemical properties of Spinach (molar extinction coefficient or quantum yield) [3]. Note that the excitation and emission spectra of Spinach and LL-Spinach coincide (Figure 3.3A). The possible influence of translating ribosomes on Spinach fluorescence can be examined by monitoring transcription activity in the presence (+R) or in the absence (ΔR) of ribosomes. The apparent kinetics of mRNA synthesis is the same whether the transcript is translated or not (Figure 3.3F), indicating that translation does not affect the folding propensity of LL-Spinach. Furthermore, a change of DNA concentration from 0.74 nM to 7.4 nM was accompanied by a 2.5-fold increase of the transcription rate (compare Figure 3.3B withFigure 3.3D).

We also investigated the influence of the insertion of the Spinach tag dowsntream the YFP coding sequence of the mRNA on the translation activity. Two interfering mechanisms could be envisaged: First, the larger depletion of ATP and GTP upon transcription of the longer mYFP-LL-Spinach gene compared to that with the mYFP template may lead to a shortening of the protein synthesis duration since ATP and GTP are also consumed during tRNA aminoacylation and translation, respectively. Second, the global folding of mRNA can differ in the presence of LL-Spinach, which may alter the rate of translation despite the helicase activity of the ribosome. The results indicate that the apparent kinetics of protein synthesis and the final concentrations are nearly identical for the mYFP-LL-Spinach and mYFP genes (Figure 3.3E), consistent with a neutral effect of the Spinach reporter on the yield and cessation time of protein production. We had already verified in the previous study that the DFHBI chromophore does not interfere with transcription and translation reactions, and that the crosstalk of the mYFP fluorescence in the Spinach channel is negligible (0.3%) compared the improved Spinach fluorescence signal (Figure 2.12 [2]). Thus, Spinach labeling of mRNA is orthogonal to the essential reactions occurring in the PURE system and can be used in conjunction to fluorescent protein reporters for unbiased detection of the gene expression dynamics.

IMPROVED FOLDING TIME OF SPINACH APTAMER

We suspected that changing the nucleotide sequences upstream and downstream the Spinach could influence the folding time and thus the time delay between messenger production and Spinach fluorescence detection. This time delay can be decomposed into three consecutive events: the aptamer folding, the binding of DFHBI and the fluorescence emission, the latter occurring at a shorter time scale than the earlier two. To

investigate the lag time, DNase can be added to a running PURE*frex* reaction to immediately stop further mRNA production and the residual increase of the LL-Spinach fluorescence is monitored, similar to the experiment in Figure 2.4. It is important to inject the DNase while the Spinach intensity is linearly increasing and has reached sufficiently high signal-to-noise ratio for accurate measurement. The characteristic time of the increase in the remaining signal corresponds to the time delay defined above. We found that upon DNase injection the signal levels off instantaneously given our 30-s temporal resolution, resulting to a time delay estimate < 1 min. This value is significantly lower than the 2.6 min delay found using the YFP-Spinach construct [2]. This result suggests that Spinach folding is the rate-limiting step for fluorescence detection, not the aptamer-DFHBI binding, which was anticipated since DFHBI is present in large excess.

QUANTIFYING THE LEVELS OF MRNA

The quantitation of gene expression requires to convert arbitrary fluorescence intensity values into absolute mRNA and protein concentrations. In our previous study, we had quantified the mRNA concentration by gel analysis, which is complemented in this study by real-time quantitative PCR (RT-qPCR). The work-flow starts with a transcription only or coupled transcription-translation experiments conducted in the absence (ΔR) or in the presence (+R) of ribosomes, respectively (Figure 3.3F). After about 100 min mRNA production is stopped by adding DNase while the fluorescence signal is continuously monitored. The end-point fluorescence intensity is determined as the mean value of the Spinach signal during the last 10 min. Then, the reaction solution is harvested for further quantification using RNA gel and real-time quantitative PCR (RT-qPCR) analysis.

GEL ANALYSIS OF MRNA CONCENTRATION

Two to six microliters of purified RNA samples were loaded on a 1.2% agarose gel containing EtBr and a voltage of 90 V was applied for 1.5 h. The gel was then imaged and the band intensities were analyzed using the ImageLab software (Figure 3.4A). A calibration curve was generated by plotting the measured band intensity values of purified RNA (reference RNA) against their predetermined amounts. The standard curve is then used to calculate the concentration of mRNA produced in the PURE system (Figure 3.4B). Comparing the mRNA band intensity in +R and ΔR expression conditions, we noticed that a lower amount is detected in the presence of ribosomal RNA (rRNA) (Figure 3.4A, B), suggesting an rRNA-dependent loss of messenger during purification (though we paid attention that the purification column was not overloaded with total RNA) or less efficient migration through the gel. Therefore, PURE*frex* ΔR produced mRNA samples only were used for quantification. A conversion factor corresponding to 10.0 ± 2.1 nM mRNA per fluorescence a.u. was obtained by plotting the mRNA band intensity onto the calibration curve and normalizing the end-point fluorescence intensity measured with the spectrofluorometer.

REAL-TIME QUANTITATIVE PCR ANALYSIS

Absolute quantification of RNA samples can also be performed by RT-qPCR. We used the Eco Real-Time PCR System from Illumina. Three microliters of PURE*frex* ΔR produced mRNA were harvested from a PURE system reaction treated with DNaseI, diluted 100-fold in RNase-free water and stored at -80°C until used. No purification is needed. The

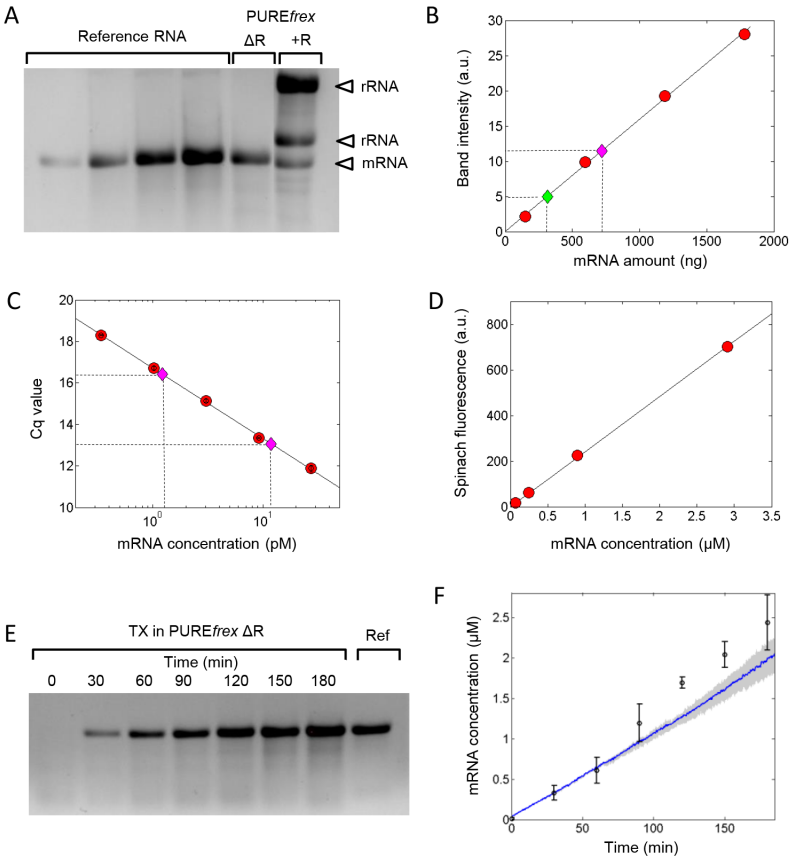


Figure 3.4: Quantification of mRNA synthesis in PURE system reactions starting from 7.4 nM of the mYFP-LL-Spinach DNA template. (A) RNA samples loaded on an agarose gel. The band intensities of mRNA produced in a PURE*frex* reaction (Fig. 4F) with (+R) or without ribosome (ΔR) were compared to reference RNA samples of known concentrations. The 1.5- and 2.9-kb ribosomal RNA bands are visible in the +R reaction condition. (B) Calibration curve plotted as the measured band intensities of reference RNA samples versus their predetermined concentrations (circles). The mRNA band intensities of the PURE*frex* samples shown in (A) were appended on the calibration curve (diamonds), after which the amount of synthesized transcript can be determined. (C) Reference RNA samples were analyzed by RT-qPCR and their C_q values plotted as a function of concentration. The obtained standard curve has a typical equation of y = -1.479ln(x) + 16.723; R^2 = 0.998. The measured C_q values of diluted samples from PURE*frex* ΔR reactions were appended on the calibration curve and their concentrations were determined. Two samples of 10-fold different dilution factors are displayed. (D) Calibration curve consisting of Spinach fluorescence intensities measured for different concentrations of reference RNA solutions. The slope gives a conversion factor of 4.1 nM / a.u. (E) Time series gel analysis of mRNA produced in a PURE*frex* ΔR reaction. The band intensities were compared with that of a reference RNA (right-most lane). (F) The dynamics of transcription was reconstructed by plotted the concentrations of mRNA as determined in (E) at different time points. Error bars indicate SEM, n = 3. For comparison the apparent kinetics obtained by monitoring the Spinach fluorescence in real-time is overlaid. The blue curve is the mean of three independent measurements and the gray shaded area denotes the min and max deviation.

samples were further diluted and 1 μL was added in 10 μL of RT-qPCR reactions corresponding to a final dilution factor of 10^5 or 10^6. The Power SYBR Green RNA-to-CT 1-Step kit (Applied Biosystems) was used according to the supplier's recommended protocol. Primers were designed to amplify a 267-bp long region of the mYFP gene. The forward (5'-CACCTACGGCAAGCTGACC-3') and reverse (5'-TTCAGCTCGATGCGGTTC-3') primers (Biolego) were used at 100 nM each. Prior mixing with the Power SYBR Green RNA-to-CT 1-Step kit, the reverse primer was incubated with the target RNA for 5 min at 65 °C and left for 10 min at room temperature. Reaction samples of 10 μL were loaded on a microplate (Eco Sample Dock, Illumina) and spun down for about 10 sec at 3,000-4,000 rpm (Eppendorf 5810R centrifuge). Each sample was analyzed in triplicate. Concentrations of mRNA were determined using a standard curve generated by serial dilution in autoclaved milliQ water of reference RNA at final concentrations ranging from 0.33 pM to 27.5 pM (five points, each in triplicate) (Figure 3.4C). A total number of seven independent samples (gene expression reactions performed separately) were analyzed leading to a conversion factor of 9.0 ± 5.2 nM (mean ± standard deviation) of mRNA per fluorescence a.u., in close agreement with the value extracted from gel analysis.

MRNA QUANTIFICATION FROM SPINACH FLUORESCENCE OF REFERENCE RNA

Reference RNA samples were diluted at different factors in the PURE system buffer (50 mM HEPES, 100 mM potassium glutamate, 13 mM magnesium acetate, pH 7.6), heated for 5 min at 65 °C and the solutions were allowed to cool down to room temperature in the presence of DFHBI. The Spinach fluorescence intensity of each sample was measured at 37°C with the spectrofluorometer and its value plotted as a function of the corresponding RNA concentration (Figure 3.4D). The obtained calibration curve served to determine the concentration of mRNA synthesized in the PURE*frex*, for which the endpoint Spinach fluorescence signal was measured (Figure 3.3F). A value of 4.1 nM mRNA per fluorescence a.u. was found, which is markedly lower than the conversion coefficient obtained with the gel and RT-qPCR methods. Though the exact reason of this discrepancy remains to be clarified, it is likely that the folding propensity of Spinach, and thus the fluorescence intensity, differs whether the mRNA is gradually produced *in situ*, i.e. in the PURE*frex* reaction, or is presynthesized and subsequently exposed to DFHBI. Therefore, we recommend to opt for the RNA gel and RT-qPCR as more reliable methods since the Spinach signal is also measured in the PURE*frex*.

QUANTIFYING THE LEVELS OF PROTEIN

We have quantified the mYFP concentration by fluorescence correlation spectroscopy and absorbance measurements (Figure 3.5A and B), essentially following the procedures as described in Chapter 2 [2]. In this study, we decided to use the conversion factor calculated from FCS measurements, 0.33 nM / a.u., for further quantitative analysis of mYFP kinetics. This choice is motivated by two facts: First, in contrast to our previous work [2], the mYFP contains the mutation that makes the protein stable in the monomeric form, which is more suitable for FCS quantification. Second, FCS measurements can directly be performed in diluted PURE*frex*, whereas absorbance measurements require the substitution of the expression system for the PURExpress. The higher FCS conversion factor means that we actually detect three times more fluorescent YFP molecules

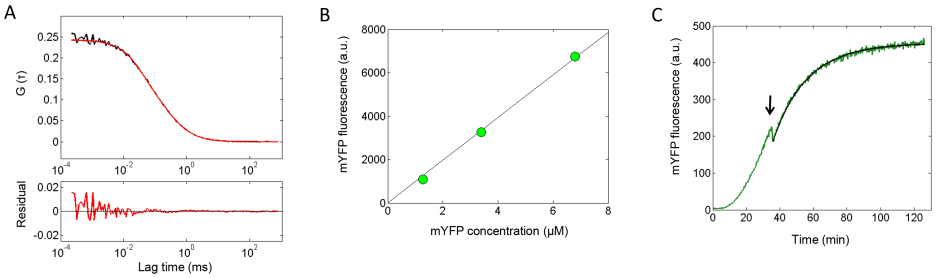


Figure 3.5: Quantification of YFP concentration and maturation time. (A) Fluorescence autocorrelation curve of a 50-fold diluted solution of mYFP synthesized in the PUREfrex and analyzed by FCS. The fit to Eq. 1 (red line) and the fit residual (bottom graph) are shown. (B) Quantification of PURExpress-synthesized mYFP concentration by absorbance measurements. The mYFP fluorescence intensity measured on the spectrofluorometer was plotted against the concentration determined by absorbance. The slope gives a conversion factor of 1.01 nM / a.u. (C) The maturation time of mYFP was experimentally measured by adding 3 μL of the translation inhibitor chloramphenicol (arrow). The residual increase in fluorescence was fitted with a monoexponential function (black curve), giving a maturation time of 20 min.

with this method compared to the absorbance measurements, suggesting that the latter method is less sensitive. Though we have not tried to apply FCS to quantify Spinach concentration, the technique is potentially of interest. However, the additional binding dynamics of DFHBI within the detection volume and the yet unknown triplet lifetime of spinach will complicate the interpretation of the autocorrelation curve.

3.2.1. QUANTITATIVE ANALYSIS OF MRNA AND PROTEIN CONCENTRATION VERSUS TIME

Having determined the conversion factors between fluorescence intensity values and absolute concentrations of synthesized mRNA and YFP, one can quantitatively analyze the transcription and translation reactions. A complete quantitative understanding of the gene expression dynamics would require further mathematical modeling and computer simulations, but these studies are not within the scope of the work presented here. The sensitivity of the two-reporter system can be assessed by calculating the lowest concentration of LL-Spinach and mYFP measured by spectrofluorometry. The lowest detected concentration of mRNA was calculated as two times the standard deviation of the LL-Spinach fluorescence signal collected within the first 20 min of expression after applying a baseline correction to eliminate the contribution of the rising average component of the signal. A value of 10 nM mRNA was obtained. The lowest detected concentration of mYFP was estimated as 2.3 nM using the same approach, except that a 20-min window taken in the linearly rising phase of the kinetics was used to compute the standard deviation. To verify that the apparent kinetics reported by Spinach fluorescence reflects the actual profile of mRNA synthesis, it is essential to quantify on RNA gel the amount of transcript produced at different time points in PURE*frex* ΔR reactions. The mYFP-LL-Spinach DNA template was used. Two microliter samples were collected at 30-min intervals, diluted 50-fold in RNase-free water, purified and gel analyzed according to the protocol described in sections 3.2 and 3.3. The fluorescence and gel-based

kinetics are nearly identical over the full time window covered (Figure 3.4F), thus validating the use of the LL-Spinach tag as a reliable reporter for mRNA synthesis dynamics. Given that the fluorescence emission signal of mYFP is not instantaneous upon synthesis owing to folding and chromophore maturation, experiments to determine the time delay between mYFP production and its fluorescence detection were performed. We supplemented the PURE*frex* reaction operating in the linear protein production regime with chloramphenicol (260 μg / ml final concentration), a translation inhibitor, and measured the residual increase of mYFP fluorescence (Figure 3.5C). Fitting the fluorescence signal to a first-order (mono-exponential) kinetic equation led to a time delay of 20 min, corresponding to a maturation time of the mYFP chromophore of 0.05 min^{-1} (Figure 3.5C). This value is similar to that previously determined in the PURExpress with a different method [9]. Importantly, the actual kinetics of protein synthesis can now be reconstructed from the measured apparent kinetics accounting for 20 min time delay [2]. The final amount of produced YFP and mYFP is between 0.4 and 0.5 μM (Figure 3.3C and E), assuming all the synthesized proteins emit fluorescence. This concentration is about 20-fold lower than the theoretically achievable production of 10.7 μM calculated using an initial concentration of 0.3 mM for each amino acid [10]. The low yield of protein synthesis contrasts with the more efficient production of mRNA that reaches the maximum theoretical concentration of 3.3 μM after approximately 7 h reaction when starting with 7.4 nM of DNA template. In this calculation, the initial concentration of CTP and UTP (1 mM) was used, since they are not involved in translation-associated processes. This good agreement indicates that CTP and UTP depletion by reacting with the nucleoside diphosphate kinase for regenerating ATP and GTP is negligible. Together these results indicate that only a small fraction of transcript is translated, which should definitely be taken into account for further improvement of the PURE system efficiency.

3.2.2. DFHBI-1T INCREASES SPINACH FLUORESCENCE IN THE PURE SYSTEM

We have tested the utilization of the DFHBI-1T (DFHBI with a 1,1,1-trifluoroethyl substituent) molecule, by comparing the fluorescence of Spinach complexed with DFHBI, or DFHBI-1T when mYFP-LL-Spinach is expressed from a pooled PURE*frex* ΔR system. We find that the Spinach complex with the DFHBI-1T molecule is indeed more fluorescent than with the DFHBI molecule, by a factor 1.5 (data not shown). The optimal excitation and emission wavelengths measured in the PURE system for the DFHBI-1T are 470 nm and 505 nm, respectively (Figure 3.6). Even at the optimal wavelengths of the DFHBI molecule, 460 nm for the excitation and 502 nm for the emission, the DFHBI-1T fluorescence is measured to be 1.37 times higher. We note that the optimal excitation wavelength in the PURE system buffer differs slightly from the reported maximal excitation wavelength, 482 nm, of the DFHBI-1T [4]. However, when using the optimal excitation and emission wavelengths for the DFHBI-1T, the possible crosstalk with the YFP protein increases, which will give a contribution of 1.6% of the maximum fluorescence signal of the YFP measured at its optimal wavelengths (λ_{exc}=514 nm and λ_{em}=528 nm). One should take this cross-talk into account when quantifying RNA levels using DFHBI-1T in fluorescence kinetics measurements.

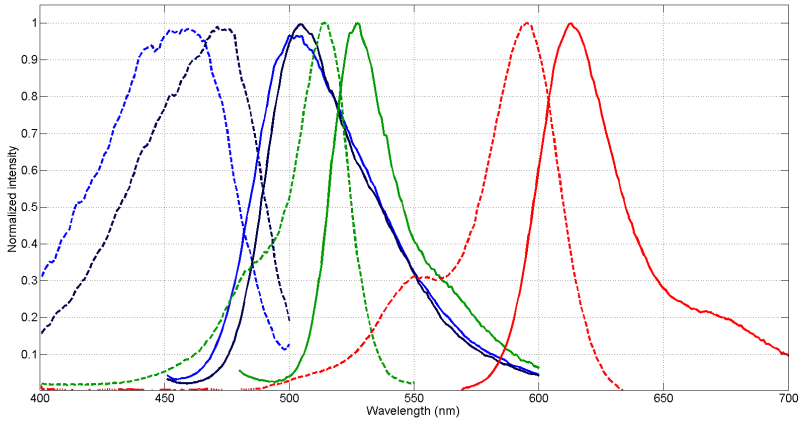


Figure 3.6: Excitation (dashed) and emission (solid) spectra of the Spinach in complex with DFHBI (light blue) or DFHBI-1T (black), the latter being slight red-shifted. The spectra of YFP (green) and Texas Red (red), the dye that colors the liposome membranes, are also displayed.

3.2.3. IMPROVING GENE EXPRESSION BY GENE-OPTIMIZING MYFP-LL-SPINACH DNA

The degeneracy in the genetic code, where one amino acids can be coded by different triplet sequences, varies from organism to organism. The origin of the green fluorescent protein, and its many engineered variants, is the jellyfish *Aequoria Victoria* [11]. The PURE system contains a tRNA mix from the *E.coli* organism[1] and its tRNA abundance and condon usage is thus not optimized for the expression of jellyfish genes. Nowadays, the construction of nucleic acid sequences is customized which allows a great flexibility in gene design for optimal protein expression [12]. We have therefore chosen to investigate if a gene-optimized version of the mYFP-LL-Spinach could improve gene expression in the PURE system. We have opted for the gene-optimization procedure and construction by the company Eurogentec. The sequence optimization is performed from the START till the STOP codon and involves improving the codon usage to the host expression system and avoiding secondary structures in the RNA, among others (not all processes are disclosed by the company). We term this gene the mYFPgo-LL-Spinach. Expression in the PURE system results in a final yield of YFP synthesis that is 1.5 times higher for the gene optimized mYFPgo-LL-Spinach compared to mYFP-LL-Spinach, and the Spinach fluorescence is increased around 2.5-fold (Figure 3.7). Interestingly, although the Spinach sequence did not change, the fluorescence of the Spinach complex improves when optimizing the sequence of only the YFP gene. This suggests that intramolecular RNA interactions affect the folding or stability of the Spinach aptamer.

[1] *E. coli* strain MRE 600

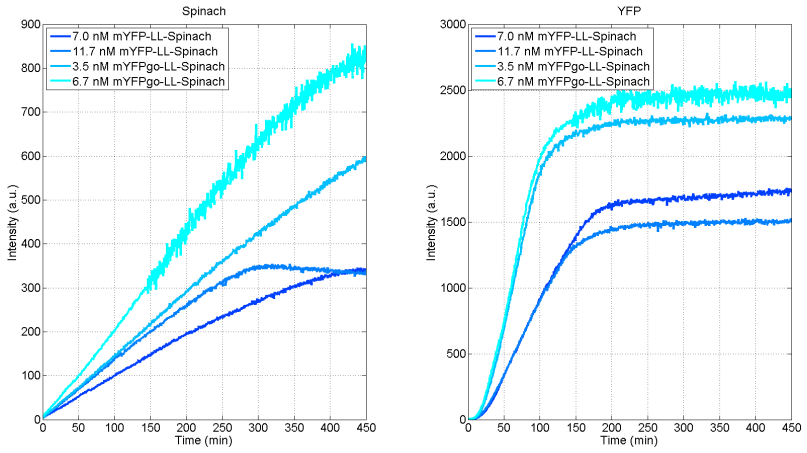


Figure 3.7: The gene optimized mYFPgo-LL-Spinach produces higher Spinach and YFP fluorescence when expressed in the PURE system compared to that with the mYFP-LL-Spinach construct (expressed at 37 °C).

3.3. QUANTIFYING GENE EXPRESSION INSIDE SEMIPERMEABLE LIPOSOMES

In this research, we report on gene expression kinetics of the optimized construct inside liposomes. We perform a quantitative analysis that allows us to monitor actual concentrations of produced RNA and proteins inside liposomes and compare these values with bulk measurements. New insights of these results will be discussed. The recently established protocol to trigger biosynthesis of proteins inside surface-tethered lipid vesicles as described in [2, 13, 14] was used to detect gene expression of the mYFP-LL-Spinach and its variants. In short, the protocol is based on gentle re-hydration of a lipid film (Figure 3.8). This method is compatible with a diversity of natural and synthetic lipids, and it is oil-free. The vesicle membrane can be equipped with a number of functionalities, such as biotin-PEG lipids (PEG = poly(ethylene glycol)) for liposome immobilization on neutravidin-coated surfaces and TRITC-conjugated lipids (TRITC = N-(6-tetramethylrhodaminethiocarbamoyl)) for membrane localization using fluorescence imaging. Moreover, the lipid composition can be tailored to tune the bilayer phase transition temperature or to regulate membrane permeability as a response to osmotic stress. The preparation of lipid-film coated beads, liposome formation and encapsulation of the PURE system, surface functionalization, liposome immobilization and triggering of gene expression we refer to the protocol in Chapter 2 and in [15]. The following lipid compositions (approximate mole %) are used for DMPC/DMPG vesicles: 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC, 80%), 1,2-dimyristoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DMPG, 20%), N-(6-tetramethylrhodaminethiocarbamoyl)-1,2-dihexadecanoyl-sn-glycero-3-phospho-ethanolamine (TRITC-DHPE, 0.5%) and 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[biotinyl(polyethylene glycol)-2000] (DSPE-PEG-biotin, 0.5%). Alternatively, a different lipid mixture was used for DOPC/DOPG vesi-

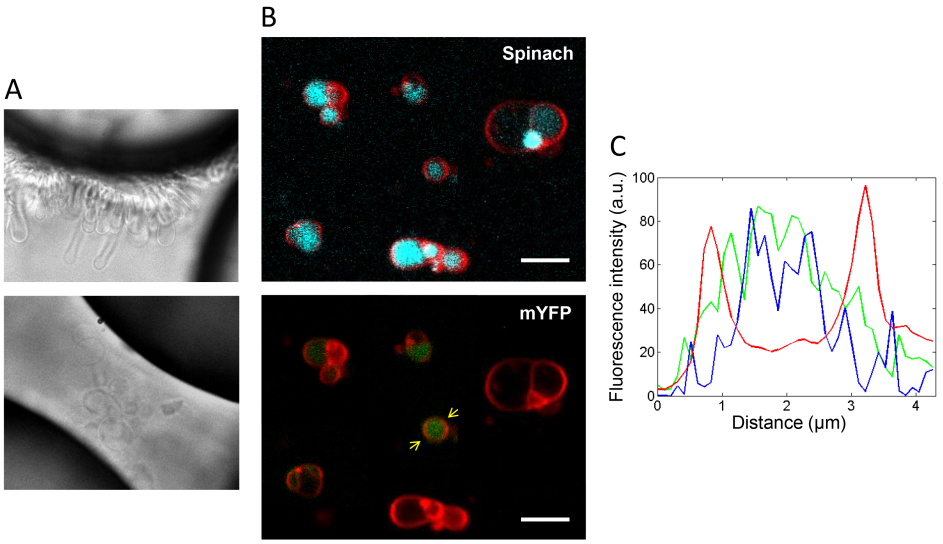


Figure 3.8: Fluorescence imaging of liposomes. (A) Phase contrast micrographs of lipid film swelling from glass bead surfaces (top). The tethered tubular liposomes eventually detach from the glass surface and remain trapped within the bead cavities (bottom) until the bead stack is gently disassembled for liposome harvesting. Adapted from [13]. (B) Fluorescence confocal images of surface-tethered liposomes (dioleoyl phospholipids, composition 2) post expression of the mYFP-LL-Spinach gene. The vesicle membrane (red) is localized using TRITC-labeled phospholipids. Scale bar is 5 μm. (C) Fluorescence intensity profiles of the TRITC, Spinach and mYFP signals measured along the line defined between the two arrows on (B). Color coding is the same as in (B).

cles containing 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC, 80%), 1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt) (DOPG, 20%), TRITC-DHPE (0.5%) and DSPE-PEG-biotin (0.5%). All lipids were purchased from Avanti Polar Lipids except for TRITC-DHPE, which was from Invitrogen.

Compared with the YFP-Spinach construct, imaging of synthesized mRNA inside liposomes was greatly enhanced using the LL-Spinach as shown by the high fluorescence intensity measured in transcriptionally active vesicles (Figure 3.8B, C). Similar improvement was observed with both dimyristoyl- or dioleoyl-containing lipid compositions. For the remaining liposome experiments described in this chapter, we have utilized an YFPgo-LL-Spinach template to investigate gene expression kinetics in a quantitative manner. This construct has the same sequence as the described mYFPgo-LL-Spinach, but it lacks the A206K mutation. Without this mutation the YFP has the tendency to form dimers instead of monomers, which we suspect will be beneficial to retain the synthesized protein inside the vesicles. The mutation does not change the spectral properties of the YFP.

3.3.1. GENE EXPRESSION KINETICS

Kinetics of gene expression of the YFPgo-LL-Spinach in the PURE*frex* system encapsulated inside liposomes were measured with time-lapse fluorescence confocal microscopy, as displayed in Figure 3.9. We are interested in the question: How does gene expression in liposomes differ from gene expression in bulk? The fundamental dissimilarities between the batch-mode and vesicle-confined reaction include:
1) the lipid vesicle is a semi-open system. Indeed, small molecules are expected to diffuse through the semi-permeable membrane. This includes uptake of feeding compounds and removal of waste products.
2) The encapsulation of components and the formation of liposomes are stochastic processes and therefore the initial conditions in each liposome are different.
3) Reaction components are exposed to surface to surface effects.
4) Confinement effects play a role in femtoliter reaction volumes.

CALIBRATION OF YFP AND SPINACH SIGNAL IN CONFOCAL MICROSCOPY MEASUREMENTS

To investigate this question quantitatively, we converted the arbitrary intensity values of the YFP and Spinach channels in the confocal microscopy experiments into concentrations. The fluorescence intensity of PURE system samples containing either Spinach or YFP fluorescence at different levels were first measured in the spectrofluorometer and subsequently with the fixed excitation and emission settings of the confocal microscopy set-up. The resulting calibration curve is displayed in Figure 3.10. Assuming the fluorescence intensity increases linearly with concentration over the whole range of the microscope (with exception of threshold and saturated values of intensity), the linear fit of the points in Figure 3.10 is used as an intermediate conversion. This results for the Spinach fluorescence in a first conversion formula:

$$I_{spectrometer} = 4.9\,(\pm 0.3)\cdot I_{microscope} - 1.6\cdot 10^{2}\left(\pm 0.4\cdot 10^{2}\right) \tag{3.1}$$

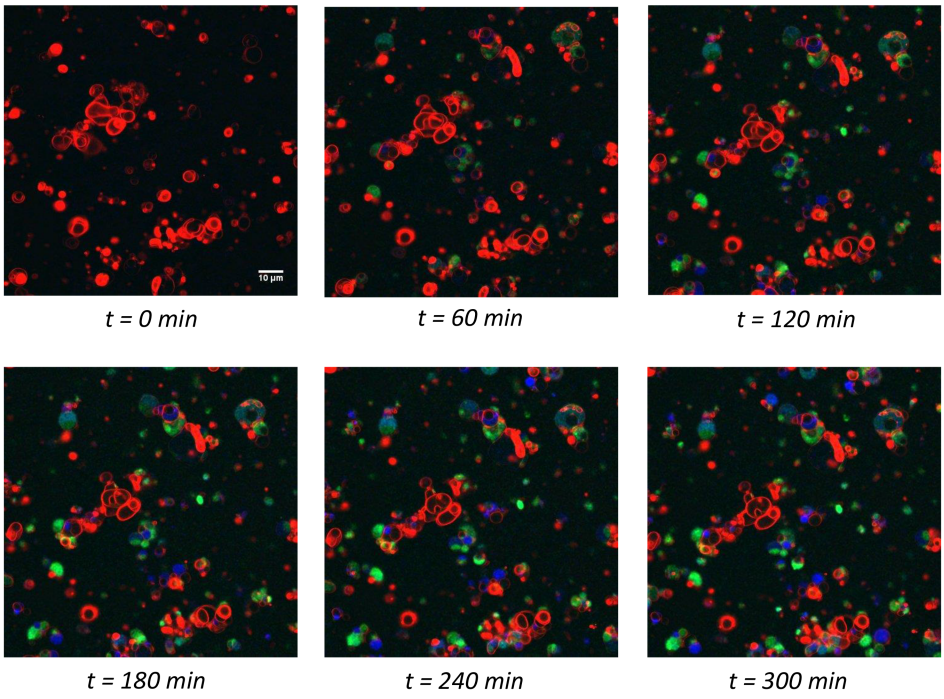


Figure 3.9: Liposomes containing PUREfrex expressing Spinach and YFP, imaged at six different time points. t = 0 is defined at the start of the image acquisition, 12 min after triggering gene expression. The liposome membrane is visible in red, the lumen shows the expression of Spinach (dark blue) and YFP (green). Raw data has been adjusted with ImageJ to enhance visibility. The image shown is only a small part of the total imaged field of view. The number of liposomes increases over time due to the fact that liposome immobilization is not perfect but gradual and vesicles still reach the glass surface during the measurement time.

and for the YFP we found:

$$I_{spectrometer} = 17.2\,(\pm 0.2)\cdot I_{microscope} - 3.3\cdot 10^2\left(\pm 0.8\cdot 10^2\right) \tag{3.2}$$

The spectrofluorometer units are converted to concentrations as described in the sections 3.2 and 3.2. The arbitrary Spinach fluorescence units are converted into concentrations by the same method as described in section 3.2. A conversion factor of 3.9 ± 0.9 nM mRNA per fluorescence a.u. was found for the (m)YFPgo-LL-Spinach construct. Although the maximum fluorescent signal of the Spinach complex is higher here than for the mYFP-LL-Spinach (see also Figure 3.7), the conversion factor is decreased and the theoretical maximum concentration of 3.3 μM RNA is indeed not exceeded in Figure 3.7. Since the spectral properties of YFP and mYFP are identical, the 0.33 nM / a.u. is applicable (as found in section 3.2). Combining these conversion factors with Equation 3.1 and Equation 3.2 results in the conversion formulas from arbitrary fluorescence units on the microscope to mRNA and YFP concentrations:

$$[mRNA] = 19.1\,(\pm 5)\cdot I_{microscope} - 6.2\cdot 10^2\left(\pm 2\cdot 10^2\right) \tag{3.3}$$

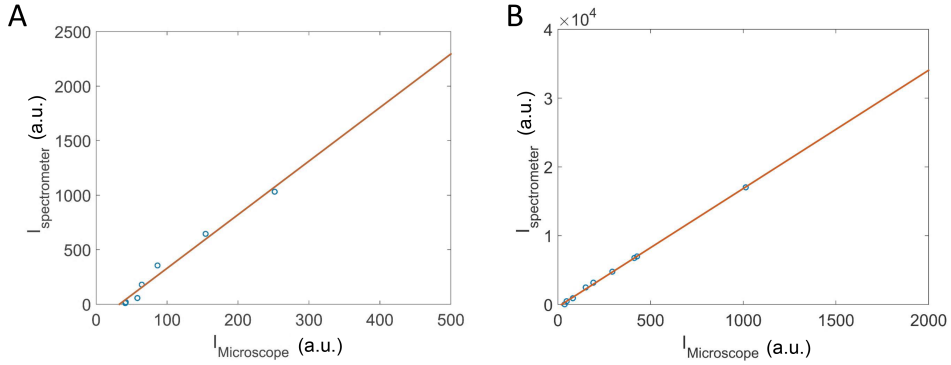


Figure 3.10: Calibration curves of the Spinach (A) and YFP (B) signals of the microscope versus the spectrofluorometer. The linear fits are shown by Equation 3.1 and Equation 3.2 respectively . This calibration is performed at the lower intensity ends of the range of the microscopy, and we have used this to cover the whole range.

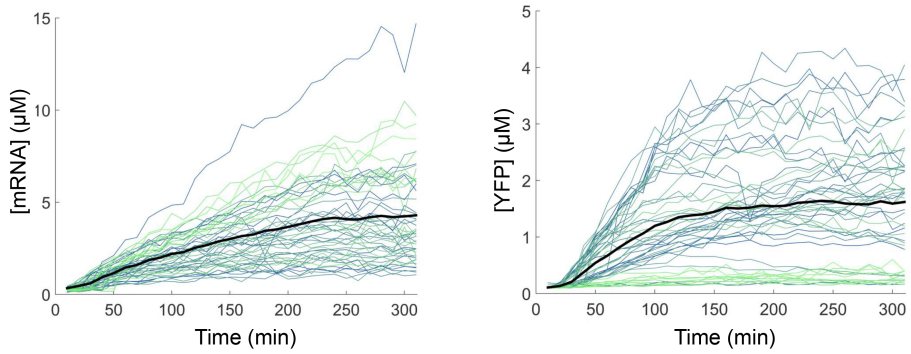


Figure 3.11: Apparent kinetics of gene expression of the YFPgo-LL-Spinach template in the PURE system encapsulated in liposomes (N=50). Each line corresponds to one liposome. The superimposed black lines respresent the mean of the curves, and is typical to bulk concentrations.

$$[YFP] = 5.7\,(\pm 0.7)\cdot I_{microscope} - 1.1\cdot 10^2\left(\pm 0.3\cdot 10^2\right) \tag{3.4}$$

Applying these conversion formulas to the gene expression kinetics of 50 individual liposomes of Figure 3.9, results in Figure 3.11.The production of mRNA slows down after approximately four hours, while the YFP concentration reaches a plateau after 2.5 h. Important to note is that the start of mRNA and protein production is synchronous for all liposomes, an indication that gene expression is triggered at the same moment by this experimental approach. Both mRNA and protein expression kinetics have a broad spread in the concentration curves, indicating a high vesicle-to-vesicle variability. In Figure 3.12A the amount of protein produced per mRNA (the ratio of apparent YFP concentration and Spinach concentration) is plotted for the analyzed vesicles. An upper bound of approximately two proteins per mRNA is found. The broad distribution of this ratio indicates the existence of vesicles with low mRNA and high protein concentrations,

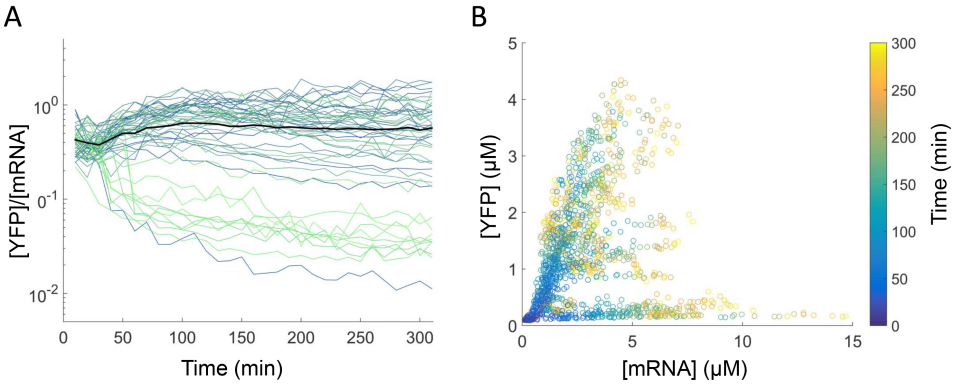


Figure 3.12: The graphical representation of the relation between mRNA and YFP concentration in individual liposomes shows a high liposome-to-liposome variability. A) YFP concentration divided by the mRNA concentration from the liposomes analyzed in Figure 3.11. B) scatter plot of YFP versus mRNA concentration, where each point corresponds to a specific time-point in an individual liposome. Color changes from blue to yellow with increasing time.

and vice versa. When plotting the mRNA concentration against the protein concentration for each time point (Figure 3.12B), these two categories of expressing vesicles are clearly visible.

A characteristic feature of protein expression is the time window in which expression takes place before it ceases and reaches a plateau. Among the different definitions one can use for determining and comparing this characteristic time, we chose to fit the data with a sigmoidal fit of the response curve typically used in synthetic biology [16]:

$$C(t) = k' + k\frac{t^n}{t^n + K^n} \tag{3.5}$$

where n is a measurement of the steepness, k' a measurement of the start height, k the predicted final protein yield and K a measurement of delay for the tipping point in the curve. Given that the time lag for YFP signal due to chromophore maturation is fixed (ca. 20 min), an indication for the translation lifespan (duration of gene expression) is a relation of the delay time K and the steepness of the curve (determined by n). We defined this duration as the time up to the moment the tangent of the curve in point K crosses the value of the plateau yield, leading to $T_{plateau} = 2 \cdot \frac{K}{n} + K$ (see Section 3.5 for detailed explanation). In Figure 3.13A, a histogram of the plateau times $T_{plateau}$ for the kinetic YFP curves from Figure3.11 was plotted. As a comparison, the $T_{plateau}$ delay time of the bulk kinetic curves in Figure 3.7 is 130±28 min. Remarkable is the broad spread in delay times, indicating that individual vesicles have different expression behaviours. The most extreme case are five vesicles that do not reach a plateau during the 5-h measurement; their kinetic curves are highlighted in Figure 3.13B. Close-ups of these vesicles are shown in Figure 3.14. Visual inspection of these vesciles do not reveal any particular feature that could explain long-term expression; most vesicles in Figure 3.9 are also fused to other membranes.

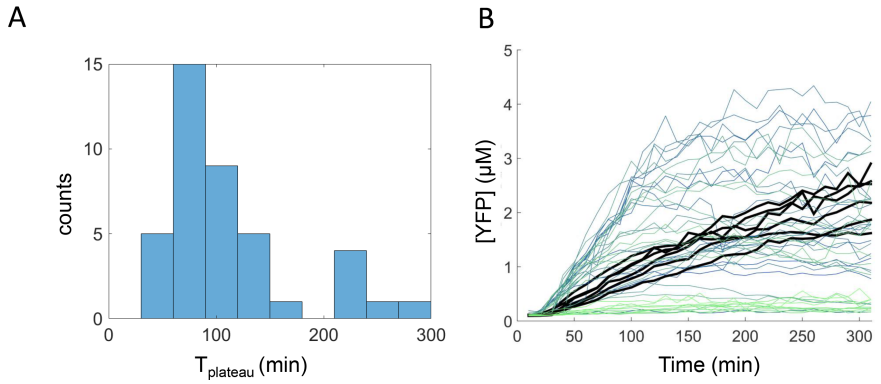


Figure 3.13: Expression time of the YFP varies between liposomes. A) Histogram (counts) of the plateau times $T_{plateau}$ of the YFP kinetic curves shown in Figure 3.11. Plateau times are determined as in Section 3.5. Plateau times higher than 300 min are left out of this histogram (see Figure 3.16 as they do not represent an actual experimentally measured plateau time. B) Some of the kinetic curves that do not reach a plateau during the time of measurement are highlighted in black.

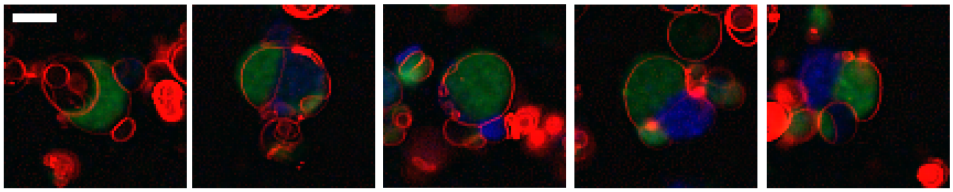


Figure 3.14: Close-ups of five liposomes that did not reach a plateau of YFP expression after five hours. The corresponding kinetics were highlighted in Figure 3.13B. The liposomes of interest are in the center of the images showing YFP expression (green). The scale bar equals 5 μm.

3.4. DISCUSSION

We reported in this chapter on the enhanced Spinach fluorescence owing to improved folding of the LL-Spinach sequence. The LL-Spinach fluorescence signal enables reliable and real-time monitoring of mRNA production in gene expression reactions. Our studies on gene expression kinetics of the YFPgo-LL-Spinach in the PURE system encapsulated inside lipid vesicles led to the following interesting observations:

- There is a significant vesicle-to-vesicle variability in gene expression.
- Average kinetic curves are similar to bulk expression.
- Some vesicles produce significantly higher amounts of protein compared to bulk reactions.

3.4.1. POSSIBLE ORIGINS OF VESICLE-TO-VESICLE VARIABILITY IN GENE EXPRESSION

Gene expressing liposomes exhibit large signal disparities for both LL-Spinach and YFP reporters, reflecting intrinsic heterogeneity in the levels of synthesized mRNA and protein. We showed previously that the vesicle-to-vesicle variability in YFP fluorescence intensity is correlated neither with the amount of encapsulated DNA templates [14] nor with the mRNA level [2]. This feature is not specific to the natural-swelling method of forming liposomes, as significant variability in protein expression between individual liposomes formed by a water-in-oil emulsion method is likewise reported in literature [17]. In this research, giant liposomes containing the PURE system were formed by a spontaneous transfer of water-in-oil microdroplets through an oil/water interface and fluorescent protein kinetics were monitored. The results showed that variability emerges at both transcription and translation steps. Investigation of gene expression noise of two proteins expressed by the PURE system in liposomes formed by the water-in-oil emulsion transfer method, led to the conclusion that these artificial microreactors contain less correlated noise but similar levels of uncorrelated noise as bacterial cells [18]. The authors found that the in the range of 1-50 fl the mean concentrations of produced proteins were similar and the main source for correlated noise came from the fluctuation of encapsulated DNA molecules as every other component of the PURE system is present at more than 100 copy numbers in a 10-fl vesicle (assuming that encapsulation is a Poisson process). Interestingly, based on their results they proposed that only part of the mRNA was active.

The main difference between the two methods to form liposomes is that the water-in-oil emulsion transfer method allows direct encapsulation of the reaction solution, whereas the protocol described here relies on spontaneous formation of vesicles. The encapsulation of components and the forces driving the vesicle formation might therefore differ. Spontaneous crowding of ribosomes and proteins inside vesicles prepared by the film hydration method was observed in literature [19]. When a ribosome solution was encapsulated inside vesicles with an average size of 100 nm, the ribosome occupancy distribution did not follow a Poisson statistics but suggested a power law partitioning. No strong interaction between the lipid membrane and the ribosomes was reported. However, the authors hypothesized that such a macromolecular crowding is a result of the interplay between the vesicle formation mechanism and weak solute-solute and solute-

membrane interactions. A possible excluded volume effect was also invoked as entropic driving force. However, these entrapment events might occur differently with our approach to form liposomes. Here, the enzymes and DNA are encapsulated during the natural swelling process, whereas the nutrients, buffer components, and tRNAs are supplied from the outside solution. Although the precise mechanism of molecular diffusion across the lipid bilayer remains to be explored, it is very likely that the osmolarity mismatch between the inside and outside liposome solutions generated when supplementing the feeding mixture leads to transient defects in the liposome membrane. Remarkably, these defects enable the uptake of tRNAs and nutrients, whereas the different PURE system constituents, engaged into functional macromolecular complexes, remain trapped inside the vesicles. We attribute this heterogeneity (or stochasticity) primarily to the low-copy number of some constituents of the biosynthesis machinery confined within (sub-)micrometer sized liposomes, which leads to a large compositional diversity of vesicles and, thus, to a great disparity in the transcription and translation rates between liposomes. As the resources have to translocate from the environment, another factor influencing the yield of internal production is the lamellarity of the liposome membrane, with lower exchange efficiency for vesicles with more bilayers.

When the 16 carbon-acyl chain saturated phospholipid DPPC (bilayer in the liquid-ordered phase at 37 °C) are used, liposomes fail to express fluorescent proteins [13] indicating that the physico-chemical properties of the lipid bilayer, in particular its mechanical response to osmotic pressure, govern the molecular exchange with the environment and thus the allocation of resources. Importantly, the efflux of toxic reactional products, such as pyrophosphate, across the vesicle membrane could contribute to the prolonged expression duration observed in some vesicles compared to batch mode reactions.

3.4.2. SUGGESTIONS FOR IMPROVEMENTS

We foresee two research directions to achieve long-lived expressing vesicles:
1. Enhance the molecular exchange between the liposome reactor and the feeding environment. This could be accomplished by incorporating pore-forming peptides or proteins in the vesicle membrane [20].
2. Improving performance of the PURE system by restoring transcription-translation coupling [21], or by increasing the amount of transcriptionally active mRNA, e.g. by enhancing translation initiation. Another improvement could come from limiting accumulation of inhibiting reaction products, by adding peroxidases or e.g. by preventing the stalling of ribosomes with addition of EF-P [22], or by retrieving tRNA by peptidyl hydrolase when stalling of the ribosome results in a tRNA-peptidyl chain [23]).

3.5. MATERIALS AND METHODS

For the experiments described in the Figures up to Figure 3.8 the materials and methods are the same as described in Chapter 2 and [2, 15], unless indicated otherwise. New image acquisition and analysis were performed for the liposome experiments presented in Figure 3.9 ans later, which will be described here.

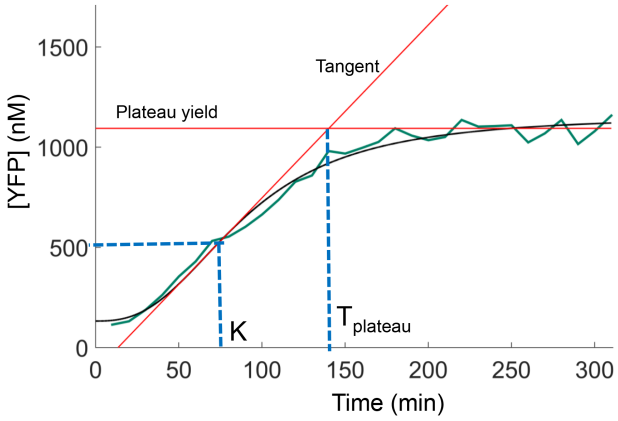


Figure 3.15: Schematic representation of the determination of the plateau time. The green curves shows an arbitrary data set. The black curve is the fit according to Equation 3.5. The x-coordinate of the intersection of the two red lines (tangent and plateau yield) is the plateau time.

IMAGE ACQUISITION AND ANALYSIS

For the fluorescence confocal microscopy experiments, an A1R Nikon Confocal Microscope equipped with an SR Apo TIRF 100x oil immersion objective is utilized. To image Texas Red, Spinach and YFP fluorescence, lasers with wavelengths of respectively 561 nm, 457 nm, and 514 nm are used. Since no dichroic mirror compatible with these three wavelengths was available, the Texas Red is first measured individually, after which the YFP and Spinach signal are measured. Texas Red is detected with a 595/50 band-pass emission, Spinach is detected with a 482/35 band-pass emission and YFP is detected with a 540/30 band-pass emission. Fixed laser settings have been used for each experiment, as reported in [24]. The temperature of the sample is regulated by a controller and is set to 37 °C. Images have been acquired every ten minutes for five hours. Manual image analysis using NIS Elements software is performed [25]. Liposomes containing a lumen surrounded by a border membrane that shows expression at the last time frame in at least one of the channels were selected for analysis. A region of interest (ROI) was defined, and its fluorescence signal was tracked backwards in time to confirm its presence in all the frames. The ROI, if necessary, was redefined when shape or location were slightly changed in between the time frames. The data obtained in this way by applying the *Edit ROIs in Time* and the *Measure* function were further processed in MATLAB.

DETERMINING PLATEAU TIME

The plateau time is calculated from the fit parameters of Equation 3.5. First, the derivative at the tipping point $t = K$ and its corresponding tangent are determined. The point where the tangent is equal to the plateau yield of the fit, $k' + k$, is the plateau time. This is visualized in Figure 3.15, and step by step calculations are given below. The derivative of Equation 3.5 is:

$$\frac{dC(t)}{dt} = 0 + k\frac{dt^n}{dt} \cdot \frac{1}{t^n + K^n} + k \cdot t^n \cdot \frac{d}{dt}(t^n + K^n)^{-1} = \frac{knt^{n-1} \cdot K^n}{(t^n + K^n)^2} \tag{3.6}$$

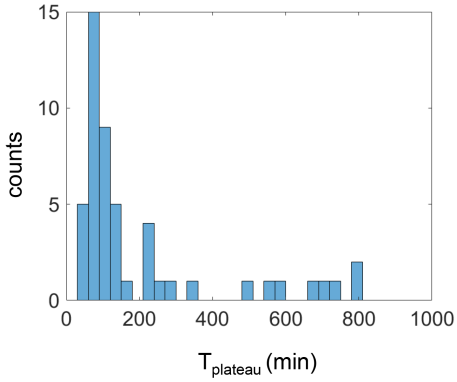


Figure 3.16: Histogram (counts) of the plateau times $T_{plateau}$ of all the YFP kinetic curves shown in Figure 3.11. Plateau times are determined as in Section 3.5.

The value at the tipping point $t = K$ is:

$$C(K) = k' + k\frac{K^n}{K^n + K^n} = k' + \frac{k}{2} \tag{3.7}$$

and the slope in the tipping point is given by:

$$\frac{dC(t=K)}{dt} = \frac{knK^{n-1} \cdot K^n}{(K^n + K^n)^2} = kn \cdot \frac{K^{2n-1}}{(2K^n)^2} = \frac{kn}{4K} \tag{3.8}$$

The tangent has the general formula of $y = ax + b$, where in our case $x = K$, a is the slope in K found in Equation 3.8 and y is the value $C(k)$ found in 3.7. The value for b is determined to find the tangent:

$$y = k' + \frac{k}{2} = ax + b = \frac{kn}{4K} \cdot K + b = \frac{kn}{4} + b \rightarrow b = k' + \frac{k}{2} - \frac{kn}{4} \tag{3.9}$$

The formula for the tangent in point K is thus:

$$y = \frac{kn}{4K}x + k' + \frac{k}{2} - \frac{kn}{4} \tag{3.10}$$

To find the plateau time $x = T_{plateau}$, the tangent is set equal to the maximal yield $y = k' + k$ in Equation 3.10:

$$y = \frac{kn}{4K}x + k' + \frac{k}{2} - \frac{kn}{4} = k' + k \rightarrow x = 2 \cdot \frac{K}{n} + K = T_{plateau} \tag{3.11}$$

Plateau times of YFP expression in liposomes are calculated for all YFP curves in Figure 3.16.

3.6. ACKNOWLEDGEMENTS

We thank Jan van Esch, Rienk Eelkema and Jos Poolman from the department of Chemical Engineering at the Delft University of Technology for synthesizing DFHBI. Thanks to Christophe who wrote most of the invited paper for Methods in Enzymology, this paper is substantially described in this chapter and included the Figures 3.1 till 3.5 and Figure 3.8 [15]. I would like to thank Zohreh for setting up the liposome protocols for the lab stimulated by Christophe, and both for providing highly relevant and interesting information from previous research that helps interpreting current and future liposome experiments. I am grateful to Alicia who performed the RT-qPCR measurements, to Roeland for preliminary characterization of the mYFP-LL-Spinach construct, to Ilja for assistance in designing and preparing the constructs and to Sabine for performing the chloramphenicol experiments. I would like to thank Duco for enthusiastically picking up how to perform liposome experiments during his master thesis, resulting in the work represented in the Figures 3.9 till 3.14, and for continuing these investigations successfully during his PhD. Thanks to Emma and Katy for preliminary characterization of kinetic curves and to Anne for gaining highly interesting insights through her modeling and experimental work on the PURE system.

REFERENCES

[1] J. S. Paige, K. Y. Wu, and S. R. Jaffrey, *RNA mimics of green fluorescent protein.* Science (New York, N.Y.) **333**, 642 (2011).

[2] P. van Nies, Z. Nourian, M. Kok, R. van Wijk, J. Moeskops, I. Westerlaken, J. M. Poolman, R. Eelkema, J. H. van Esch, Y. Kuruma, T. Ueda, and C. Danelon, *Unbiased tracking of the progression of mRNA and protein synthesis in bulk and in liposome-confined reactions.* Chembiochem : a European journal of chemical biology **14**, 1963 (2013).

[3] R. L. Strack, M. D. Disney, and S. R. Jaffrey, *A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA.* Nature methods **10**, 1219 (2013).

[4] W. Song, R. L. Strack, N. Svensen, and S. R. Jaffrey, *Plug-and-Play Fluorophores Extend the Spectral Properties of Spinach.* Journal of the American Chemical Society **136**, 1198 (2014).

[5] J. S. Paige, T. Nguyen-Duc, W. Song, and S. R. Jaffrey, *Fluorescence Imaging of Cellular Metabolites with RNA,* Science **335**, 1194 (2012), arXiv:NIHMS150003 .

[6] R. L. Strack, W. Song, and S. R. Jaffrey, *Using Spinach-based sensors for fluorescence imaging of intracellular metabolites and proteins in living bacteria.* Nature protocols **9**, 146 (2014), arXiv:NIHMS150003 .

[7] M. Zuker, *Mfold web server for nucleic acid folding and hybridization prediction,* Nucleic Acids Research **31**, 3406 (2003).

[8] F. Chizzolini, M. Forlin, D. Cecchi, and S. S. Mansy, *Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength.* ACS synthetic biology (2013), 10.1021/sb4000977.

[9] R. Iizuka, M. Yamagishi-Shirasaki, and T. Funatsu, *Kinetic study of de novo chromophore maturation of fluorescent proteins,* Analytical Biochemistry **414**, 173 (2011).

[10] Y. Shimizu, a. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda, *Cell-free translation reconstituted with purified components.* Nature biotechnology **19**, 751 (2001).

[11] R. Heim and R. Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer,* Current Biology **6**, 178 (1996).

[12] B. K.-S. Chung and D.-Y. Lee, *Computational codon optimization of synthetic gene for protein expression.* BMC systems biology **6**, 134 (2012).

[13] Z. Nourian, W. Roelofsen, and C. Danelon, *Triggered Gene Expression in Fed-Vesicle Microreactors with a Multifunctional Membrane,* Angew. Chem. Int. Ed **51**, 3114 (2012).

[14] Z. Nourian and C. Danelon, *Linking genotype and phenotype in protein synthesizing liposomes with external supply of resources.* ACS synthetic biology **2**, 186 (2013).

[15] P. van Nies, A. S. Canton, Z. Nourian, and C. Danelon, *Monitoring mRNA and protein levels in bulk and in model vesicle-based artificial cells.* Methods in enzymology **550**, 187 (2015).

[16] J. Ang, E. Harris, B. Hussey, R. Krill, and D. R. Mcmillen, *Tuning Biological Response Curves for Synthetic Biology,* ACS synthetic biology **2**, 547 (2013).

[17] H. Saito, Y. Kato, M. Le Berre, A. Yamada, T. Inoue, K. Yosikawa, and D. Baigl, *Time-Resolved Tracking of a Minimum Gene Expression System Reconstituted in Giant Liposomes,* ChemBioChem **10**, 1640 (2009).

[18] K. Nishimura, S. Tsuru, H. Suzuki, and T. Yomo, *Stochasticity in Gene Expression in a Cell-Sized Compartment,* ACS Synthetic Biology **4**, 566 (2015).

[19] T. Pereira de Souza, F. Steiniger, P. Stano, A. Fahr, and P. L. Luisi, *Spontaneous crowding of ribosomes and proteins inside vesicles: A possible mechanism for the origin of cell metabolism,* ChemBioChem **12**, 2325 (2011).

[20] V. Noireaux and A. Libchaber, *A vesicle bioreactor as a step toward an artificial cell assembly.* Proceedings of the National Academy of Sciences of the United States of America **101**, 17669 (2004).

[21] M. B. Iskakova, W. Szaflarski, M. Dreyfus, J. Remme, and K. H. Nierhaus, *Troubleshooting coupled in vitro transcription-translation system derived from Escherichia coli cells: synthesis of high-yield fully active proteins.* Nucleic acids research **34**, e135 (2006).

[22] S. Ude, J. Lassak, A. L. Starosta, T. Kraxenberger, D. N. Wilson, and K. Jung, *Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches.* Science (New York, N.Y.) **339**, 82 (2013).

[23] S. Vivanco-Domínguez, J. Bueno-Martínez, G. León-Avila, N. Iwakura, A. Kaji, H. Kaji, and G. Guarneros, *Protein synthesis factors (RF1, RF2, RF3, RRF, and tmRNA) and peptidyl-tRNA hydrolase rescue stalled ribosomes at sense codons,* Journal of Molecular Biology **417**, 425 (2012).

[24] D. Blanken, *Measuring cell-free gene expression kinetics in individual liposomes,* Master Thesis TU Delft (2015).

[25] NIS, *Elements Advanced Research 4.3,* in *Nikon Instruments Europe B.V.* (NIKON, 2015).

4

DNA REPLICATION CANDIDATES FOR THE MINIMAL CELL

In this chapter an overview of several in vivo and in vitro developed DNA replication systems is given, that may serve as inspiration for a DNA replication system for the minimal cell and as introduction for the remaining chapters of this thesis.

The minimal cell displays the essential functions of a living cell (see Section 1.1) and has the minimal set of genes that are necessary and sufficient to sustain a functioning cell under ideal conditions, in the presence of unlimited amounts of all essential nutrients and in the absence of any adverse factors [1]. Several approaches to construct a minimal cell were discussed in Chapter 1, including the top-down and synthetic biology bottom-up approach that both are inspired by comparative genomic studies. The main message is that a minimal cell based on our knowledge of modern biology will comprise around 150-250 genes encoded in its DNA [1]. In this chapter, we give an overview of several DNA replication systems that could be potential candidates for the minimal cell framework.

4.1. DNA REPLICATION

DNA replication is at the core of information transfer, vital to the propagation of life. In this section, the properties of the DNA polymerase, the key enzyme involved in DNA replication, as well as a typical prokaryotic DNA replication machinery, are introduced. After obtaining an understanding of the general mechanism of DNA replication, we argue where to look for DNA replication candidates for the minimal cell.

4.1.1. DNA POLYMERASE

The DNA polymerase (DNAP) is the enzyme that synthesizes new DNA, by polymerizing incoming deoxynucleotides (dNTPs, consisting of a triphosphate linked to a 5-carbon sugar molecule and a purine or pyrimidine base) complementary to the template strand according to Watson-Crick base pairing (adenine bonds with thymine :A–T and guanine with cytosine: G–C). For a detailed description of how DNA polymerases work we refer to the excellent books of DNA replication [3], DNA polymerases [4] and Genome Duplication [5]. In this section, some of the main similarities and differences between various DNA polymerases are described. First of all, DNAPs need a DNA template, a primer:template junction (PTJ) and dNTPs to perform the polymerization reaction. In the polymerization reaction, a phosphodiester bond is formed when the 3'-OH group of one nucleotide reacts with the phosphoric acid of the incoming nucleotide, linking the 5' carbohydrate to the 3'-carbon of the previous nucleotide (Figure 4.1A). No DNAP can initiate DNA synthesis *de novo*, in which they differ from RNAPs. The dNTPs are incorporated as monophosphates to the 3'-OH group at the growing end of the primer (Figure 4.1B). Polymerization is unidirectional and chain growth is exclusively in the 5' to 3' direction, which is anti-parallel to the template strand. DNAPs can make mistakes in the basepairing when incorporating a new nucleotide and have so-called different fidelities[2]. Spontaneous correct basepairing occurs with a ratio of 10 to 1 corrects versus incorrect [4], whereas the positioning of incoming nucleotides on the template inside the enzyme can increase this fidelity already to only 1 mistake in 1000 000 nucleotides. Many DNAPs contain a 3'-5' exonuclease domain that performs a proofreading activity by removing a mismatched incorporated nucleotide, increasing the fidelity further by

[1] This estimate would imply a minimal genome size of around 100 kb, assuming a simpler cell would have gene sizes on average half the size in modern bacteria [2]

[2] The degree of exactness with which DNA is copied or reproduced [5]

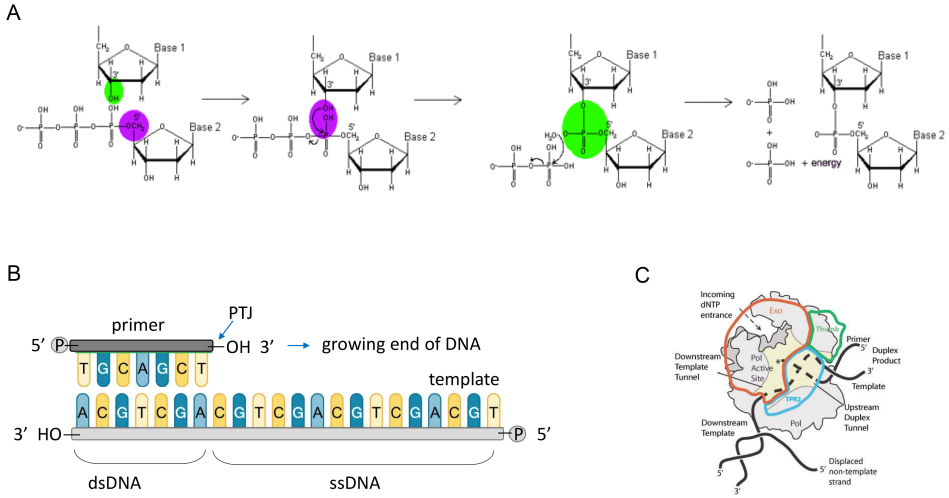


Figure 4.1: Schematic representations of the DNA polymerization process. A) The chemistry of the polymerization of two nucleotides is displayed, linking the 3'-carbon of one nucleotide with a phosphodiester bond to the 5'-carbon molecule of the incoming nucleotide. Pyrophosphate is released in this reaction. B) DNAPs need a primer:template junction (PTJ) to initiate the polymerization process. Chain growth is exclusively in the 5' to 3'-direction. C) Schematic surface representation of phi29 DNA polymerase with DNA substrate, adapted from [6]. In here, a narrow tunnel allows the uncopied downstream template into the active site while a large pore provides a path for incoming dNTP. Upstream product duplex exits from the polymerase active site through a tunnel of intermediate dimensions.

.

an order of magnitude. Different DNAPs vary in their synthesis rate from 1-1000 nt/s and are often accompanied by an accessory protein that increases the polymerization rate and the binding to the DNA. The processivity of a DNAP is defined as the number of dNTPs added for a primer template junction binding event before it falls off. Accessory proteins that increase the rate and processivity of a DNAP are therefore often called processivity factors. Apart from the 3'-5' exonuclease proofreading activity, some polymerases also contain a domain with a 5'-3' exonuclease activity. These polymerases can continue synthesis when they encounter a duplex DNA ahead of them, by cleaving off the nucleotides from the incoming strand. An example is the DNA Pol I in *E. coli*, that has the role to excise RNA primers with its 5'-3' exonuclease activity in DNA replication. Other polymerases are able to produce strand displacement, where duplex DNA is separated during the polymerization process. A good example is the viral phi29 DNA polymerase, of which a cross-section of the structure of the DNAP with the upstream and downstream template channels and NTP channel is shown in Figure 4.1C [6].

4.1.2. COMPARATIVE GENOMICS: DNA REPLICATION MECHANISMS IN PROKARYA

In order to replicate DNA, many accessory proteins are necessary besides the DNA polymerase. In this section, an overview is given of the main enzymes involved in DNA replication in *E. coli*, a relevant organism in synthetic biology and the basis for the PURE system [7]. The *E. coli* genome is a circular plasmid and has an asymmetric mode of replication performing leading and lagging strand synthesis at the replication fork[3]. The term leading strand means that the direction of the replication fork moves in the same direction, and continuous polymerization from 5'-3' by the DNAP takes place. The lagging strand has the opposite orientation and therefore can only be made in a discontinuous manner and covalently joined later. In total 228 proteins are involved in the faithful replication process and we will here discuss the most important processes. A minimal DNA auto-replicative system[4] has been proposed in literature that found the necessary genes by comparison of the genome of *E. coli* with those of 25 extremely reduced endosymbiotic bacteria [9]. A set of 17 genes is proposed, comparable to the number of proteins mentioned in the minimal gene set[5] [10].

DNA replication starts in two directions at a sequence that is recognized as an origin, *oriC*, enabling an initiatior protein, *dnaA*, to bind and unfold the initation site. A DNA helicase, *dnaB* unwinds the DNA strands during the elongation process. A DNA primase, *dnaG*, synthesizes the RNA primers that serve as an initiation site for the elongation by the DNA polymerase, once for the leading strand and multiple times on the lagging strand. In *E. coli*, the DNA polymerase is the Pol-III, with its main subunits alpha (*dnaE*), catalyzing DNA synthesis, and epsilon (*dnaQ*) that carries out proofreading. In the Pol-III holoenzyme two clamps assist the elongation process, the sliding clamp and the clamp loader. The sliding clamp is a DNA processivity factor[6] and consists of two

[3] for an elaborate overview of the processes at the replication fork, read Genome Duplication chapter 3[8]

[4] Defined by the authors as: a genetic system comprising the minimum number of DNA components, including regulatory elements and gene products necessary for the auto-replication of the DNA molecule on which they are encoded, functioning in an *in vitro* condition.

[5] Based on comparative genomics of just two bacteria: the parasite *M. genitalium* and the *H. Influenza*.

[6] The processivity of DNA Pol III alone is around 100 bp, whereas the processivity of the holoenzyme complex

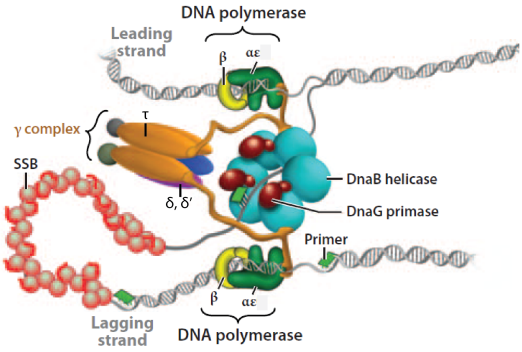


Figure 4.2: Schematic representation adapted from [11] to show the proteins necessary for the replication fork in a minimal DNA auto-replicative system as proposed in [9]. The oriC, DnaA, gyrase, ligase, and Tus-ter system, not depicted here, complete the minimal DNA auto-replicative system.

beta units (*dnaN*), forming a closed ring around the dsDNA and tightly binds the DNA polymerase to prevent it from falling off its template. The clamp loader assists in loading the sliding clamp onto the DNA at the primer:template junction of the DNA-RNA hybrid. It is composed of the subunits tau (*dnaX*), gamma (*dnaX*), delta (*holA*) and delta' *holB*). The other role is to bind the DNA polymerases[7] and to anchor the DNA helicase. DNA synthesis on the leading strand is continuous, and on the lagging strand Okazaki fragments are synthesized. The single strand stabilization protein (*ssb*) prevents the folding on itself or other interactions of the ssDNA in the lagging strand. A DNA ligase (*ligA*) covalently links the different Okazaki fragments after RNA primer excision by DNA pol I (*polA*). A type II topoisomerase, a DNA gyrase complex (*gyrA* and *gyrB*) is necessary to relax the DNA from the superhelical turns that are introduced during replication as a response to the change in topology. Finally, termination of replication is enabled by a termination utilization substance protein *tus* that binds with an extremely strong affinity to the *terB* and *terC* sites, blocking the helicase. This proposal of a minimal DNA auro-replicative system leads to a mini-chromosome of ca. 25 kb [9] and the components of the replication fork are schematically depicted in Figure 4.2 [11].

This minimal DNA replication system is based on the existing system in *E. coli*, and one could question what other minimal DNA replication systems can be found when starting from different organisms. In the emerging field of the minimal cell, *E. coli* is almost always taken as a starting point for bottom-up mimicry. This is logical because it is the most well-studied bacteria with proven potential in biotechnology and synthetic biology. On the other hand, the great thing of the bottom-up approach of constructing a minimal cell is that it does not constrain us to combine modules from only one organism like *E. coli*, so I am of the opinion we can be more open exploring other systems. Unless we aim to build a minimal bacteria like *E. coli*, in which case we should spend more

is greater than 50 kb.

[7] Three of them, one on the leading strand and two for the lagging strand synthesis.

attention to the insights gained from the top-down approach. In the next sections, we will not follow the implicit bias *'What is true for E. coli is true for a minimal cell'*[8], and will explore other DNA replication systems.

4.1.3. DIVERSITY OF DNA REPLICATION MACHINERIES: ORIGIN IN VIRUSES?

Tracing evolution back in time will undoubtedly lead to simpler, less complex forms of cellular life than we observe today. We cannot trace evolution further back than LUCA, though if we could, we would at one point find the first universal common ancestor, which would be a great example of a minimal cell. In this light, one also assumes that we will find a minimal DNA replication system if we trace back evolution in time. Unfortunately, although some parts of the replicative DNA apparatus appear universal in all three domains (clamp, clamp loader), there are large differences between Bacteria and Archea/Eukaryotes. This makes it hard to reconstruct the replicative machinery at the early stage of LUCA. The question 'How did different DNA replication proteins evolve?' is preceded by the fundamental question: 'When did DNA originate?' Hypothesis to address these vital questions are elaborately discussed in literature, and the main ideas introduced by Forterre and other critical thinkers and creative scientists will be summarized here [13–17].

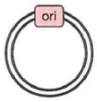
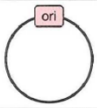
DNA could only have originated after the invention of modern complex proteins, in an already elaborated protein/RNA world. DNA cannot have originated in an RNA world[9] because the reduction of ribose requires the formation of stable radicals and this complex chemistry can degrade the ribozyme itself [10]. A major advantage of DNA versus RNA as the genetic material is its higher stability, by preventing the 2'-oxygen attack on the phosphodiester bond that can occur in RNA, and this paved the way to the formation of larger genomes. But in what scenario was DNA selected to replace RNA? This is were viruses comes into play. That viruses were already widely spread during the time of LUCA is corroborated by multiple observations, e.g that there are common features between viruses that infect members of different domains such as homologous capsid proteins and ATPases for protein packaging, and that some structural features of viral proteins, like the typical double-jelly-toll fold, is absent in cellular proteins, confirming that these structures originated from a common ancestral protein. The idea is that selection pressure would have favoured the stepwise RNA to DNA transition in viral lineages, as a strategy to resist cellular attack against viral genomes by RNases. In steps, viruses could first have evolved U-DNA and later T-DNA. This hypothesis is supported by the observation that many viruses encode proteins that are critical for the first steps of RNA-to-DNA transition, such as ribonucleotide reductase, reverse transcriptase and thymidylate synthase. The production of new genes in viruses could occur when their genomes replicate and/or recombine during the intracellular phase of the life-cycle.

[8] An adaption of the original quote from the Delft microbiologist Kluyver *'From elephant to butyric acid bacterium–it is all the same'* that has later been paraphrased by another famous microbiologist, Monod, to *what is true for the Gram-negative bacterium E. coli is also true for "E. lephant"[12].*

[9] RNA world: a period in the biosphere when both information and enzymatic activities were contained in RNA molecules.

[10] another recent hypothesis is that DNA could have originated in the RNA world by the reversal of deoxyriboaldolase step in deoxyribonucleotide salvage (DERA) pathway that does not include chemistry with free radicals. [15]

Table 8-1. The Five Types Of Replicon

Replicon	Chromosome	Examples
Type I Circular dsDNA with internal replicators		Most bacteria, bacteriophage λ[a], plasmids, archaea, papillomaviruses, polyomaviruses, herpes simplex virus[a], Epstein–Barr virus[a], mitochondria
Type II Linear dsDNA with internal replicators		Some bacteria, bacteriophages T7 and T4, budding and fission yeast, flies, vertebrates
Type III Linear dsDNA with terminal replicators		Adenoviruses, bacteriophage ϕ29
Type IV Circular ssDNA with internal replicators		Icosahedral bacteriophages, filamentous phages, geminiviruses, circoviruses
Type V Linear ssDNA with terminal replicators		Parvoviruses

[a]These viral chromosomes are linear in their virions, but circular in their host cell.

Figure 4.3: Five different DNA replication mechanisms found in nature [18].

The relevance of the scenario relying on a viral origin of DNA is that it would explain why genomes and associated replication mechanisms are more diverse in viruses than in cells. According to this 'out-of-virus' scenario, many different DNA replication machineries first originated in the viral world before transferring to the cellular world. The two types found in modern cells represent a subset that originally appeared in early RNA-based cells and their viruses.

This should then stimulate us, in our quest to find a minimal DNA replication system, to be inspired and learn from viral DNA replication strategies.

To give the reader an idea of the different DNA replication mechanisms nature has invented, it is highly recommended to browse the book of Genome Duplication by De-Pamphilis and Bell, where they have summarized them in the following table (Figure 4.3) [18].

4.2. CANDIDATES FOR A MINIMAL DNA REPLICATION SYSTEM

Now that we are aware of the different DNA replication types existing in nature, we will focus in this section on a few relevant mechanisms before we argue at the end of this chapter which minimal DNA replication systems we chose to study in this thesis.

4.2.1. VIRAL DNA REPLICATION

More than 100 years have past since the first discovery of bacterial viruses and we know now that they are very abundant [19]. For example, in oceans the number of free virions out compete any form of marine life by an order of magnitude [20]. This means that there are a lot of unknown viruses out there, and thus new viruses, viral proteins and

even DNA replication mechanisms are discovered regularly [21] In the following paragraphs we will focus on bacterial viruses of which the DNA replication mechanism is well understood. we focus on bacterial viruses rather than archeal or eukaryotic viruses because our platform for the minimal cell, the reconstituted gene expression system, is based on the *E. coli* bacteria. Of interest for our understanding is how the different DNA replication systems initiate the reaction mechanism and how the so-called termination paradox is solved in the case of linear DNA with internal origins.

LINEAR DNA REPLICATION IN T4 AND T7 VIRUSES

The T-phages are a collection of viruses, assembled in 1945, that infect *E. coli* [22]. They are well known for their beautiful mind-blowing head-tail shapes as electron microscopy images revealed to us, similar to the viral particle of the phi29 in Figure 6.1A. The T4 virus has a linear genome of 170 kb and around 300 genes. At least 20 proteins are involved in DNA replication, plus another 25 proteins fulfilling tasks in the modification of the DNA and digestion of the host DNA to provide nucleotides for the genome synthesis [3]. The T4 replisome has less proteins than the *E. coli* replisome (8 vs. 13) but overall resembles the architecture with its main components: the polymerase (gp43) in conctact with the clamp (gp45) and clamp loader (gp44/62) forming the holoenzyme[11], a helicase-primase complex is formed by the proteins gp41 and gp61, the helicase loading on the replication fork is aided by protein gp59 and the SSB (gp32) complements the replisome, as depicted in Fig. Initiation of replication of the genome occurs by a variety of mechanisms and does not have a single origin [24]. One mechanism for initiation is the formation of R-loops[12] RNA at a promoter site that is followed by an AT-rich downstream unwinding element, where the RNA after processing by RNase H can prime leading strand DNA synthesis. Another mode is the recombination-dependent replication, where initiation happens on the invading 3'-end of recombination generated D-loops structures[13]. Recombination plays an important role in the overall process of maturating DNA replication as well: the T4 genome contains terminal redundancy of ca. 2 kb inverted repeats, that allow the formation of multiple-length linear DNA upon recombination for the packaging process [25].

Although the complete process of T4 genome duplication seems rather complex, the *in vitro* reconstitution of the eight T4 replisome proteins provides a fast and processive isothermal DNA amplification system [23]. Schaerli et. al demonstrated that a covalently closed circular plasmid of 4.7 kb could be amplified 1100-fold in 1 h by a circular nicking endonuclease dependent amplification. In this system, an endonuclease nicks a covalently closed plasmid after which the T4 replisome initiates DNA polymerization replacing the nicked strand. Since the endonuclease also nicks the newly synthesized strand, a ssDNA copy of the strand that is released is then primed by a specific primer to initiate second strand synthesis. The resulting DNA is thus a linear copy of the circular plasmid. Interestingly, this study also showed that whereas a covalently closed plasmid cannot be amplified, linear genomic DNA can be unwound and primed at its ends by

[11] Polymerization rate ca. 250 nt/s and processivity around 20 kb [23]

[12] An R-loop is formed when a complementary RNA-DNA hybrid is formed, displacing the other DNA strand as a single stranded-loop.

[13] D-loops consist of three strands of DNA when duplex DNA takes up an additional strand.

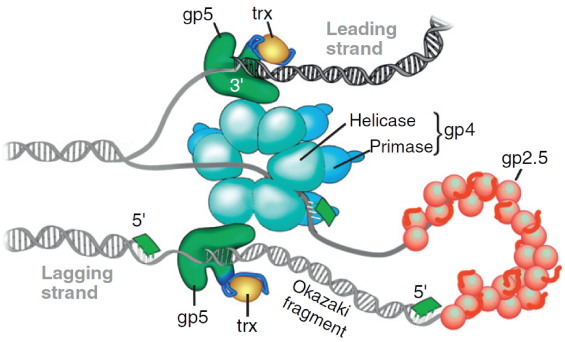


Figure 4.4: Schematic representation adapted from [26] to show the proteins and their interactions in the T7 replisome. Other proteins involved in full-length replication of the genome are the exonuclease (gp6), ligase (gp1.3) and T7 RNAP (gp1).

the T4 replisome. This primase-based whole genome amplification led to 200-fold amplification in 2 h, yielding products of different lengths that were suitable for sequencing applications.

Another well studied virus of the T-series is the simpler T7 phage that has a linear genome of 40 kb and is special in the sense that it encodes its own RNAP and all, except one, of the necessary proteins for DNA replication. The T7 replisome is constituted by only four proteins: the DNAP (gp5) with its host processivity factor thioredoxin, a fused helicase-primase (gp4), and an SSB (gp2.5). The thioredoxin substitutes the role of the sliding clamp in the aforementioned replisomes by binding and reorganizing a flexible domain of the DNAP to better grasp the DNA, increasing the processivity by two orders of magnitude to around 800 nt [26]. The activity of the DNAP is furthermore enhanced by the interaction with the helicase, which is a reciprocal stimulation. The gene 4 specifies the two overlapping primase and helicase proteins, and they are expressed in a 1:1 ratio by restarting translation in the same reading frame. The polypeptide of 63 kDa contains both primase and helicase, the smaller 56 kDa peptide only the helicase protein [27]. The helicase protein moves uni-directionalyl in the 5'-3' direction, hydrolyzing NTPs as an energy source[14] [28]. The primase activity is not dependent on a specific DNA sequence, though the gp4 protein binds preferably to 5'-(G/T)GGTC-3' and (G/T)TGTC that are found at the origins of replications in the T7 genome. The SSB protein, apart from its main role of coating the unwound DNA of the lagging strand, also physically interacts with the DNAP and gp4 to stimulate their activities. DNA replication initiates bi-directionally at the primary origin of replication by a transcription mediated DNA unwinding. The primary 200-bp origin contains two T7 promoters, followed by a 61-bp AT-rich region that functions as a DNA unwinding element [29]. The T7 RNAP is rapidly

[14] The preferred *in vitro* substrate is dTTP, though the helicase hydrolizes all NTPs except the CTP. Since *in vivo* the ATP concentration is an order of magnitude higher than that of dTTP, ATP is probably the most utilized substrate [28].

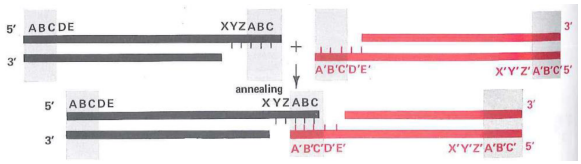


Figure 4.5: Model of how the termination paradox can be solved by terminal redundancy and the formation of concatemers as intermediates in replication of a linear dsDNA genome, adapted from [3]. The letters ABCDEXYZ represent sequences and the A'B'C'D'E'X'Y'Z' their complementary counterpart. The unreplicated 3'end region anneals to another molecule at the redundant region and the gaps can be filled.

displaced by the T7 DNAP that initiates leading strand DNA synthesis on the short RNA primer, aided by the SSB that prevents re-annealing. The primase-helicase can bind the opposite unwound strand of DNA to complete the replisome assembly and synthesizes primers for the lagging strand synthesis. The RNA primers are later removed by the exonuclease (gp6), filled in by the DNAP and ligated by the phage-encoded (gp1.3) or host ligase. From this example, it should become clear what the termination paradox is. The gap left by the excision of RNA primers, can only be filled in again by the DNAP if there is a 5'-3' primer present upstream of that location. At the ends of the linear template, this will not be the case, and since there is no polymerase that can extend in the other direction the gap remains. The T4 phage solved this problem by recombination of its genome and the T7 has a similar solution. The T7 genome has likewise a terminal redundancy in the form of a 160-bp repeat. This stimulates the formation of concatemers, where the overhanging 5'-ends of the genome anneal to each other (Figure 4.5). The gaps can now be filled in by a polymerase. The further processing of the concatemers requires several proteins *in vivo* [3].

PROTEIN-PRIMED DNA REPLICATION

Instead of priming synthesis from RNA, it is possible to utilize the (3-)hydroxyl group of an amino acid to prime nucleotide incorporation by a DNAP. This mechanism containing terminal origins is found in the phi29 virus and the family of adenoviruses. We will discuss DNA replication of the phi29 virus that has a linear genome of 19 kb and infects the bacteria *Bacillus subtilis*. There are four proteins involved in *in vivo* genome replication, which is schematically depicted in Figure 4.6 [30]. However, complete replication of the genome *in vitro* can be performed by only two proteins, the terminal protein (TP) and the DNAP [31]. The terminal protein forms a 1:1 complex with the DNAP in solution, stimulated by the presence of the positive ammonium ion. In the presence of Mg^{2+} and dATP, the formation of a covalent bond between dAMP and the OH-group of the amino acid serine of the TP acts as a primer for DNA synthesis. The initiation occurs at the second nucleotide of the template and a sliding back mechanism of the initiation complex sets the dAMP base-paired to the 3'-terminal nucleotide [32][15]. After the DNAP has synthesized 10 nt, dissociation of the TP occurs resulting in DNAP-DNA interactions and elongation of the newly created DNA primer [33]. The phi29 DNAP binds very tightly to ssDNA and contains a strand displacement activity allowing completion of the full

[15] Posing the requirement for a replication template to contain two similar bases at position 1 and 2. The phi29 genome contains a 12bp inverted terminal repeat: 5'-AAAGTAAGCCCC-3'.

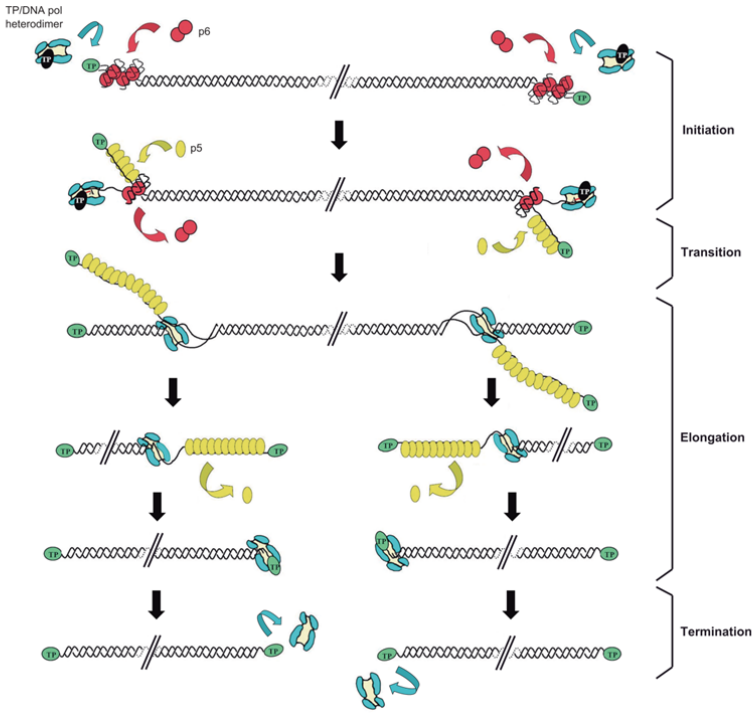


Figure 4.6: Schematic representation of phi29 DNA replication *in vitro*, adapted from [30].

genome without accessory proteins. The two other proteins involved in DNA replication are an SSB (p5) and a double-stranded binding protein (DSB, p6) that increase the efficiency of DNA replication enormously. The p6 protein binds to DNA in a dimeric and non-cooperative manner all along the viral genome as a histone-like protein, inducing a conformational change. The binding restrains positive super-coiling[16] and can locally open the duplex DNA, stimulating the initiation of DNA replication at the terminal origins. The SSB protein protects the ssDNA from nucleases present in the host, but also directly contributes to efficient DNA replication by its helix destabilizing activity [34]. In addition to preventing the non-productive binding of replication proteins to the displaced strands, it appears to be critical for the progression (maturation) of the multiple replication forks initiated at both phi29 DNA replication origins. An *in vitro* reconstruction of the DNA amplification system of the four phi29 proteins was developed by the group of Prof. Salas and this system can amplify the natural phi29 genome and a heterologous DNA template containing the same origin sequences as the phi29 genome successfully [35].

[16] The protein has a binding preference for less negatively super-coiled DNA which probably helps in binding to the viral DNA more than to the bacterial DNA.

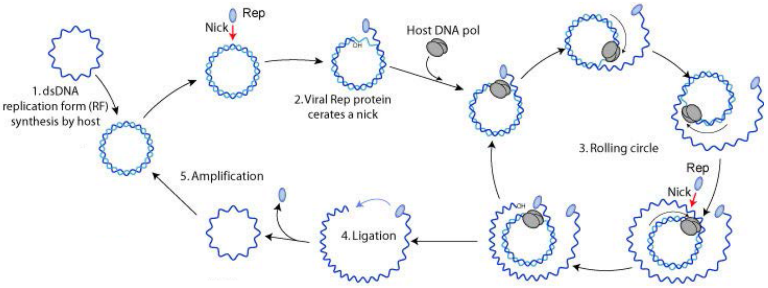


Figure 4.7: Replication of single stranded viral plasmids is initiated by a viral Rep protein and performed by the host DNA replication system, adapted from [37].

4

CIRCULAR PLASMID DNA REPLICATION VIA SINGLE-STRANDED INTERMEDIATES

As discussed in the case of *E. coli*, replication of circular chromosomes in bacteria is bi-directional starting from a single origin and produces after termination at the *ter* sites two double-stranded chromosomes. There are circular plasmids, like R100 in bacteria that are replicated by a different mechanism, namely via a single-stranded intermediate, which we will explain concisely here. Single-stranded DNA viruses, such as the pili filamentous M13 virus and the small phiX174 virus, are also replicated via this mechanism by host enzymes. The initiation of DNA replication is performed by a plasmid or viral encoded Rep protein that has sequence specific DNA binding and type I topoisomerase activities. It creates a nick at a defined origin, after which the host replication machinery (helicase, SSB and DNAP) perform leading strand synthesis until the strand has been fully displaced. Other Rep proteins perform a sequence of cleaving and rejoining events resulting in a relaxed closed circular plasmid and a circular ssDNA leading strand [36]. The released ssDNA is converted into dsDNA by host proteins, including the synthesis of an RNA primer by RNA polymerase or primase. This is schematically depicted in Figure 4.7.

FROM A SINGLE-STRANDED RNA TO A DOUBLE-STRANDED DNA GENOME

Retroviruses like the HIV are viruses that use RNA as their carrier of genetic information, which is converted into dsDNA in the host cell. The dsDNA is integrated in the host genome for subsequent expression of the essential viral proteins by the host cell transcription and translation machinery. The enzyme that carries out the reverse transcription of the viral RNA is the reverse transcriptase (RT). DNA reverse transcribing viruses also exist[17], in which the DNA is first transcribed by a host RNA polymerase to produce the genomic RNA that is then further processed by a RT reaction. The RT possesses three activities in one enzyme: (1) an RNA dependent DNA polymerase activity, (2) a DNA dependent DNA polymerase activity and (3) an RNase H function which can degrade RNA from the intermediate DNA:RNA hybrid template. The mechanism of reverse transcription by the RT is schematically described in Figure 4.8. In short, the reverse transcription reaction is primed by a tRNA that binds to the RNA template, to create a DNA sequence

[17] For example, the Caulimoviridae family infects plants with virions containing up to 8 kb open circular ds DNA genomes.

that after strand transfer serves as primer for first strand synthesis. A second DNA sequence is created during this process that can then serve as a primer for second strand synthesis, resulting in a double-stranded DNA template without loss of sequence information.

REPLICATION OF AN RNA GENOME BY RNA DEPENDENT RNA POLYMERASES

Considering the possibility that LUCA possessed an RNA genome [2], and was predated by a simpler form of a ribo-cell, it is interesting to investigate the idea of an RNA genome for the minimal cell. There exists a variety of RNA viruses with either a ssRNA genome (+ strand or - strand) or a dsRNA genome, with the longest genome sizes around 30 kb for dsRNA viruses (reoviridae) and ssRNA viruses (coronaviridae) [39]. Transcription from dsRNA templates proceeds very different than in dsDNA viruses, where a DNA dependent RNAP recognizes promoter and termination signals on the template. Transcription in dsRNA virus is performed by the viral RNA-dependent RNA polymerase (RdRP) that is also responsible for replication of the RNA genome. Transcription and replication are actually part of the same process, because the transcribed mRNA serves both for translation and replication. The RdRP produces the mRNA by a strand displacement reaction from the genome or its segments. This results in most viruses in polycistronic mRNA templates that will be translated by the host ribosome into polypeptides and processed by a protease to obtain the separate proteins. Another strategy of dsRNA viruses is to encode its proteins in different segments, like the Reoviridae family that has a genome of 10 to 12 different segments. The segment size ranges from 0.2 to 3.0 kb and results in one mRNA per protein[18]. Similarly applying viral replication and transcription strategies of dsRNA viruses to the minimal cell concept seems cumbersome, as these processes are not decoupled. In the extreme case it leads to the scenario where one mRNA encodes for all the proteins without control on the protein stoichiometry and with difficulty in producing separate rRNA. A more viable approach would be to segment the genome, though it does not avoid the coupling of the transcription and replication processes. Naturally, these viral strategies are likely different than the mechanism(s) present in the predates of LUCA. All we would need is a separate RdRP that can recognize initiation and termination signals. This RNAP is not found in modern day cells or viruses[19], but it could be evolved from well-studied RNAPs of which their crystal structure is known [40]. Good candidates can be the T7 RNAP, that has been shown to be flexible in template and substrate recognition or the RNAP found in LUCA [41–44].

4.2.2. *In vitro* DEVELOPED DNA AMPLIFICATION SYSTEMS

We can now appreciate the four different ways we have seen for initiating DNA replication: by DNA nicking, protein-priming, DNA transcription, or with a DNA helicase loader and primase. Inspired by nature, several *in vitro* DNA amplification methods have been invented by scientists over the past three decades. In most of these methods the introduction of primers and heat-denaturing steps are utilized to bypass the need of accessory proteins to initiate DNA replication and solve the termination paradox. The most

[18] one segment can encode more proteins with the leaky-scanning mechanism or other protein processing strategies

[19] To my best knowledge.

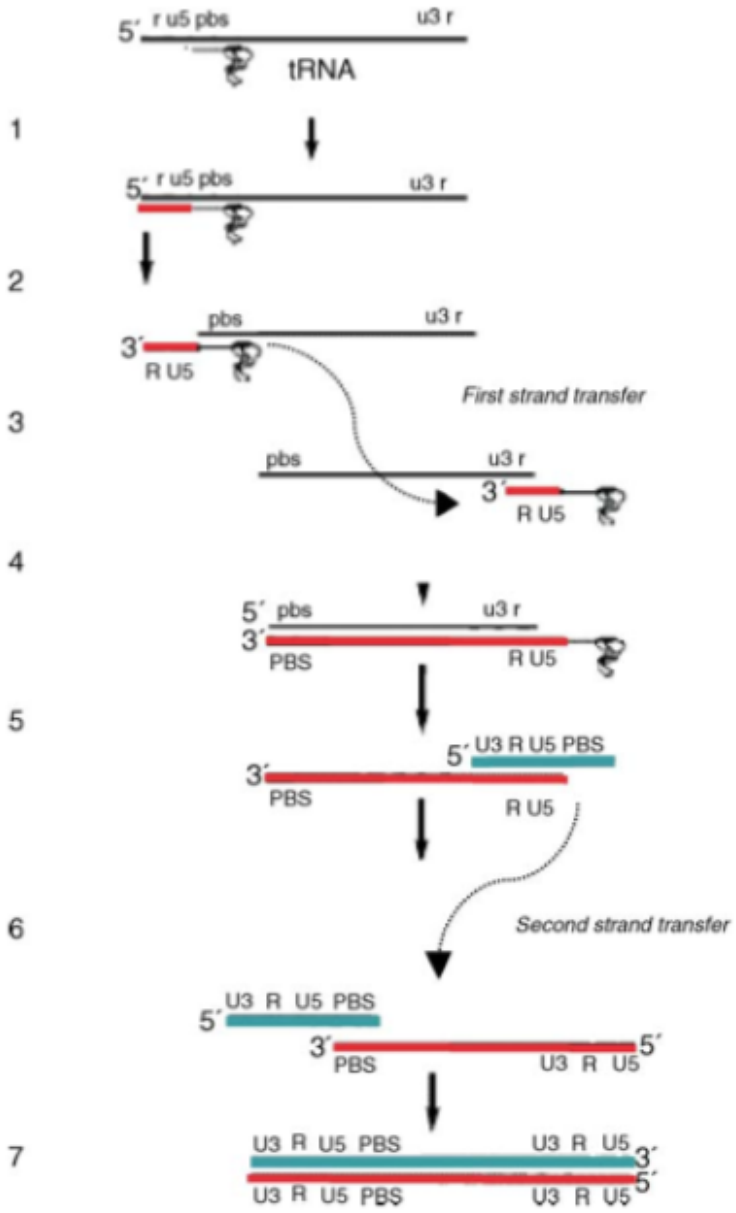


Figure 4.8: Mechanism of reverse transcription in retroviruses, adapted from [38]. The DNA is indicated by thick lines, the RNA is indicated by thin lines. Step 1) First strand (minus-strand) synthesis on a viral RNA construct starts near the 5'-end of the RNA, where a tRNA binds to the primer binding site (PBS) and the 3' end of the tRNA primer is extended until the 5' end of the RNA by the RT. Step 2-3) The RNA of this short DNA:RNA hybrid is degraded by RNase H and the resulting short DNA strand, called the minus-strand strong-stop DNA (-sssDNA), subsequently binds the 3' end of the RNA at a complementary sequence. Step 4-5) The RT resumes first strand synthesis at the -sssDNA and meanwhile degrades the RNA. Some sequences are more resistant to RNase H degradation, and these tracts serve as primer for second strand (plus-strand) synthesis. The plus-strand is extended until the initial tRNA primer is reverse transcribed, generating a short DNA sequence called plus-strand-stop sequence (+sssDNA). Steps 6-7) The +sssDNA is transferred to the complementary sequence at the 3' end of the plus-strand and the DNA-dependent DNA polymerase activity of the RT completes the second strand resulting in a dsDNA.

widely used *in vitro* DNA amplification method is the polymerase chain reaction (PCR) developed in 1987 [45], which amplifies a linear target sequence. It requires thermo-cycling to denature the double-stranded DNA, hybridization of primers and incorporation of dNTPs in the new strands by a thermophilic DNA polymerase. In our context of the minimal cell framework, thermo-cycling is detrimental to gene expression since high temperatures will denature proteins. We are therefore only interested in isothermal DNA amplification methods, not exceeding temperatures of ca. 40 °C. In this section, a few ideas that are relevant when one wants to design, rather than copy an existing, DNA replication system for the minimal cell are touched upon. For a more elaborate comparison of isothermal amplification methods and PCR, as well as their applications, see reference [46].

SELF-SUSTAINED SEQUENCE REPLICATION (3SR)

The self-sustained sequence replication mimics the retro-viral strategy of RNA replication by using RNA as a template to create DNA by the enzymatic activities of a reverse transcriptase, RNaseH and a DNA-dependent RNA polymerase [47]. The specific enzymes used here are the avian myeloblastosis virus (AMV) reverse transcriptase, the *E. coli* RNase H and the T7 RNA polymerase. In the first step of the cycle, a primer hybridizes to the 3'-end of the target sequence. The reverse transcriptase, being both an RNA and DNA dependent DNA polymerase, elongates the primer to create the complementary DNA (cDNA). During elongation, the RNase H subunit of the AMV already digests partly the RNA and further processing is completed by the additional RNaseH enzyme. A second primer can now bind to the 3'-end of the single-stranded cDNA and will be extended by the DNA dependent DNA polymerase activity of the AMV to make a double-stranded DNA of the initial RNA target sequence. Both primers contain a 5'-extension with the sequence for the T7 promoter, so that the dsDNA can be transcribed producing both new sense and anti-sense RNA transcripts. These RNA transcripts can enter the cycle again, resulting in the self-sustained nucleic acid sequence amplification of both RNA and DNA as depicted in Figure 4.9. A 214-nt target of the HIV-1 RNA could be amplified isothermally at 37C a million-fold with input concentrations as low as 10^{-5} fmol [47]. It is also reported that an efficient 3SR reaction can be performed when only one primer contains the promoter. The 3SR reaction is mainly efficient and optimized for sequences below 250 bp [48], though in principal it should be possible to complete such a cycle from RNA to DNA up to ca. 12 kb in length, which is the larger genome size of retro-viral viruses.

ROLLING CIRCLE AMPLIFICATION (RCA)

Multiply-primed rolling circle amplification is developed to amplify circular DNA from single colonies or plaques, bypassing lengthy growth periods and traditional DNA isolation methods [49]. RCA starts with a heating step to denature the DNA. Random hexamer primers then hybridize to the DNA and are extended by a DNA polymerase. The proof-reading phi29 DNA polymerase is chosen for its strand displacement capability and high processivity (more than 70 kb on ssDNA [50]). Because the 3'-5'exonuclease activity of the phi29 DNAP could reduce the yield by degrading the hexamer primers, they were made exo-resistant by introducing thiophosphate linkages for the two 3'-terminal nucleotides. After the DNAP has completed the polymerization of one full circle, the prod-

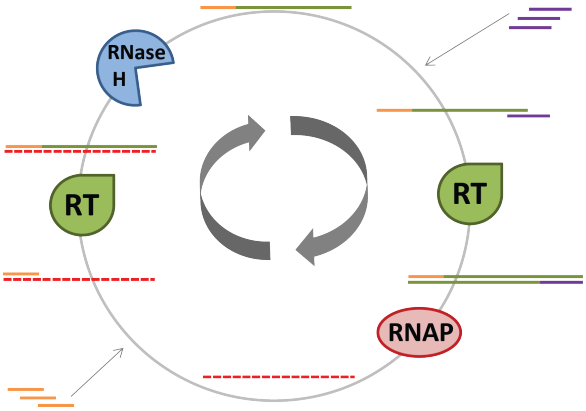


Figure 4.9: Scheme for self sustained sequence amplification, starting with an RNA sequence. For simplicity in this scheme only one primer contains a promoter sequence (sense amplification cycle). RNA = red dashed line, DNA primer tha binds to the 3'-end of the RNA = orange line, synthesized DNA by the reverse transcriptase = green line, primer containing promoter that binds to the 3'-cDNA = purple line.

uct strand is displaced and this ssDNA serves as new hybridization site for the primers, allowing the synthesis of the other strand. Multiple replication forks exist and final products can differ greatly in length. To obtain copies of the full-length sequence of the initial plasmid, final products can be cleaved by restriction endonucleases as successfully shown for the M13 with EcoRI [49]. This method can easily achieve thousand-fold amplification of ng input DNA in several hours. It has been successfully used in combination with gene expression in an *in vitro* transcription and translation system[51]. The authors aimed to answer the question if RCA-generated DNA, that contains interruptions in both strands and branched structures, can serve as a template for *in vitro* transcription. After performing multiply-primed RCA[20] on 100 pg input DNA containing a gene for GFP, IVTT components[21] were added to the same tube and the protein production was found to be comparable in efficiency to the input of 500 ng plasmid DNA. This is a promising result for applications such as high-throughput screening of bacterial colonies for a desired protein [51]. How to circumvent the heating step for application to the minimal cell framework will be discussed in the next paragraph.

STRAND DISPLACEMENT AMPLIFICATION (SDA)

The strand displacement amplification (SDA) was initially developed as an isothermal method to diagnostically detect tuberculosis as an alternative to PCR detection methods [52]. This method relies on the strand displacement activity of an exonuclease-deficient DNA polymerase that can extend the 3'-end of a nicked template. The requirement for a nicked template is cleverly solved by the use of HicII restriction endonuclease that recognises a 6-nt palindromic sequence and does not cut strands containing modified deoxynucleotides. The first step of the cycle is a heat-denaturation step, allowing the primers containing the restriction signal to bind to the target sequence. Modified nu-

[20] With the GenomiPhi V2 kit from GE Healthcare Lifesciences.
[21] Roche *E. coli*

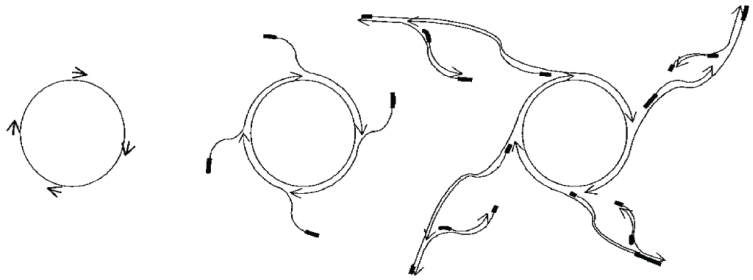


Figure 4.10: Scheme for multiply-primed rolling circle amplification, adapted from [49].

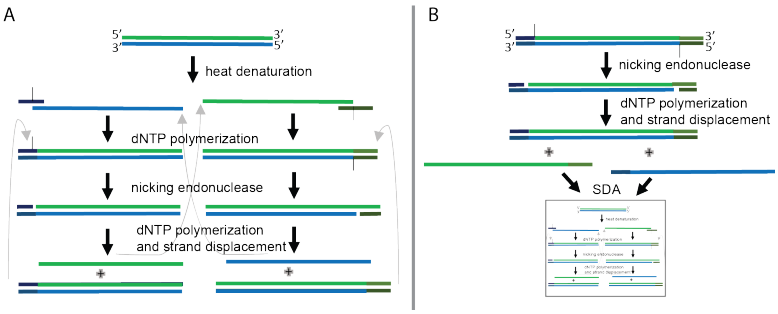


Figure 4.11: Strand Displacement Amplification (SDA) reaction with heat-denaturation step (A), reproduced from[53]. B) SDA without heat-denaturation step can be achieved by a nicking endonuclease.

cleotides, alpha-thio-dATPs, are introduced in the elongation step by the DNAP (Klenow exo-) and are therefore not present in the primer. The double-stranded DNA is targeted by HincII that nicks the unprotected primer strands of the recognition site, leaving the modified complementary strand intact. Next, the DNAP can initiate on the newly created 3'-end and thereby displace the downstream strand, resulting in the exact same dsDNA as before the cut, plus the displaced ssDNA that can enter the cycle again. Exponential amplification of a target sequence of 47 bp was observed for 60 min, and after 5 h a total amplification of 10^6 compared to the initial target amount could be reached [53]. If we wish to apply this method for DNA amplification in a minimal gene expression system, we can make two modifications. Firstly, to perform the reactions in a truly isothermal manner, the first heat-denaturation step should be omitted. To enter the cycle, the solution is to include already two nicking sites in the starting template, a species that is otherwise not an intermediate in the SDA cycle. After the first nicks and DNA polymerization reactions, the two strand displaced DNA can readily enter the cycle as depicted in Figure 4.11B. Note that the initial template is not amplified, but the reaction is sustainable after the first cycle. Secondly, a different restriction endonuclease should be utilized in order to eliminate the use of modified dNTPs which can otherwise affect subsequent transcription reactions. Nicking endonucleases that cut only one strand of the dsDNA target are not abundant in nature, but there are a few like the Nt.BspQ1 from *Bacillus*

stereothermophilus and a range of engineered ones commercially available[22]. If the SDA is also efficient for large templates, is a remaining question. A related proposal for DNA replication in a minimal cell context is discussed in the next section.

HELICASE-DEPENDENT AMPLIFICATION (HDA)

In the helicase-dependent amplification system, the separation of DNA strands is performed by a helicase, mimicking the *in vivo* mechanisms. The first approach for such a system was based on the helicase UvrD, that, together with the accessory protein MutL[23], has a blunt-end unwinding activity in the 3'-5' direction [54]. After the unwinding, coating of the exposed ssDNA by single-stranded binding protein (SSB, T4 gene 32) is necessary to prevent the dehybridization of the strands. Sequence specific primers can bind to the exposed ssDNA and are extended by a DNAP polymerase, here the exo$^-$ Klenow fragment of DNA polymerase I. With these enzymes carrying out the amplification cycle, depicted in Figure4.12, over a million fold amplification was achieved of a ca. 100-bp target sequence in the genome of an oral pathogen. The HDA method is also compatible with crude bacterial cell samples and can detect pathogens in the presence of human blood [54]. Further developments of this method include the engineering of a helimerase, a fusion protein of a helicase and DNA polymerase, that was able to amplify targets up to 2.3 kb [55]. This truly isothermal method is a potential candidate for the minimal cell[24], with some small modifications. To not be limited by the low processivity of the UvrD helicase (ca. 100 bp [56]), the DNAP should have high processivity and strand displacement activities, therefore the phi29 DNA polymerase is a good candidate. To prevent degradation of the primers in solution, a mutated version of the phi29 DNAP (D12AD66A[57]) without exonuclease activity can be utilized. Another option without compromising the fidelity of DNA replication is to use exo-resistant primers, as described for the RCA method.

Later on, the T7 replisome became the machinery of interest for HDA reactions, both for circular and for linear DNAs. When using the engineered 3'-5' exonuclease-deficient T7 DNAP[25], instead of the wild-type DNAP, the T7 replisome can perform a helicase-dependent strand displacement reaction similar to the rolling circle amplification, depicted in Figure 4.13. An advantage of the T7 replisome is that the gp4 contains both helicase and primase activities, so the requirement of adding primers to initiate DNA amplification can be removed in some cases [56]. Based on this, a primase-based whole genome amplification method was developed, where dsDNA templates are denatured by the gp4, omitting any heat denaturation step as well [58]. This method is the same as described for the *in vitro* reconstitution of the T4 replisome. We should keep in mind that the products are not exact copies of the input DNA, rather discontinuous or branched DNAs (due to displaced Okazaki fragments) of various sizes. This is mainly the result of decoupling leading and lagging strand syntheses by having a DNAP that can perform strand displacement activity coupled with the helicase. An interesting proposal to retain the original DNA template for circular plasmids after RCA will be described in the next

[22] see for example https://www.neb.com/products/restriction-endonucleases/hf-nicking-master-mix-time-saver-other/nicking-endonucleases/nicking-endonucleases/nicking-endonucleases-available-from-neb

[23] in *E. coli*, MutL recruits UvrD in the process of mismatch repair

[24] Accepting the introduction of oligonucleotides as external resources.

[25] Sequenase

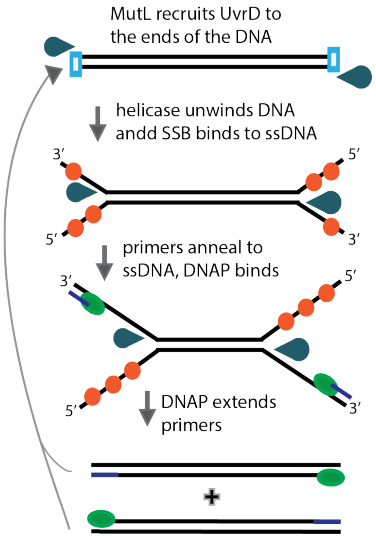


Figure 4.12: Helicase-dependent amplification (HDA) reaction, reproduced from [54].

section [59].

4.3. PROPOSED DNA REPLICATION MECHANISMS FOR THE MINIMAL CELL

We gave a brief report of several DNA replication machineries found in nature, mainly in viruses because these systems are naturally more minimal than found in higher organisms. It is not straightforward to choose a naturally occurring DNA replication -or any other- system and from there try to reach a system that performs the same task faithfully with a minimal amount of components or complexity. This is exemplified in the following remark of Forterre [14]:

> 'Fundamentally, organisms are not systems but historical products. As a consequence, it is not possible to understand them without a sound evolutionary framework.'

The idea to combine different mechanisms as is successfully applied for the development of *in vitro* amplification systems, may also be a fruitful approach for the minimal cell framework.

4.3.1. ROLLING CIRCLE AMPLIFICATION WITH RECOMBINATION

The first detailed proposed mechanism for a DNA replication scheme in a minimal cell appeared in a review from Anthony Forster and George Church in 2006 [59]. Their idea is based on a nicked circular dsDNA that undergoes rolling circle strand displacement by the phi29 DNA polymerase. This will result in a long oligomeric ssDNA (provided the exchange of DNAPs, since the processivity of phi29 DNAP is ca. 70 kb on ssDNA [50]).

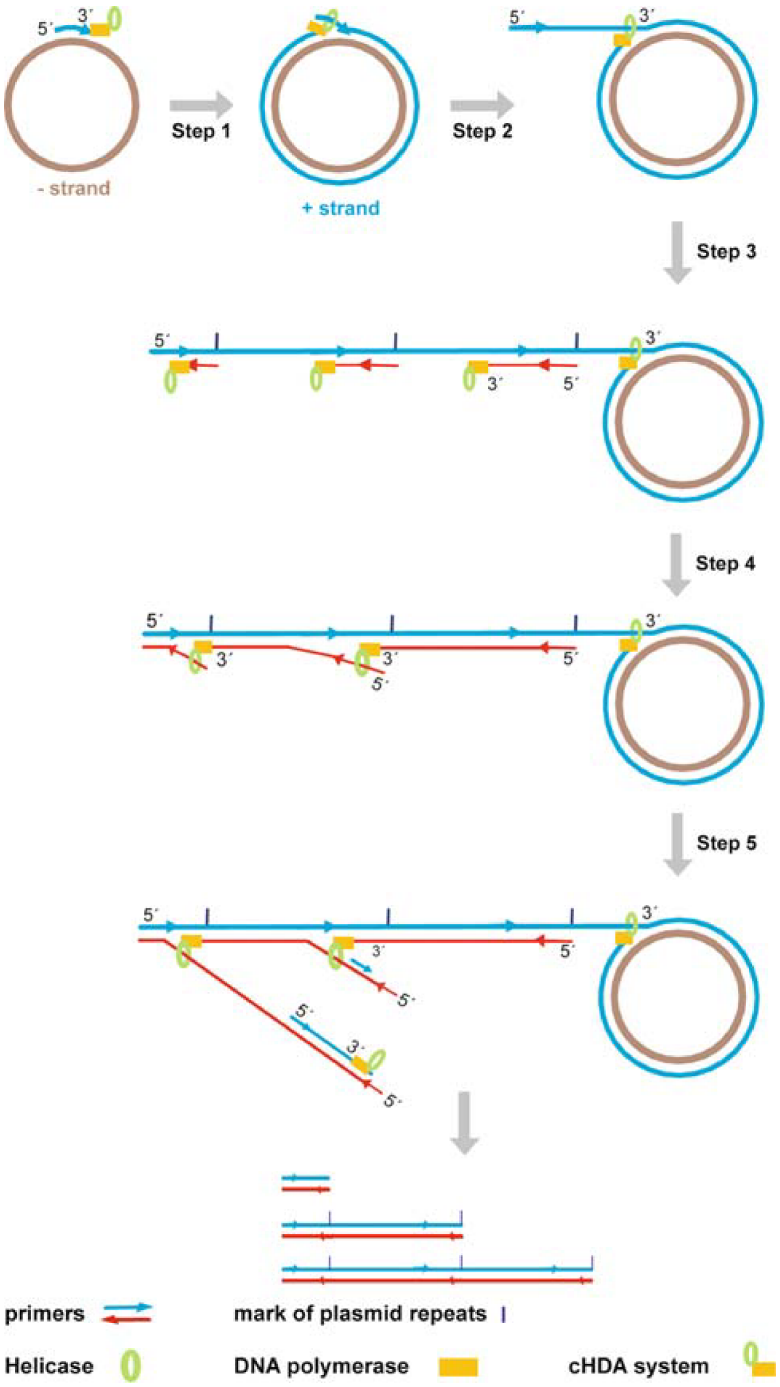


Figure 4.13: Helicase-dependent amplification of a circular template by the T7 replisome proteins, adapted from [56].

Lagging strand synthesis is then initiated by the binding of RNA primers to the ssDNA, upon which the DNAP can bind and polymerize the second strand of DNA. An additional modification step is necessary to come back to the initial replication substrate: The processing of the concatemeric dsDNA into monomeric DNA circles. They propose to utilize the homologous recombination mechanism at Lox sites catalyzed by the Cre recombinase. Cre recombinase is a Type I topoisomerase from the P1 bacteriophage which catalyzes site-specific DNA recombination at a 35bp site called LoxP [60]. When there are two LoxP sites on the same DNA molecule, the result after recombination is a covalently closed circle. The critical comments one could place for this proposal is that it is incomplete, and it is worth considering:

- The RNA primers are not replaced by DNA in the process, what are the consequences?
- A significant amount of ssDNA is formed during this process, which can form secondary structures and impede replication.
- To include the specific nicking enzyme in the gene list of the minimal cell.
- How to regulate termination of the strand displacement reaction by the DNAP?
- A big challenge mentioned by the authors themselves: how to avoid defective by-products such as linear DNAs or oligomeric circles and boost the efficiency of monomeric circular templates?

Apart from these comments, one should definitely value and appreciate this encouraging proposal. The interesting aspect of this DNA replication scheme is that it is inspired by nature and combines mechanisms found in different viruses. Both proteins come from bacteriophages and are therefore likely to be compatible in a similar chemical environment. Combining proteins from different species (in this example: bacteria and their viruses) is an advantage of approaching the minimal cell with a bottom-up synthetic biology approach. Therefore, this paper serves as a starting and reference point for many projects within the minimal cell community.

4.3.2. CURRENT RESEARCH ON DNA REPLICATION IN MINIMAL GENE EXPRESSION SYSTEMS

We will now have a look at the research performed during the past 10 years after the detailed proposal for a minimal cell, where the focus lies on bringing together gene expression systems and genome replication. Cell-free expression systems are interesting for their possibility to reconstruct *in vitro* a biological complex or process, starting from DNA templates. A striking example is the demonstration that the bacteriophage T7 can be expressed from its genome and self-assemble into its viral particle in a cell-free extract of *E. coli* [61]. DNA processing enzymes have also been successfully expressed in the more minimal PURE system,like the *E. coli* RNAP holoenzyme and a reverse transcriptase [62, 63]. These are promising examples of the last decade and inspiring for further studies into a reconstructed genome replication system. In this section, we discuss the acquired knowledge gained from recent studies in which DNA and RNA replication enzymes have been expressed in the PURE system.

E. coli REPLISOME: IMPORTANCE OF CORRECT FOLDING OF PROTEINS

Recently, the group of Prof. Nomura in Japan accomplished to synthesize all the proteins in the *E. coli* DNA polymerase III holoenzyme from its 9 genes (*dnaE, dnaQ, holE, dnaX, holA, holB, holC, holD, dnaD*) and the initiator protein (*dnaA*), the DNA helicase (*dnaB*) and the DNA primase *dnaD* in a minimal gene expression system in their active form [65]. This study has provided the minimal cell community with a variety of relevant information. First of all, it has been shown earlier that of all the cytoplasmic *E. coli* proteins that are expressed in the chaperon-free PURE system at 37 °C, around 55% are aggregation prone versus a 45% soluble fraction [64]. To solve the potential problem of aggregation prone, non-functional protein expression, reactions were carried out at 27 °C, which led indeed to a functional Pol III holoenzyme as shown in a replication assay with the G4 phage ssDNA[26]. This finding also shows that the PolIII holoenzyme can spontaneously self assemble without the need of accessory proteins. Secondly, the combined expression of the Pol III holoenzyme and the genes for the initiator protein (*dnaA*), DNA helicase (*dnaB*), helicase loading factor (*dnaC*) and DNA primase (*dnaG*) resulted in succesful replication of A-site ssDNA[27]. A GFP gene that was encoded in the same ssDNA plasmid as the A-site, was expressed after DNA replication was carried out, reaching two rounds of central dogma in a one-tube reaction. These results are promising as they show the concerted functioning of 14 expressed and functional proteins in the PURE system, the most complex system so far. Also interesting to note is that the *dnaX* gene encodes for two proteins that are produced by translational frame shifting *in vivo*. That translation frame-shifting *in vitro* by *E. coli* ribosomes on the mRNA from the *dnaX* gene works was already shown in an S30 *E. coli* extract [66][28]. Importantly, no difference in ratio of the produced tau (71 kD) and gamma (47 kD) proteins was observed when changing the temperatures in the range of 23-37C in that study. The successful expression of both proteins from the *dnaX* gene in the PURE system proves that the -1 translational frame-shift mechanism[29] can be used as a viable tool for future projects.

RCA WITH RECOMBINATION: OPTIMIZING NUCLEOTIDE RATIOS

Another Japanese group, of Prof. Yomo, has tried to realise the rolling circle amplification scheme as proposed in 2006[59, 67]. A 4.7-kb plasmid encoding for the phi29 DNA polymerase was expressed in the PURE system[30] in the presence of dNTPs and random hexamers. This system is capable of rolling circle amplification, though recombination by the Cre recombinase on Lox sites was not successful. That the phi29 DNAP can initiate RCA in this system is already interesting in itself because the starting plasmid is a ds-DNA and the random hexamers therefore cannot bind to the template as in the multiply-primed RCA reaction in which the primers hybridize to the ssDNA [49]. A possible expla-

[26] protein expression at 37 °C resulted in an inactive holoenzyme. The idea to lower the temperature came from literature were it is reported that *E. coli* lacking two chaperons can grow at temperatures below 30 °C

[27] A-site ssDNA is a circular ssDNA with a hairpin-form duplex that has a DnaA binding site and requires the same proteins for replication as chromosomal DNA [65].

[28] but did not work with the dnaX gene with eukaryotic extract translation system and is thus specific to the host ribosome/translation machinery

[29] Caused by a A AAA AAG heptamer motif and greatly enhanced by stable secondary stem-loop structure in the RNA directly downstream [66].

[30] The group of Prof Yomo reconstitutes their own PURE system and this differs slightly from the commercially available PURE*frex*. The exact composition is stated in the paper itself [67].

Table 4.1: Concentration of nucleotides (mM) in different systems. The initial concentrations in the experiment of Prof. Yomo's group [67] is given by $PURE_{in}$ and final optimized concentrations by $PURE_{opt}$. The PURE*frex* system contains no dNTPs.

dNTP	*E. coli*	$PURE_{in}$	$PURE_{opt}$	**NTP**	*E. coli*	$PURE_{in}$	$PURE_{opt}$	PURE*frex*
dATP	0.18	0.1	0.3	**ATP**	3.0	3.75	0.375	3
dGTP	0.12	0.1	0.3	**GTP**	0.92	2.5	0.25	3
dCTP	0.07	0.1	0.3	**CTP**	0.52	1.25	0.125	1
dTTP	0.08	0.1	0.3	**UTP**	0.89	1.25	0.125	1

nation that is given is that transcription likely initiates RCA, since the phi29 DNAP can utilize RNA as a primer. This scenario then resembles the *in vivo* mechanism of the T7 DNA replication initiation, where the T7 DNAP likewise utilizes the RNA primer synthesized and replaces the T7 RNAP at the initiation site. Once one such event happens, RCA is sustainable since the random hexamers can bind to the displaced strand, functioning as new primer:template junctions for the phi29 DNAP. However, even more surprising was that in the abscence of random hexamers DNA synthesis was also observed, an observation inexplicable at this moment[31]. An interesting finding of this paper is that optimizing the concentrations of rNTPs, dNTPs, tRNAs and T7 RNAP in the PURE system led to an increase in DNA replication by approximately 100-fold compared to the original conditions. The final yield of DNA amplification in optimized conditions was 10-fold after 12 hours when the reactions were incubated at 30 °C [67]. The main optimization came from decreasing the ratio of total rNTP/dNTP around 25-fold leading to dNTPs being in excess with a 0.7 final ratio. Similarly, in *in vitro* DNA replication of the phi29 genome by the phi29 DNAP and the terminal protein, DNA replication decreased 3-fold upon the addition of 0.5 mM rNTPs to the solution[32] [50]. The slowing-down of DNA polymerization is either caused by the extra time it takes to discriminate between the correct base and sugar when the pool of nucleotides is much higher, or due to the proofreading activity of the DNAP that slows down the overall incorporation rate[33] How do these ratios of ribonucleotides and deoxyribonucleotides compare with bacterial cells? In Tabel 4.1 a comparison is shown between the nucleotide pools of bacterial cells [3], the initial and optimized PURE system composition [67], and the PURE system compositions of the PURE*frex* utilized in this thesis [7]. Optimiziation of the nucleotides in the experimental system is important for future research when DNA replication, consuming dNTPs as substrates, is combined with transcription and translation, the latter processes utilizing both rNTPs as substrates. Finding a delicate balance will be a challenging task, especially with increasing complexity of processes that the minimal gene expression system has to perform towards the realization of a minimal cell.

[31] I believe the following could explain it: The purified plasmid DNA is mainly in the covalently circular closed form, but a small fraction is also nicked, as suggested in [23]. Note that second strand synthesis cannot be initiated by the produced RNA, but perhaps by non-specific binding of tRNA to the displaced strands, or by template switching of the DNAP as similarly shown for RCA by the phi 29 DNAP on circular plasmids [68].

[32] Final ratio rNTPs/dNTPs=6.

[33] The effect of adding more rNTPs to the pool of dNTPs results probably also in more base-mismatches. For some polymerases it is known that proofreading of an incorrect sugar is less efficient than proofreading an incorrect base [69] and 1 rNMP is incorporated per about 1000 dNMPs.

QBETA REPLICASE: EVOLUTION EXPERIMENTS

Over the past decades, the group of Prof. Yomo has also investigated in depth the expression and replication dynamics of an RNA dependent RNA polymerase (RdRP), the Qbeta replicase, in the PURE system amongst other cell-free expression systems. The Qbeta is a small positive ssRNA virus encoding for only four genes that infects *E. coli*. From the plus RNA the complementary minus strand RNA is synthesized and serves as a template for translation. The translated replicase subunit beta needs tree host factors to perform replication of its genome, namely ribosomal protein S1 and the elongation factors EF-TU and EF-TS [70]. The Qbeta replicase recognizes its own plus and minus RNA by a terminal loop structure, and although the host RNA is generally not copied, there is a variety of small non genomic RNA molecules that can serve as template. These self-replicating RNAs are referred to as RQ RNA and originate from several sources. One way to repress the amplification of RQ RNA is to compartmentalize reactions, where the compartments that are free from RQ RNA can undergo RNA genome amplification without inhibitory effects [71]. The idea of compartmentalizing Qbeta replication reactions in water-oil emulsion droplets has been further developed to study evolution of the Qbeta replicase sequence in gene expression environments. Studying the Qbeta replicase in a gene expression system is especially interesting because of the direct (feedback) coupling with transcription and translation: any mutations introduced during the replication of RNA by the Qbeta replicase has its effect on the newly translated Qbeta replicase from that RNA molecule. Improvement in RNA replication can thus be a consequence of a difference in RNA secondary structure, or from a better interaction with the altered protein structure resulting from sense mutations. By applying multiple manual fusion and division cycles of emulsion droplets containing the RNA gene for the Qbeta replicase and the PURE system, the RNA genes spontaneously evolved according to Darwinian principles and overcame the parasitic RNA through evolution [72]. With the same approach, the group of Yomo tried to answer the question: Is it possible for a simple self-replication system to adapt to various environments? The different environments were created by leaving out several ribosome factors that would either impair initiation (IF1, IF2, IF3, MTF) or termination (RF1, RF2, RF3, RRF), or both (low ribosome levels). They found that for all environments the replication efficiency of the artificial RNA gene increased with an increasing number of rounds, showing that a simple system can adapt to different environments [73].

This research is highly inspiring for implementing evolution experiments involving DNA replication proteins, even though we do not consider the RdRPs as a candidate for genome replication in the minimal cell.

4.3.3. DNA REPLICATION CANDIDATES INVESTIGATED IN OUR LAB

In this thesis, we will describe the investigations into several DNA replication candidates for the minimal cell.

Our first, naive, thought was to investigate if DNA replication using mRNA as an intermediate can be achieved similar to the retro-viral replication. As the reverse transcriptase would be the only extra enzyme needed[34] to turn information from DNA to RNA to

[34] Provided the RNAP is already a component of the transcription-translation system. Apart from the RT, retroviruses use host-factors such as enzymes with RNaseH function, ligases or DNAPs for genome replication.

DNA, this systems seems truly minimal. Our main investigations consisted in trying to achieve the self sustained sequence amplification cycle as depicted in Figure 4.9 for a 1-gene construct [35]. We have realised that processing mRNA for the reverse transcription reaction and for translation by the ribosome might cause detrimental interference, and therefore propose a new way to couple such DNA replication cycle in a translation environment. When analyzing the intermediate products of the replication cycle, we stumbled upon a previously unknown property of the T7 RNAP in *in vitro* transcription reactions, namely the capability to transcribe RNA:DNA hybrid templates[36].
The findings on the **self sustained sequence amplification cycle** and the **transcription by the T7 RNAP of hybrid DNA:RNA templates** are discussed in Chapter 5.

Secondly, we are interested in the system of the **T7** virus, as it encodes the minimal **replisome** found in nature so far. Literature contains a wealth of information from its genome organization to the dynamics of replication processes of the replisome, which makes it an exciting system to study [27, 74, 75]. Our first goal comprises the expression of the replisome proteins inside the PURE system and confirming their activities in comparison to purified proteins. These findings will not be reported in this thesis.

Finally, we propose the **phi29 DNA replication** system as a candidate for the minimal cell in Chapter 6 where we investigate if the DNA replication system based on the four phi29 proteins is compatible with the PURE system, our biosynthesis platform for the minimal cell. **Our aim is to copy a linear dsDNA template that encodes for its own replication proteins, completing a full round of the central dogma of molecular biology**.

[35] Literature does not report the cycle for over 250 nt, as it is mainly used as a detection tool for pathogens in the medical field.
[36] The promoter also being a complete DNA:RNA hybrid.

FURTHER CONSIDERATIONS

There are several options or considerations that are not discussed yet in this chapter. The first, most obvious one is:
• *Should the DNA genome have a linear or a circular form?*
We would like to put this question forward, and would need the input of biologists that are knowledgeable in DNA replication mechanisms to have a meaningful discussion on this topic. All the proposed 'minimal cells' or 'minimal-gene sets' so far discuss a circular genome [9, 59, 76]. We can think of the following three main advantages choosing a circular genome:
- circulization solves the termination paradox
- many life forms have circular plasmids and can serve as an example: we have mentioned the mechanism of replicating bacterial chromosomes and plasmids in *E. coli*, but there are others that are interesting as well, such as the single strand displacement mechanism in mitochrondia or different DNA replication systems involving the Rep protein.
To what extend are these really advantages compared to a linear genome? We have discussed several mechanisms that solve the termination paradox for linear DNA molecules, such as non-covalent circulization in intermediate stages of replication, protein-priming or concatenation. Other mechanisms that we have not discussed but are abundantly found in nature are the introduction of telomeres in eukaryotes, or hairpin formation at the 3'-ends of linear plasmids. One should realise that some mechanisms applied by viral replication are not always transferable to the concept of a minimal cell: the packaging of more than 1 full-length T7 genome in its head is not translatable to packaging concatamers to daughter liposomes. The most elegant way to solve the initiation and termination paradox, killing two birds with one stone, is the protein-priming mechanism. This mechanism is mainly known from the phi29 virus and the human adenovirus, but it is also found in yeast, funghi and higher plants [77]. The disadvantage of this mechanism is the generation of large amounts of ssDNA, which is especially vulnerable *in vivo* in the presence of DNases. The adenovirus has ensured that even displaced ssDNA of its full genome length of 40 knt can serve as a template for replication, by an inverted terminal repeat of 100 bp at its genome ends: the ends of the ssDNA hybridize creating a double stranded handle that serves as a new initiation site. Even though the generation of ssDNA can be problematic due to the formation of impeding secondary structures, we have at least one natural example of a genome length of 40 kb (twice the size of the phi29 genome) where this mechanism proves it success. Could the protein-priming mechanism also be viable for even larger genomes, around 100-150 kb considering the context of the minimal cell? We do not have an answer to this question at the moment, but would like to make the following remarks. The 4-gene set concept for double stranded DNA replication by the phi29 proteins is definitely the minimum number of genes one can find for a minimal DNA replication system. However, although it is composed of a minimal set of genes, it is not clear if this DNA replication is the most minimal one in terms of energy costs. The single-stranded and double-stranded DNA binding proteins (p5 and p6) are absolutely necessary to achieve efficient DNA replication [78], and the function of SSB is not limited to boosting the efficiency but also to reduce template switching activities of the DNAP causing the formation of sideproducts (cite something). In other words, the longer the DNA template to be replicated, the more SSB (and DSB) proteins

are necessary to ensure faithful full-length replication[37]. The amount of produced proteins necessary for this symmetric mode of replication thus increases more or less linear with the length of the DNA genome, in contrast to the asymmetric mode of replication where a replication fork bubble is in the order of 1 kb in length, processing leading and lagging strand synthesis simultaneously. Though the latter needs more genes encoding for proteins, the absolute amount of proteins necessary to perform faithful replication is less. Perhaps a combination of the two mechanisms, as is seen in linear plasmid replication in *Streptomyces* is an idea. *Streptomyces coelicolor* contains a 350-kb linear plasmid and internal origins from which DNA replication occurs in a bi-directional manner [25, 77]. The unreplicated part of the lagging strand has an inverted repeat that forms a secondary structure, allowing access to the terminal protein to prime DNA replication to fill in the missing DNA strand.

Other considerations are: How does DNA replication influence other processes in the minimal cell? How to optimize conditions for DNA replication? What happens when the DNA is simultaneously processed by replication and transcription enzymes and how detrimental are collisions? How to couple DNA replication with growth and division of liposomes?
We have already slightly touched upon the optimization of DNA replication dependent on the balance of nucleotide pools in this chapter, and will further elucidate this topic and the other questions in Chapters 6 and 7.

[37] the concentration of p6 proteins *in vivo* phi-29 infected bacillus cells is even in the mM range [79]

REFERENCES

[1] E. V. Koonin, *How Many Genes Can Make a Cell: The Minimal-Gene-Set Concept 1,* Annual review of genomics and human genetics **1**, 99 (2000).

[2] A. M. Poole and D. T. Logan, *Modern mRNA proofreading and repair: clues that the last universal common ancestor possessed an RNA genome?* Molecular biology and evolution **22**, 1444 (2005).

[3] A. Kornberg and T. A. Baker, *DNA replication*, 2nd ed. (Wh Freeman San Franscisco, 1992).

[4] U. Hübscher, *DNA polymerases: discovery, characterization and functions in cellular DNA transactions* (World Scientific, 2010).

[5] S. D. DePamphilis, Melvin L.; Bell, *Genome duplication* (Garland Science, 2011).

[6] S. Kamtekar, A. J. Berman, J. Wang, J. M. Lázaro, M. De Vega, L. Blanco, M. Salas, and T. a. Steitz, *Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage ϕ29,* Molecular Cell **16**, 609 (2004).

[7] Y. Shimizu, a. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda, *Cell-free translation reconstituted with purified components.* Nature biotechnology **19**, 751 (2001).

[8] M. L. DePamphilis and S. D. Bell, *Replication forks,* in *Genome duplication* (Garland Science, 2011).

[9] A. Martinez-antonio, L. Espindola-serna, and C. Quiñones-valles, *Proposal for a Minimal DNA Auto-Replicative System,* The Mechanisms of DNA Replication , 127 (2013).

[10] A. R. Mushegian and E. V. Koonin, *A minimal gene set for cellular life derived by comparison of complete bacterial genomes.* Proceedings of the National Academy of Sciences **93**, 10268 (1996).

[11] A. van Oijen and J. Loparo, *Single-molecule studies of the replisome,* Annual review of biophysics **39**, 429 (2010).

[12] R. Singleton and D. R. Singleton, *Remembering Our Forebears: Albert Jan Kluyver and the Unity of Life,* Journal of the History of Biology , 1 (2016).

[13] P. Forterre, J. Filée, and H. Myllykallio, *Origin and Evolution of DNA and DNA Replication Machineries,* The Genetic Code and the Origin of Life , 145 (2004).

[14] P. Forterre, *Why are there so many diverse replication machineries?* Journal of Molecular Biology **425**, 4714 (2013).

[15] A. M. Poole, N. Horinouchi, R. J. Catchpole, D. Si, M. Hibi, K. Tanaka, and J. Ogawa, *The Case for an Early Biological Origin of DNA,* Journal of Molecular Evolution **79**, 204 (2014).

[16] D. Raoult and P. Forterre, *Redefining viruses: lessons from Mimivirus,* Nat.Rev.Microbiol. **6**, 315 (2008).

[17] P. Forterre, *Defining Life: The Virus Viewpoint,* Origins of Life and Evolution of Biospheres **40**, 151 (2010).

[18] M. L. DePamphilis and S. D. Bell, *Replicons,* in *Genome duplication* (Garland Science, 2011) pp. 169–189.

[19] D. H. Duckworth, *"Who Discovered Bacteriophage?".* Bacteriological reviews **40**, 739 (1976).

[20] M. Breitbart, *Marine Viruses: Truth or Dare,* Annual Review of Marine Science **4**, 425 (2012).

[21] M. Oke, M. Kerou, H. Liu, X. Peng, R. a. Garrett, D. Prangishvili, J. H. Naismith, and M. F. White, *A dimeric Rep protein initiates replication of a linear archaeal virus genome: implications for the Rep mechanism and viral replication.* Journal of virology **85**, 925 (2011).

[22] R. Hausmann, *The genetics of T-odd phages,* Annual Reviews Microbiology **27**, 51 (1973).

[23] Y. Schaerli, V. Stein, M. M. Spiering, S. J. Benkovic, C. Abell, and F. Hollfelder, *Isothermal DNA amplification using the T4 replisome: circular nicking endonuclease-dependent amplification and primase-based whole-genome amplification,* Nucleic Acids Research **38**, e201 (2010).

[24] K. N. Kreuzer and J. R. Brister, *Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives.* Virology journal **7**, 358 (2010).

[25] J. W. Dale and S. F. Park, *Molecular Genetics of Bacteria,* 5th ed. (Wiley-Blackwell, 2010) pp. 122–123.

[26] S.-J. Lee and C. C. Richardson, *Choreography of bacteriophage T7 DNA replication.* Current Opinion in Chemical Biology **15**, 580 (2011).

[27] J. J. Dunn and F. W. Studier, *Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements.* Journal of molecular biology **166**, 477 (1983).

[28] S. W. Matson and C. C. Richardson, *DNA-dependent Nucleoside 5'-Triphosphatase Activity of the Gene 4 Protein of Bacteriophage T7*,* The Journal of biological chemistry **258**, 14009 (1983).

[29] C. W. Fuller and C. Richardson, *Initiation of DNA Replication at the Primary Origin of Bacteriophage T7 by Purified Proteins,* The Journal of biological chemistry **260**, 3197 (1985).

[30] I. Holguera, D. Muñoz-Espín, and M. Salas, *Dissecting the role of the phi29 terminal protein DNA binding residues in viral DNA replication.* Nucleic acids research , 1 (2015).

[31] L. Blanco and M. Salas, *Replication of phage phi 29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length phi 29 DNA.* Proceedings of the National Academy of Sciences **82**, 6404 (1985).

[32] J. Mendez, L. Blanco, J. A. Esteban, A. Bernad, and M. Salas, *Initiation of phi 29 DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication.* Proceedings of the National Academy of Sciences **89**, 9579 (1992).

[33] J. Méndez, L. Blanco, and M. Salas, *Protein-primed DNA replication: A transition between two modes of priming by a unique DNA polymerase,* EMBO Journal **16**, 2519 (1997).

[34] M. S. Soengas, C. Gutiérrez, and M. Salas, *Helix-destabilizing activity of phi 29 single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication.* Journal of molecular biology **253**, 517 (1995).

[35] M. Mencía, P. Gella, A. Camacho, M. de Vega, and M. Salas, *Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage Phi29.* Proceedings of the National Academy of Sciences of the United States of America **108**, 18655 (2011).

[36] S. A. Khan, *Plasmid rolling-circle replication: Highlights of two decades of research,* (2005).

[37] C. Hulo, E. De Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios, and P. Le Mercier, *ViralZone: A knowledge resource to understand virus diversity,* Nucleic Acids Research **39** (2011), 10.1093/nar/gkq901.

[38] S. Litvak, *Reverse Transcriptase and Retroviral Replication,* Encycolpedia of Biological Chemistry, Volume 3 , 708 (2004).

[39] C. Hulo, E. De Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios, and P. Le Mercier, *ViralZone: A knowledge resource to understand virus diversity,* Nucleic Acids Research **39** (2011), 10.1093/nar/gkq901.

[40] J. L. Ong, D. Loakes, S. Jaroslawski, K. Too, and P. Holliger, *Directed Evolution of DNA Polymerase, RNA Polymerase and Reverse Transcriptase Activity in a Single Polypeptide,* Journal of Molecular Biology **361**, 537 (2006).

[41] R. Sousa and R. Padilla, *A mutant T7 RNA polymerase as a DNA polymerase.* The EMBO journal **14**, 4609 (1995).

[42] N. Arnaud-Barbe, V. Cheynet-Sauvion, G. Oriol, B. Mandrand, and F. Mallet, *Transcription of RNA templates by T7 RNA polymerase.* Nucleic acids research **26**, 3550 (1998).

[43] G. M. Cheetham, *Structure of a Transcribing T7 RNA Polymerase Initiation Complex,* Science **286**, 2305 (1999).

[44] K. E. McGinness and G. F. Joyce, *Substitution of ribonucleotides in the T7 RNA polymerase promoter element,* Journal of Biological Chemistry **277**, 2987 (2002).

[45] K. Mullis and F. Faloona, *Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction,* Methods in Enzymology **155**, 335 (1987).

[46] P. Gill and A. Ghaemi, *Nucleic acid isothermal amplification technologies: a review.* Nucleosides, nucleotides & nucleic acids **27**, 224 (2008).

[47] J. C. Guatelli, K. M. Whitfield, D. Y. Kwoh, K. J. Barringer, D. D. Richman, and T. R. Gingeras, *Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication.* Proceedings of the National Academy of Sciences of the United States of America **87**, 7797 (1990).

[48] B. Deiman, P. van Aarle, and P. Sillekens, *Characteristics and applications of nucleic acid sequence-based amplification (NASBA).* Molecular biotechnology **20**, 163 (2002).

[49] F. B. Dean, J. R. Nelson, T. L. Giesler, and R. S. Lasken, *Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification.* Genome research **11**, 1095 (2001).

[50] L. Blanco, A. Bernads, J. M. Lazaro, G. Martin, C. Garmendia, and M. Salas, *Highly Efficient DNA Synthesis by the Phage phi 29 DNA Polymerase,* The Journal of biological chemistry **264**, 8935 (1989).

[51] G. Kumar and G. Chernaya, *Cell-free protein synthesis using multiply-primed rolling circle amplification products.* BioTechniques **47**, 637 (2009).

[52] G. T. Walker, M. C. Little, J. G. Nadeau, and D. D. Shank, *Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system.* Proceedings of the National Academy of Sciences of the United States of America **89**, 392 (1992).

[53] G. T. Walker, M. S. Fraiser, J. L. Schram, M. C. Little, J. G. Nadeau, and D. P. Malinowski, *Strand displacement amplification–an isothermal, in vitro DNA amplification technique.* Nucleic acids research **20**, 1691 (1992).

[54] M. Vincent, Y. Xu, and H. Kong, *Helicase-dependent isothermal DNA amplification.* EMBO reports **5**, 795 (2004).

[55] A. Motré, Y. Li, and H. Kong, *Enhancing helicase-dependent amplification by fusing the helicase with the DNA polymerase,* Gene **420**, 17 (2008).

[56] Y. J. Jeong, K. Park, and D. E. Kim, *Isothermal DNA amplification in vitro: the helicase-dependent amplification system.* Cellular and molecular life sciences : CMLS **66**, 3325 (2009).

[57] C. Garmendia, A. Bernad, J. A. Esteban, L. Blanco, and M. Salas, *The bacteriophage phi 29 DNA polymerase, a proofreading enzyme.* The Journal of biological chemistry **267**, 2594 (1992).

[58] Y. Li, H. J. Kim, C. Zheng, W. H. A. Chow, J. Lim, B. Keenan, X. Pan, B. Lemieux, and H. Kong, *Primase-based whole genome amplification,* Nucleic Acids Research **36**, 1 (2008).

[59] A. C. Forster and G. M. Church, *Towards synthesis of a minimal cell.* Molecular systems biology **2**, 45 (2006).

[60] B. Sauer, *Cre/lox: one more step in the taming of the genome.* Endocrine **19**, 221 (2002).

4

[61] J. Shin, P. Jardine, and V. Noireaux, *Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction,* ACS synthetic biology , 3 (2012).

[62] H. Asahara and S. Chong, *In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme.* Nucleic acids research **38**, e141 (2010).

[63] Y. Katano, T. Hisayoshi, I. Kuze, H. Okano, M. Ito, K. Nishigaki, T. Takita, and K. Yasukawa, *Expression of moloney murine leukemia virus reverse transcriptase in a cell-free protein expression system,* Biotechnology Letters **38**, 1 (2016).

[64] T. Niwa, B.-W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, and H. Taguchi, *Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins.* Proceedings of the National Academy of Sciences of the United States of America **106**, 4201 (2009).

[65] K. Fujiwara, T. Katayama, and S.-I. M. Nomura, *Cooperative working of bacterial chromosome replication proteins generated by a reconstituted protein expression system.* Nucleic acids research , 1 (2013).

[66] Z. Tsuchihashi, *Translational frameshifting in the Escherichia coli dnaX gene in vitro.* Nucleic acids research **19**, 2457 (1991).

[67] Y. Sakatani, N. Ichihashi, Y. Kazuta, and T. Yomo, *A transcription and translation-coupled DNA replication system using rolling-circle replication,* Scientific Reports **5**, 10404 (2015).

[68] C. Ducani, G. Bernardinelli, and B. Hogberg, *Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA,* Nucleic Acids Research , 1 (2014).

[69] J. S. Williams, A. R. Clausen, S. A. Nick McElhinny, B. E. Watts, E. Johansson, and T. A. Kunkel, *Proofreading of ribonucleotides inserted into DNA by yeast DNA polymerase epsilon,* DNA Repair **11**, 649 (2012).

[70] T. Blumenthal and G. G. Carmichael, *Rna replication: function and structure of the Qbeta replicase,* Annual review of biochemistry **48**, 525 (1979).

[71] H. Urabe, N. Ichihashi, T. Matsuura, K. Hosoda, Y. Kazuta, H. Kita, and T. Yomo, *Compartmentalization in a water-in-oil emulsion repressed the spontaneous amplification of RNA by Q?? replicase,* Biochemistry **49**, 1809 (2010).

[72] N. Ichihashi, K. Usui, Y. Kazuta, T. Sunami, T. Matsuura, and T. Yomo, *Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment.* Nature communications **4**, 2494 (2013).

[73] R. Mizuuchi, T. Yomo, and N. Ichihashi, *Adaptation and diversification of an RNA replication system under initiation- or termination-impaired translational conditions,* ChemBioChem **0871**, n/a (2016).

[74] L. Y. Chan, S. Kosuri, and D. Endy, *Refactoring bacteriophage T7.* Molecular systems biology **1**, 2005.0018 (2005).

[75] S. M. Hamdan and C. C. Richardson, *Motors, switches, and contacts in the replisome.* Annual review of biochemistry **78**, 205 (2009).

[76] R. Gil, F. J. Silva, J. Peretó, and A. Moya, *Determination of the core of a minimal bacterial gene set.* Microbiology and molecular biology reviews : MMBR **68**, 518 (2004).

[77] M. Salas, *Protein-priming of DNA replication,* Annual review of biochemistry **60**, 39 (1991).

[78] L. Blanco, J. M. Lazaro, M. de Vega, A. Bonnin, and M. Salas, *Terminal protein-primed DNA amplification.* Proceedings of the National Academy of Sciences **91**, 12198 (1994).

[79] A. M. Abril, M. Salas, J. M. Andreu, J. M. Hermoso, and G. Rivas, *Phage phi29 Protein p6 Is in a Monomer-Dimer Equilibrium That Shifts to Higher Association States at the Millimolar Concentrations Found in Vivo,* Biochemistry **36**, 11901 (1997).

5

DNA REPLICATION VIA RNA INTERMEDIATES AND TRANSCRIPTION-TRANSLATION OF HYBRID DNA:RNA TEMPLATES

In this study, we investigated how a nucleic acid amplification cycle (based on 3SR) could be integrated into a gene expression environment in order to combine DNA replication with transcription and translation processes. We confirmed the individual steps of the nucleic acid amplification cycle that has RNA as an input template and can potentially amplify DNA and RNA molecules in a cyclic manner, for a one-gene construct of ca. 1 kb in length. The properties of the intermediate products were investigated and it was found that DNA:RNA hybrid templates can be transcribed by the viral T7 RNAP. The hybrid templates are transcribed by the T7 RNAP with reduced efficiency depending on the nature of the template strand, in the order dsDNA > hybrid DNA_{coding}:$RNA_{template}$ > hybrid RNA_{coding}:$DNA_{template}$. The closely related viral SP6 RNAP was also able to transcribe hybrid templates of both orientation. More aborted products were generated upon transcription from hybrid templates, although this did not prevent the translation of full-length protein in a minimal gene expression system starting from hybrid templates.

Viruses have solved the challenge of copying genetic information in the most simple, efficient and elegant ways. While identifying DNA replication candidates for the minimal cell, we chose to study a few viral replication strategies in more detail. In this chapter, we investigate the idea of utilizing RNA as an intermediate information carrier for reconstructing genes and expressing them in a minimal gene expression system.

5.1. INTRODUCTION

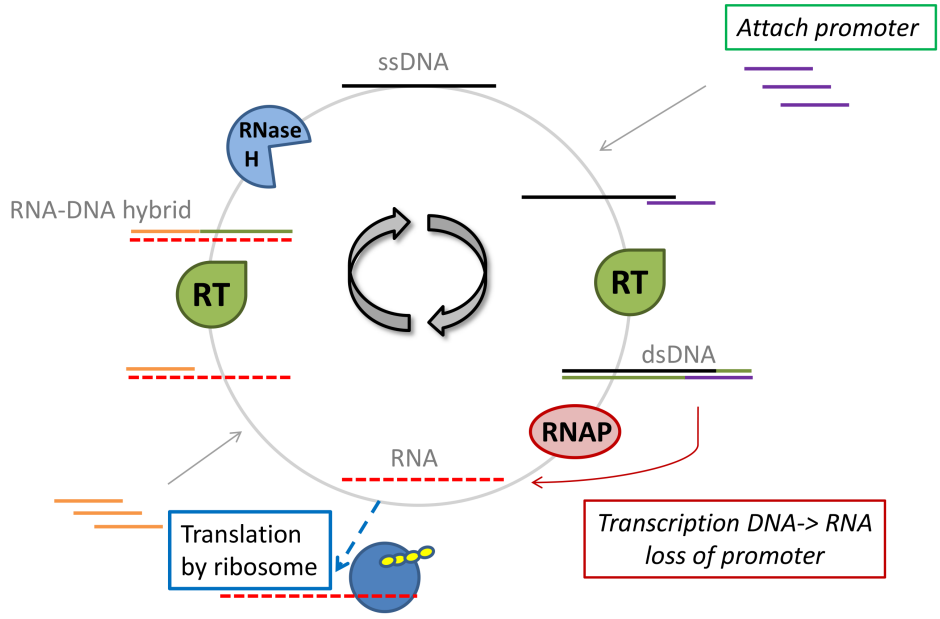


Figure 5.1: Scheme for self-sustained sequence amplification, starting with an RNA sequence that can be translated by the ribosome (blue). For details see text. Color coding: RNA = red dashed line, DNA primer that binds to the 3'-end of the RNA = orange line, synthesized DNA by the RT in the previous step = green line, primer containing promoter sequence that binds to the 3'-cDNA = purple line. The RNA and DNA molecules are not in scale.

In Chapter 4 was discussed how retroviruses use RNA as their carrier of genetic information. The viral RNA is converted into dsDNA in the host cell. This process is performed by a single enzyme, the reverse transcriptase (RT). The RT has surprisingly complex functions and possesses the three catalytic activities: (1) an RNA-dependent DNA polymerase activity, (2) a DNA-dependent DNA polymerase activity and (3) an RNase H function which can degrade RNA from the intermediate DNA:RNA hybrid template. Based on these versatile properties of the RT, several *in vitro* nucleic acid replication methods were developed [1–5]. Most of these methods are aimed at the amplification and subsequent detection of RNA sequences, always with a length of less than 250 nucleotides. The isothermal self-sustained sequence replication, or 3SR cycle, mim-

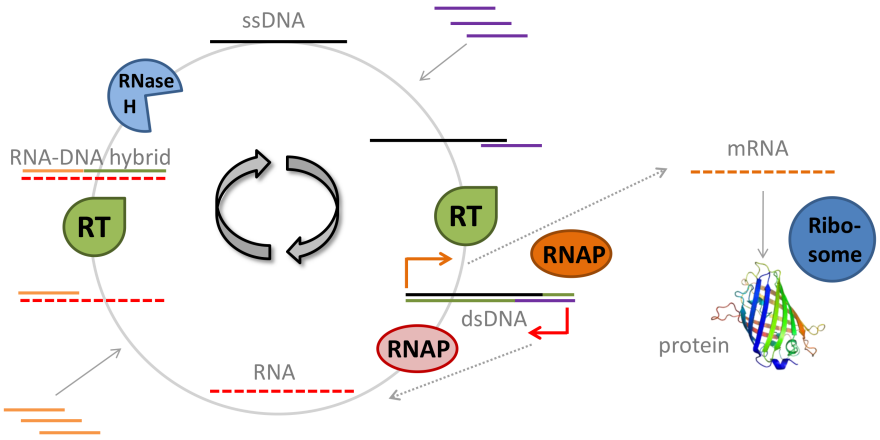


Figure 5.2: Schematic of the proposed amplification cycle of nucleic acids in a transcription-translation environment, decoupling DNA amplification and translation reactions. The self-sustained sequence amplification is presented in the left cycle. The DNA molecule now contains two promoters, and RNA molecules are transcribed by orthogonal RNAPs in opposite directions. Color coding is the same as in Figure 5.1.

ics the retro-viral strategy of RNA replication by using RNA as a template to create DNA by the combined enzymatic activities of a reverse transcriptase, RNase H and a DNA-dependent RNA polymerase [1]. While the main motivation to develop such an amplification scheme was to amplify RNA sequences, our interest is to amplify DNA sequences combined with subsequent transcription and translation reactions. Adapting the 3SR scheme for amplification of a single gene, where RNA is used as an intermediate to generate a new dsDNA gene, results in Figure 5.1. This scheme starts with transcription from the promoter on the initial dsDNA template by an RNAP into mRNA. The mRNA is both a template for translation by the ribosome and cofactors, and an intermediate species in the nucleic acid amplification cycle. A complementary DNA oligonucleotide hybridizes to the 3'-end of the mRNA, and subsequently the RT elongates the primer synthesizing the complementary DNA (cDNA) strand resulting in a hybrid RNA-DNA template. The RNA strand is digested by the RNase H function of the RT and / or by an additional RNase H enzyme, releasing the cDNA as a ssDNA template. The promoter sequence, the information part lost upon transcription of the DNA, is re-introduced by the second oligo that hybridizes to this ssDNA. As a final step, the reverse transcriptase completes the cycle by elongating the primer in both directions recovering the initial double strand DNA without loss of sequence information.

In this scheme, the mRNA is utilized for both the translation and the reverse transcription reactions, which might lead to competition of the two processes on the same template. To avoid interference we propose to extend the amplification scheme such that DNA replication and gene expression reactions are decoupled. Decoupling the amplification cycle from translation ensures that these processes can be regulated and tuned separately. Moreover, it enables the amplification of a template containing multiple

genes. To achieve this, the mRNA that serves as a template for translation and the RNA that enters the amplification cycle should be transcribed by different RNAPs and in opposite directions (Figure 5.2). This is in concordance with the reported finding that two single subunit RNA polymerases approaching each other in a head-on collision on the same DNA may pass by one another [6]. Hence, translation will only occur from mRNA coding genes, that at the same time have no binding sequence for the DNA oligos used in the amplification cycle.

This chapter describes the *proof of principle* experiment that shows that one round of the nucleic acid amplification cycle, starting from an RNA template, could be completed for a single gene (a derivative of the YFP-Spinach template, around 1100 bp). We then describe the design of a template containing a single gene that could potentially be amplified and translated according to Figure 5.2. While conducting these experiments, we discovered that the T7 RNAP was capable of transcribing the DNA-RNA hybrid intermediate, a surprising finding. This is of primary importance for the viability of the cycle as a transcriptionally active hybrid DNA:RNA template would compete for the pool of RNAPs and NTPs. Examples have been reported in literature, where the T7 RNAP could initiate transcription on a T7 promoter containing a fraction of ribonucletoides [7]. However, transcription from T7 promoter sequences where the template or coding strand fully consists of ribonucleotides have not been studied. Therefore, we explored this phenomenon further by investigating the transcription efficiency of hybrid RNA:DNA and DNA:RNA templates by the viral T7 RNAP and SP6 RNAP in a transcription-translation environment, which represents the main part of this chapter.

5.2. RESULTS OF NUCLEIC ACID AMPLIFICATION CYCLE

In this research, two different commercially available reverse transcriptase enzymes were utilized, namely the well characterized RT from the Avian Myeloblastosis Virus (AMV[1]) and the engineered RNase H-deficient AccuScript enzyme (derivative from the RT of Moloney murine leukemia virus[2]). We first confirmed the DNA-dependent DNA polymerase activity of these enzymes by a performing a primer elongation assay (step 1 in Figure 5.3A). A ssDNA lacking a promoter sequence hybridizes partly to a 35-nt oligonucleotide containing an SP6 promoter sequence in its overhang. The elongation of the 3'-end of the template and of the primer leads to a dsDNA on which transcription by SP6 RNAP (step 2 in Figure 5.3A) can be initiated. The observed conversion of ssDNA to dsDNA by the RT and subsequent transcription by the SP6 RNAP corroborates the activities of both enzymes in our assay (Figure 5.3B). Secondly, we tested if the mRNA, transcribed by the SP6 RNAP from the SP6p-T7p-YFP-Spinach-polyA template, served as a substrate for the RT reaction (step 3 in Figure 5.3A). After hybridization of a poly-dT primer, the RT elongates the primer with its RNA-dependent DNA polymerase activity. No mRNA remains at the end of the experiment and conversion to full-length hybrid DNA:RNA templates is achieved by both enzymes (Figure 5.3C).

[1] AMV is a (+) ssRNA virus with a 7.2-kb genome that produces leukemia in birds.
[2] MMLV produces leukemia in mice.

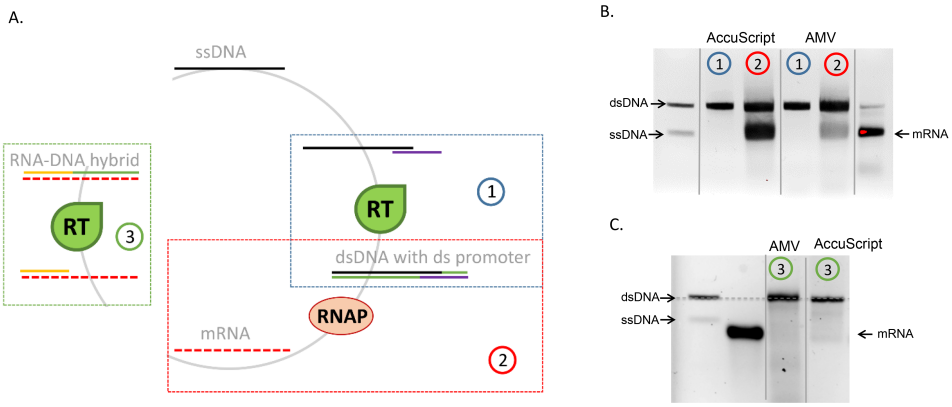


Figure 5.3: Individual steps of the amplification cycle are confirmed to assay the activity of the RTs and suitability of the YFP-Spinach template of interest. A) Reaction scheme highlighting the first three steps. B) Steps 1 and 2 are confirmed consecutively on nucleic acid gel and C) step 3 separately for the AMV and AccuScript reverse transcriptases.

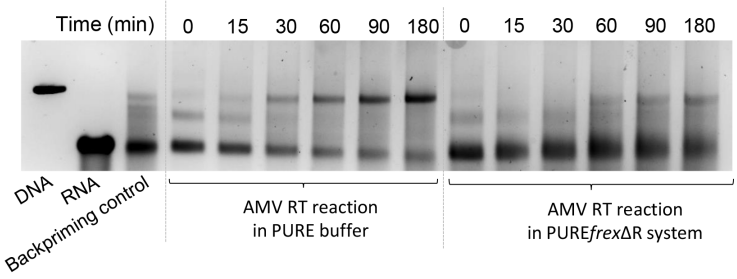


Figure 5.4: The reverse transcription reaction by the AMV RT generates two times more product after 3 h at 37 °C in the PURE buffer in comparison with the PURE system (without ribosomes). The input RNA template is purified from a transcription reaction by the SP6 RNAP on the SP6p-T7p-mYFP-LL-Spinach-polyA template. The backpriming control is a reverse transcription reaction by the AMV RT in the absence of the (dT) primer, where a small fraction (15% after 3 h compared to PURE buffer reaction) of product is formed.

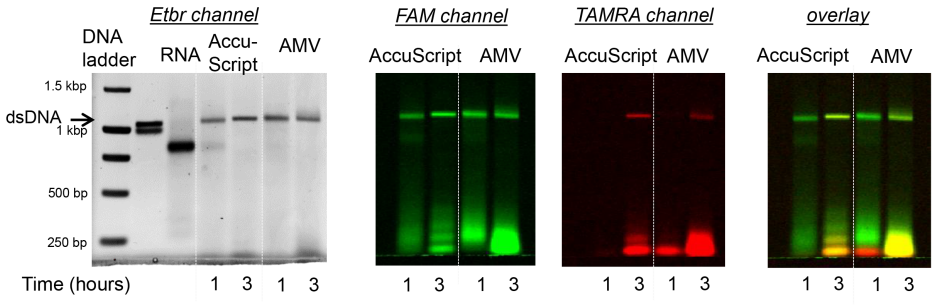


Figure 5.5: The nucleic acid amplification cycle starting from mRNA till the dsDNA template achieved by the RT and RNase H enzymes is analyzed on an 1.1% agarose gel. Incorporation of the primers is monitored by their fluorescent labels. The first (dT) primer is labeled at its 5'-end with FAM and the second (promoter-containing) primer is 5'-end labeled with a TAMRA dye. RNase H is added after 1h to the reaction and its activity proves to be necessary for incorporation of the second primer into the dsDNA template.

5

We then asked whether the reverse transcription reaction would be efficient in the PURE system environment, where binding of translation proteins to the mRNA might compete with the RT activity. We compared the formation of reverse transcription product generated by the AMV RT in the PURE buffer and in the PURE*frex*ΔR (PURE*frex* lacking ribosomes). It was found that the amount of dsDNA detected after 3 h of reaction is a factor two less in the PURE system (Figure 5.4), indicating interference with some PURE components.

Next, we completed the steps of DNA amplification cycle on the 1.1-kb SP6p-mYFPgo-LL-Spinach-polyA template with the enzymes AMV RT or AccuScript, RNase H and SP6 RNAP in a sequential manner in the PURE buffer (Figure 5.5). To demonstrate that the final product is indeed a full-length dsDNA template, we monitored the completion of the steps with two fluorescent dyes attached to the 5'-ends of the primers. Starting from the RNA template YFP-Spinach-polyA with hybridized dT primer (Figure 5.2, the RT reaction was incubated for 1 h before RNase H was supplemented to the reaction solution to degrade the RNA:DNA hybrid. Samples before and 2 h after addition of the RNase H were analyzed on gel. After 1 h, almost all RNA is converted to a reverse transcription product by the AMV and AccuScript enzyme, visible on the EtBr-stained gel at the same height as dsDNA in Figure 5.5. At this stage, only the first (dT) primer, labeled with a FAM dye, is present in the reverse transcription product. The second primer, labeled with a TAMRA dye, incorporates into the reverse transcription product only after the reaction is supplemented with RNase H. This indicates that the RNase H function of the AMV RT alone is not sufficient to degrade the full RNA:DNA hybrid and the cycle can only be completed with an accessory RNase H enzyme [1]. This *proof of principle* experiment shows that the nucleic acid amplification cycle, starting from an RNA template and achieving dsDNA as final product, could be performed on a gene-coding construct of length around 1 kb.

5.3. DISCUSSION OF NUCLEIC ACID AMPLIFICATION CYCLE

Although we showed that the steps of the nucleic acid amplification cycle work in a consecutive manner, we have not achieved the full cycle in a one-pot reaction yet (SI Figure 5.13). We suspect the following mechanisms to hinder efficiency: RNA secondary structures, self-priming of reverse transcriptase [8], formation of small undesirable byproducts [9], limited lifetime of the enzymes [3], degradation of RNA by RNase H upon primer hybridization [1] and non-specific binding of PURE system components that could reduce the reaction rates. Some of the mentioned challenges were observed in our experiments. In Figure 5.5 the formation of small (around 200 bp) sideproducts may cause an inefficient reaction. In Figure 5.4, a small fraction (ca. 15%) of RNA was reverse transcribed by the AMV RT in the absence of an oligo dT primer, which we attribute to self-priming. Self-priming of the RT can occur via the mechanism of snapped-back priming, in which the 3' end of an RNA loops back upon itself to form short and transient RNA-RNA duplexes [8], or when a 3'-hairpin structure such as the T7 terminator sequence is already present in the nascent RNA molecule. When the oligo dT primer is present in excess, we expect a negligible amount of backpriming, though self-priming of the RT on RNA molecules that have no intrinsic hybridization site might still occur. Finally, the lower rate of the RT reaction in the PURE*frex*ΔR system in comparison with the PURE buffer environment in Figure 5.4, presumably arises from the high ratio of rNTP/dNTP and non-specific binding of PURE system components to the RNA template.

Optimizing this reaction by tuning enzyme, oligo and rNTP/dNTP concentrations would be the next step, as it was done successfully for other techniques that apply a similar scheme of nucleic-acid sequence-based amplification (NASBA) [2, 4]. We should mention that NASBA is optimized as a detection tool of RNA sequences, and for efficient amplification the recommendation is to limit the target to approximately 100 to 250 nucleotides [5], which is an order of magnitude smaller than the gene template in our experiment. One could also envisage to choose another reverse transcriptase enzyme. For example, the RNase H function of the HIV-1 RT digests the RNA in shorter pieces and it was shown that this led to purer 3SR products [10]. Likewise, an RT with a higher processivity than the AMV (ca. 100 nt [11]) could pave the way for reverse transcribing longer RNA molecules. Interesting candidates are: the retrontransposable R2 element RT that has a higher processivity than retroviral RTs and has the ability of displacing RNA molecules annealed to the RNA template during cDNA synthesis [12], the recently engineered thermostable and processive MMLV RT variant that has a processivity of 1500 nt [13], or the obtained thermostable group II intron RT fusion proteins that lack an RNase H domain but are able to template switch directly to the 3' ends of new RNA templates[3] [14].

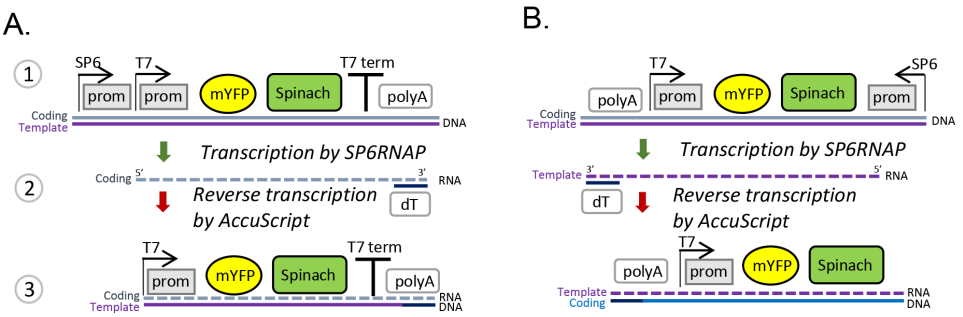


Figure 5.6: Scheme of the stepwise production of templates resulting in $DNA_{template}$:RNA_{coding} (A) or $RNA_{template}$:DNA_{coding} hybrids (B). The use of the SP6 RNAP for the transcription step enables the inclusion of a functional T7 promoter in the final template (or vice versa). C) The hybrid template is imaged on an agarose gel together with the starting double stranded DNA, the intermediate RNA, a control sample, DNA (left) and RNA ladders (right) .

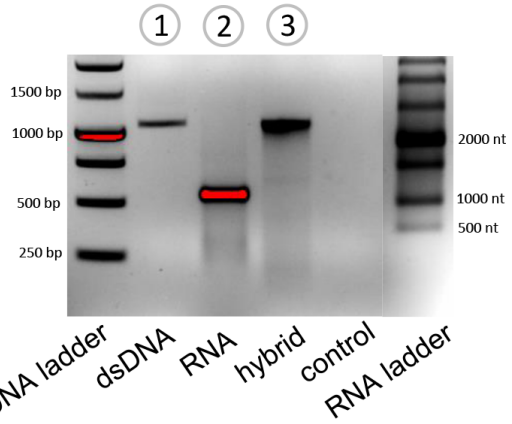


Figure 5.7: The hybrid template, constructed stepwise according to the scheme of Figure 5.6A, is imaged on an agarose gel together with the starting double stranded DNA, the intermediate RNA, a control sample, DNA (left) and RNA ladders (right).

5.4. RESULTS FOR TRANSCRIPTION AND TRANSLATION OF HYBRID DNA:RNA TEMPLATES

In this section, we have analyzed in more detail whether reverse transcription products (that appear to migrate at the same height as dsDNA in Figure 5.3C) are transcriptionally active. The product of the RNase H deficient AccuScript RT reaction is a hybrid DNA:RNA template, where the DNA represents the template strand and the RNA the coding strand for a T7 RNAP that binds to the promoter sequence (Figure 5.6A), or vice versa (Figure 5.6B). Note that the dsDNA template presented in Figure 5.6B is a potential starting template of the scheme in Figure 5.2 that combines the nucleic acid amplification cycle with gene expression. In contrast to the AccuScript RT product, the product of the AMV RT is not a full hybrid, as its RNase H subunit degrades the RNA during, and after, polymerizing the first strand of complementary DNA. The RNase H activity is nonetheless not processive and mainly results in incomplete digestion, cutting once every 100-200 nucleotides, though shorter cleavage products are also generated [11]. The remaining hybridized pieces of RNA serve as a primer template junction for second strand synthesis by the DNA dependent DNA polymerization activity of the AMV. Because the polymerization subunit of the AMV does not possess any nuclease or effective strand displacement activities [12], several RNAs can serve as an initiation and termination site for the polymerization activity, resulting in a discontinuous strand. The final product after the reverse transcription reaction on the initial RNA template is thus a double stranded nucleic acid consisting of one continuous DNA strand (first strand synthesis) and a discontinuous second strand containing a mix of RNA and DNA nucleotides.

Because the reverse transcription products of the T7-YFP-Spinach-polyA RNA sequence produced by the AMV RT and the AccuScript RT contains a DNA:RNA hybrid promoter we tested its transcriptional activity by the T7 RNAP. This hybrid construct consisting of a DNA template strand and (partially) RNA coding strand will be referred to as T7D. The Spinach detection method was applied to monitor transcription in the PURE system, as described in Chapters 2 and 3. To our surprise we found that both the T7D produced by the AMV RT (SI Figure 5.12) and by the AccuScript RT (Figure 5.8) were transcribed by the T7 RNAP, though with a lower rate than dsDNA. Moreover, the corresponding mRNA underwent subsequent translation. Next, we wondered whether the transcribability of a hybrid template depends on its orientation, in other words, if a hybrid template where the RNA is the template strand and the DNA is the coding strand ($RNA_{template}$:DNA_{coding}, or T7R), can also be transcribed by the T7 RNAP. The latter is relevant as it is an intermediate product of the nucleic acid amplification cycle proposed in Figure 5.2. We systematically prepared hybrid templates of both orientations and compared the transcription efficiency to that of a dsDNA template.

[3] This enables synthesis of a continuous cDNA that directly links an adapter to a target RNA sequence without RNA ligation [14].

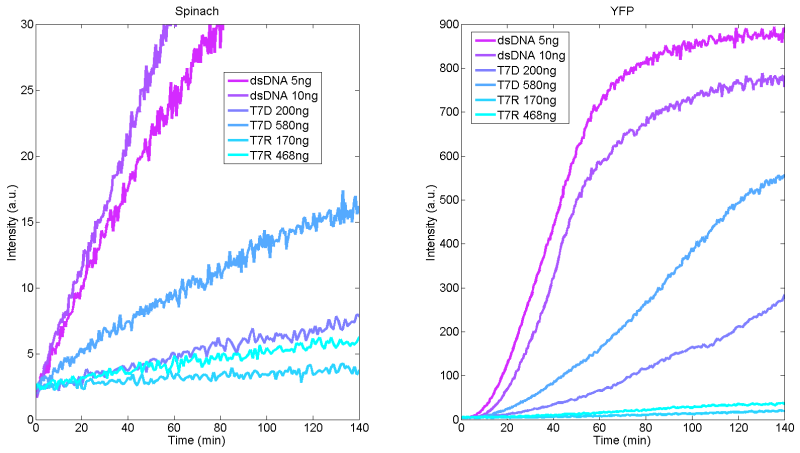


Figure 5.8: Representative kinetics of the simultaneous detection of RNA and protein expressed in the PURE system from the DNA (dsDNA), the $DNA_{template}$:RNA_{coding} hybrid (T7D) and the $RNA_{template}$:DNA_{coding} hybrid (T7R) templates. Full curves are displayed in SI Figures 5.15,5.14 and 5.16, respectively.

5

5.4.1. HYBRID TEMPLATES ARE TRANSCRIBED BY THE T7RNAP WITH REDUCED EFFICIENCY DEPENDING ON THE NATURE OF THE TEMPLATE STRAND

Purified hybrid templates are produced by an in vitro transcription reaction followed by a reverse transcription reaction performed by the AccuScript RT as shown in Figure 5.6A and B and described in *Materials and Methods*. A necessary requirement to obtain functional hybrid templates is the presence of the T7 promoter sequence in the RNA molecule. This is achieved by transcribing RNA with the SP6 RNAP, which is closely related to the T7 RNAP but has an orthogonal promoter. The resulting hybrid template contains the same sequence as the starting DNA template, only lacking the SP6 pro-

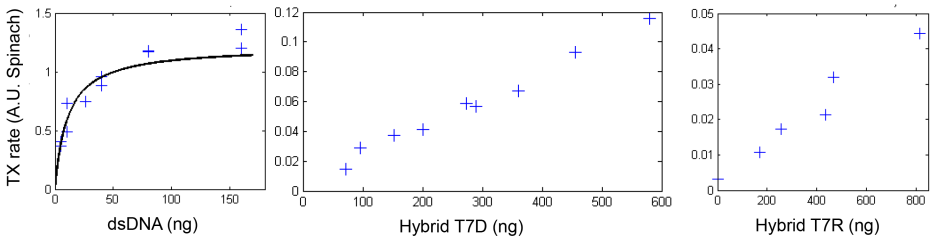


Figure 5.9: Relative transcription rates measured by linearly fitting the increase of Spinach signal in the initial 40 minutes of the reaction for the different templates. Left: transcription rates of the dsDNA template. The black line displays the fitted Michealis-Menten curve. More than 10-fold lower transcription rates are measured for the $DNA_{template}$:RNA_{coding} (middle) and the $RNA_{template}$:DNA_{coding} (right).

moter sequence, and therefore migrates at the same height as the dsDNA in gel electrophoresis (Figure 5.7). To compare the efficiency of transcription by the T7 RNAP on the different constructs, templates were expressed in the PURE*frex* system and RNA and protein production were monitored in real-time. Transcription was monitored by the fluorescence of the Spinach signal, which has proven to be a direct representation of the actual amount of produced (near) full-length transcripts [15]. We found that transcription of both the hybrid templates occurs and that the efficiency decreases in the order dsDNA > hybrid DNA_{coding}:$RNA_{template}$ > hybrid RNA_{coding}-$DNA_{template}$. To obtain quantitative information about the transcription rates, we performed reactions with a range of template concentrations (Figure 5.9). Initial phase of transcription of the dsDNA can be described by Michealis-Menten kinetics, similar to previous reports in the literature [16–18]. Assuming a conversion factor between Spinach fluorescence intensity and actual mRNA concentration of 3.9 nM / a.u. (corresponding to the value found for the DNA of mYFPgo-LL-Spinach in Chapter 3), we obtain a maximum transcription rate of v_{max} =4.7 nM / min (v_{max} = 1.2 a.u./min). Half of the maximum transcription rate is reached at a K_m of 0.8 nM for the dsDNA. In contrast, the initial transcription rates for the hybrid templates do not reach a plateau in the measured concentration ranges, indicating that K_m values for T7D and T7R are orders of magnitude higher than with dsDNA. A quantitative comparison is therefore arbitrary to the amount of starting template. As an example we compare the rates at 14 nM template (ca. 200 ng) and find that the quantitative decrease in transcription normalized to the maximum rate of dsDNA transcription, is 1/30 and 1/70, for the hybrid $DNA_{template}$:RNA_{coding} and the hybrid $RNA_{template}$:DNA_{coding}, respectively.

5.4.2. SP6 RNAP CAN ALSO TRANSCRIBE HYBRID TEMPLATES

Next, we questioned if the ability to transcribe hybrid templates is special to the T7 RNAP, or is also found in other single subunit phage RNA polymerases. The SP6 RNAP is of the same family and closely related to the T7, also commonly used for large scale in vitro transcription assays. Hybrids containing the SP6 promoter (Figure 5.10A) are constructed with similar design principles as for the T7 hybrids, by interchanging the T7 and SP6 promoters in Figure 5.6 and reversing the order of the T7 terminator and polyA tail for the steps in the $DNA_{template}$:RNA_{coding} construction protocol. We find that the SP6 RNAP displays transcription activity on both the $DNA_{template}$:RNA_{coding} and the $RNA_{template}$:DNA_{coding} hybrid (Figure 5.10B), whereas translation only occurs starting from the $DNA_{template}$:RNA_{coding} template (SI Figure 5.18). The measured transcription rate for 21 nM of the $DNA_{template}$:RNA_{coding} hybrid is 0.024 ± 0.003 A.U./min (or 0.094 nM / min, n=3), which is 50 times slower compared to the maximum transcription rate of dsDNA (by the T7 RNAP).

5.4.3. TRANSCRIPTION FROM HYBRID TEMPLATES RESULTS IN MORE INTERMEDIATE PRODUCTS

To further investigate the RNA products generated from the hybrid templates, we analysed the transcription products on nucleic acid gels. To clearly distinguish RNA products from template on the gel we added fluorescently labelled CTP to the reaction mixture that can be incorporated by both the T7 and SP6 RNAPs. Transcription was per-

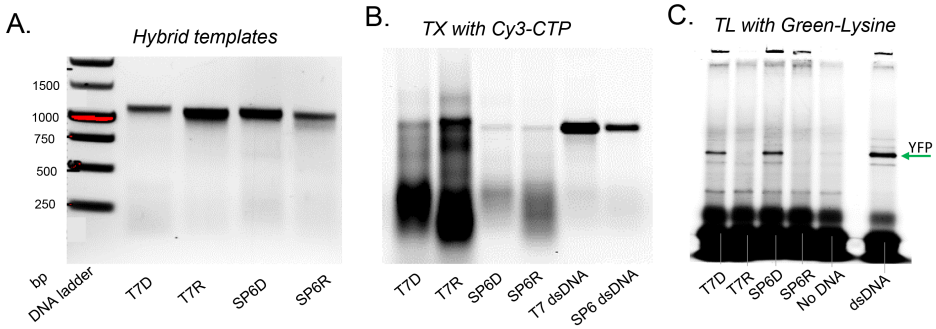


Figure 5.10: Hybrid templates are transcribed in an *in vitro* transcription (TX) system or transcribed and translated in an *in vitro* transcription-translation (TX-TL) system A.) Hybrid templates are visualized on an 1.0% agarose gel and stained with EtBr. B.) Transcription of the hybrid templates is performed in an *in vitro* TX kit in the presence of Cy3-CTP that is imaged with a fluorescence scanner. More non full-length RNA products are visible in the TX from hybrid lanes compared to TX from dsDNA. C.) Transcription and subsequent translation of hybrid templates is performed in the PURE*frex* system in the presence of a fluorescently labelled lysine amino acid (Green-Lysine). Full-length YFP peptides are produced when starting from the T7D and SP6D hybrid templates.

5

formed in the IVT kits using the dsDNA templates, the $DNA_{template}$:RNA_{coding} or the $RNA_{template}$:DNA_{coding} hybrids and the resulting amount of synthesized RNA is displayed in Figure 5.10B and SI Figure 5.19. Whereas the RNA produced from the double stranded DNA templates are all full-length transcripts, transcription from hybrid templates leads to intermediate sized products. Specifically, we observe more aborted transcripts of 100-500 nt for the $RNA_{template}$:DNA_{coding} hybrid templates, and the ratio of full length products versus total RNA products is lowest for these templates. This might explain why the YFP production in the PURE system experiments (Figure 5.8) is that low for the $RNA_{template}$:DNA_{coding} template. Indeed, the translation machinery is depleted on short RNA transcripts since there is no rescue mechanism (compared to *E.coli* that contains tmRNA and peptidyl-tRNA hydrolase among others) present to salvage stalled ribosomes on truncated or damaged mRNA transcripts. Non full-length peptides will not emit fluorescence that can be detected in the spectrofluorometer experiments (Figure 5.8), and we therefore opted for visualization of translation producys with an incorporated fluorescently labelled amino acid on a protein gel. In Figure 5.10 we qualitatively confirmed that transcription of $DNA_{template}$:RNA_{coding} hybrid templates leads to subsequent translation of the YFP. Also here, the full-length protein production from the $RNA_{template}$:DNA_{coding} templates is below the detection limit, while production of shorter peptides is likely not visible[4]. Further investigation is required to understand this last observation.

[4]The distinct lower band that is visible in the T7D, T7R, SP6D, SP6R, and no DNA lanes is probably from the remaining background GreenLys tRNA.

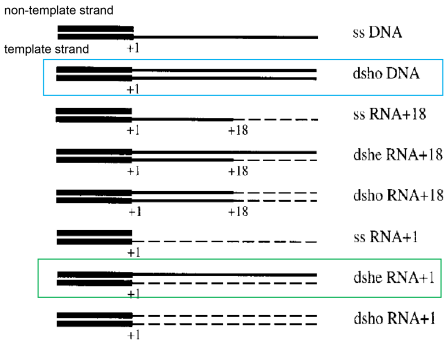
5.5. DISCUSSION FOR FOR TRANSCRIPTION AND TRANSLATION OF HYBRID DNA:RNA TEMPLATES

This research confirms that the T7 RNAP can start transcription on fully hybrid promoters, where the template strands consists of deoxynucleotides and the non-template strand of ribonucleotides or *vice versa*. This observation is important in the context of the proposed DNA amplification cycle. Besides, it provides new insights on the mechanical properties of this familiy of RNAPs.

5.5.1. T7 RNAP IS AN ENZYME WITH FLEXIBLE PROPERTIES

The T7 RNAP is known as an DNA-dependent RNA polymerase, but it also exhibits a range of other DNA- and RNA- dependent properties. They are discussed here to provide a context for this new finding. An example of these other properties is that a specific short RNA instead of a DNA molecule could serve as a template for efficient and accurate transcription by the natural T7 RNAP [20]. Furthermore, a mutant of the T7 RNAP showed greatly reduced specificity towards it substrate. This mutant (Y639F) can incorporate both deoxyribo- and ribonucleoside triphosphates into the newly synthesized strand, leading to transcripts with a mix of dNMP/rNMP composition [21]. Interestingly, it was also capable of unprimed RNA directed RNA synthesis and of unprimed (*de novo*) reverse transcription, dependent on the substrates offered. The reported functional flexibility of the T7 RNAP and its mutants raised the question if the T7 RNAP polymerase is capable of recognizing and transcribing a promoter-template that is a mix of RNA and DNA nucleotides. If the RNAP can still efficiently transcribe a hybrid template consisting of one strand of RNA and one of DNA, it opens up the possibility for a new reaction scheme that aims to amplify RNA without knowledge of its 5'-end, by ligating an RNA-DNA promoter containing oligonucleotide to the 5'-end of that molecule and performing first strand DNA synthesis from the 3'-end by a RT enzyme resulting in a hybrid template[7]. A milestone study on template specificity showed that the T7 RNAP can synthesize transcripts from ssDNA, ssRNA, or hybrid RNA-DNA templates, after initiation from a double stranded DNA promoter [19]. Their main findings are displayed in Figure 5.11A and comprised the fact that initiation and transition to elongation on RNA templates occurred but that processivity decreased or abortive initiation increased in these cases. Also, duplexes might form between the RNA template strand and the transcript strand, consistent with the higher stability of RNA-RNA hybrids as compared to DNA-RNA hybrids (that were absent). The effect of the non-template strand on the stability of the elongation complex is studied as well .Although the results presented here suggest that T7 RNAP forms a highly stable complex on ssDNA, this is not a general rule and it can be sequence dependent. In another study it was shown that transcription initiation (till +5 nt) decreases when ribonucleotides were substituted into the T7 promoter element [7] (Figure 5.11B). The authors found that there is a requirement for a greater number of DNA residues within the template strand compared to the non-template strand. Although transcription still occurred when DNA residues in the promoter were partly substituted by RNA residues, they reported no detectable activity when full RNA substi-

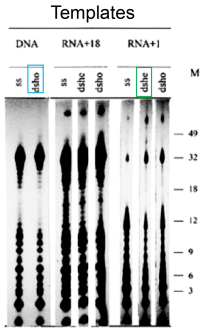
A.



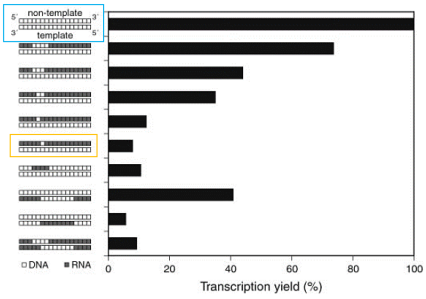
Quantitation of transcription products synthesized by T7 RNAP from templates

Templates	Runoff (pmol)	Abortive (pmol)	Intermediate (pmol)	Initiation events[a]	Efficiency[b]
ss DNA	240	1760	416	242	1/10
dsho DNA	80	1312	128	152	1/18
ss RNA+18	800	4480	1216	650	1/8
dshe RNA+18	208	2496	384	309	1/14
dsho RNA+18	176	2240	304	272	1/16
ss RNA+1	13	1280	112	141	1/108
dshe RNA+1	3.2	576	32	61	1/227
dsho RNA+1	6.4	1024	64	110	1/157

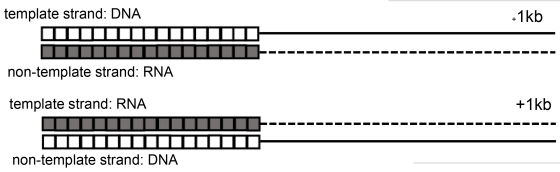
B.



C.



D.



Transcription by	T7 RNAP	Efficiency to dsDNA	SP6 RNAP
+1kb	✓	1/30	✓
+1kb	✓	1/70	✓

Figure 5.11: Studies of T7 RNAP capabilities to transcribe various forms of hybrid DNA-RNA templates. A) Figures adapted from [19]. Left: Templates of DNA (solid lines) or RNA (dashed line), or a combination, were tested for transcribability by the T7 RNAP starting on a dsDNA promoter. Template strand consists of the consensus promoter and initiation region (from +1 to +6) plus 26 bases downstream sequence. Templates related to our interest are the reference double stranded homoduplex DNA (dsho DNA, indicated by blue box), the double stranded heteroduplex (dshe) RNA +1 (indicated by green box) and RNA + 18 templates. Rights: Results of transcription from different templates was analyzed. Transcription efficiency corresponds to the number of initiation events that lead to runoff RNA versus total number of initiation events. Note that the dsho RNA+1 has more than 10-fold less runoff products than the reference dsho DNA template. B) Figure showing experimental results from (A). Transcripts incorporating [α^{32}P]ATP were analyzed on gel and the generation of aborted initiation events (till 12 nt), intermediate products (between 12-32 nt), run-off products (32-34 nt) and hybridized products (top) are visible. The RNA+18 template produces most intermediate and hybrid products. C) Figure adapted from [7]. Ribonucleotides were introduced at various locations within the template or non-template (coding) strand of the promoter which reduced transcription yield. There appears to be a critical window of deoxyribonucleotides necessary to maintain transcribability. D) In this study we found that full hybrid templates (visualized in relation to representations in A and B) can be transcribed by both the T7 RNAP and SP6 RNAP with reduced runoff efficiency compared to transcription from a dsDNA.

tution of the promoter strand took place under the assay conditions[7][5]. In Figure 5.11C, an important finding of our study is that the T7 RNAP can initiate transcription on full hybrid promoters ($DNA_{template}$:RNA_{coding} and $RNA_{template}$:DNA_{coding} hybrids), with a reduced runoff efficiency, as determined by analyzing the Spinach signal that reports on the transcription rate, whose values are 1/30 and 1/70 compared to dsDNA, respectively. We also find that more intermediate products are generated during the elongation process when RNA is the template strand as opposed to DNA being the template strand (Figure 5.10B and Figure 5.19). We hypothesize that this could be due to the stable $RNA_{transcript}$:$RNA_{template}$ duplex that could prevent dissociation and elongation by the RNAP. Intermediate products are however also generated for the $DNA_{template}$:RNA_{coding} template, suggesting that the non-template strand also plays a role in this process. Another single subunit viral RNAP, the closely related SP6 RNAP, can likewise initiate transcription on these templates and generates less aborted products in this process (see Figure 5.10B).
Placed in this context of the many different properties of the T7 RNAP, it is not that surprising that this enzyme is able to initiate transcription on a fully hybrid promoter. The structural difference between a DNA:DNA duplex molecule that adopts the B-form (10.5 bp/turn) and a RNA:DNA duplex molecule is its A-form (11 bp /turn) and the higher stability of the latter [22], which might hinder the separation of the strands in the initiation phase. The sequence specific recognition of a duplex promoter is accomplished by a combination of direct and indirect readout in adjacent major and minor grooves, where for the T7 RNAP the main recognition sites are four protein side chains that interact with the four bases G-9, A-8, G-7 and G-11 [23], which remain the same bases in RNA as in DNA. This is to our best knowledge the first report that provides information on the transcribability of full hybrid templates, complementing the reported properties of the T7 RNAP in literature, plus the first investigation of the possibility to utilize these templates in gene expression systems.

5.5.2. TRANSLATION FROM HYBRID TEMPLATES

In this research we showed that hybrid molecules (T7D and SP6D) can be used as templates for gene expression. The final yield of fluorescent protein expression from the T7R template is negligible (34 ±12 a.u. (mean ± standard deviation) of YFP fluorescence) compared to the expression of both the T7D template (613 ± 124 a.u.) and the T7 dsDNA template (859 ± 84 a.u.). The final expression yield from the hybrid $DNA_{template}$:RNA_{coding}, albeit lower than the yield from dsDNA expression, falls in a similar range, despite the much lower production of RNA (after 2 h an order of magnitude less). That lower RNA levels can lead to higher protein-to-RNA ratio is a general observation in the PURE system, ascribed to the saturation of translation machinery when mRNA is too abundant [16]. Interestingly, the low production of RNA seems to have a second advantage: the expression time is clearly prolonged for translation of the hybrid T7D template compared to the T7 dsDNA template (compare YFP curves in SI Figure 5.14 with SI Figure 5.15).
It is unfortunately too early to state strongly that transcription and subsequent translation of hybrid templates lead to similar protein expression levels as starting from dsDNA

[5] This paper studied only the initiation step of transcription, with a template strand that was 5 nt longer than the non-template strand.

templates, since we observed this only for one gene template (that had lower expression levels than similar templates studied in Chapter 3). We also note that the chance of errors in RNA transcripts of the hybrid templates is higher, an inevatible effect that is inherent to the process of producing the hybrid templates. The error rate of SP6 and T7 RNAP in transcription is around $1{\cdot}10^{-4}$, which is higher than the error rate for RT enzymes (AMV: $6{\cdot}10^{-5}$ and AccuScript $1{\cdot}10^{-5}$). This means that around 1 out of 10 templates contains a mutation in the sequence of the hybrid template which may lead to the incorporation of a different amino acid by the ribosome upon translating the transcribed RNA molecule. It is thus expected that less full-length peptides with the correct amino acids are produced from hybrid templates than from dsDNA. This can be a disadvantage; however it can also be turned into an advantage to introduce errors in gene sequences for evolution experiments.

We find that transcription from the T7D template produced by the reverse transcription process of the AMV RT (Figure 5.12) is more efficient than the transcription from T7D produced by the AccuScript RT, namely 0.17 a.u/min (0.66 nM / min) versus 0.04 a.u./min (0.16 nM / min). The possible explanation for this discrepancy lies in the nature of the final reaction product. As explained in Section 5.4 the final product after the reverse transcription reaction by AMV RT on the initial RNA template is a double stranded nucleic acid consisting of one continuous DNA strand (first strand synthesis) and a discontinuous second strand containing a mix of RNA and DNA nucleotides. The promoter site itself (29 nucleotides from the 5'-RNA end in the T7D template) may also consist of a mix of ribonucleotides and deoxynucleotides in the second strand or in even a few cases a fully double stranded DNA promoter, which would drastically increase transcription efficiency in the bulk measurement. This hypothesis can also explain the observation that no protein is produced from the transcripts coming from T7R AMV product: the template strand is discontinuous, which results in deletions of nucleotides at the discontinuous sites [24] in the RNA product, something detrimental for correct translation of the YFP. Similarly, protein expression from the SP6 templates occurs only from the SP6D template.

5.5.3. RELEVANCE OF THE TRANSCRIPTIONAL ABILITY OF RNA:DNA HYBRID IN THE CONTEXT OF THE PROPOSED AMPLIFICATION CYCLE

In the proposed nucleic acid amplification cycle of Figures 5.1 and 5.2, an intermediate species of the cycle is an DNA:RNA hybrid. In the scheme of Figure 5.1 the one-gene hybrid template has the orientation of $DNA_{template}$:RNA_{coding}, whereas the hybrid in Figure 5.2 is an $RNA_{template}$:DNA_{coding}. From this research, we learn that the transcription from a $RNA_{template}$:DNA_{coding} by the T7 RNAP is lower than transcription from a $DNA_{template}$:RNA_{coding} template. Therefore, the additional advantage of the decoupling of the amplification cycle from translation processes as proposed in Figure 5.2, is that the intermediate RNA:DNA hybrid will not serve as a competing template for transcription by the RNAP, as transcription from this template compared to transcription from dsDNA templates is negligible.

5.5.4. POSSIBLE APPLICATION: RNA AMPLIFICATION SCHEME

A possible application for this mechanism is, as discussed in the previous section, an amplification scheme for an RNA molecule without a priori knowledge of the 5'-sequence of the molecule. An RNA (or DNA-RNA) oligonucleotide containing a promoter sequence could be ligated to that RNA molecule by an RNA ligase, or in the process of reverse transcription by the group II intron RT fusion protein mentioned in section 5.3. First strand DNA synthesis by a reverse transcriptase initiated from a hybridized oligonucleotide on the 3'-end of the RNA molecule will then produce a hybrid $DNA_{template}$:RNA_{coding} template. This template can be transcribed by an RNAP like the SP6 RNAP (less aborted intermediate products are formed than for the T7 RNAP), and the produced RNA can enter the cycle again, leading to amplification of the initial RNA molecule over time.

5.5.5. EVOLUTIONARY PERSPECTIVE

As hypothesized in the RNA world scenario (see also Chapter 1 and 4), RNA preceded DNA as genetic material, which implies an early emergence of the reverse transcription process. This led to the prediction of prokaryotic reverse transcriptases, which is consistent with the protovirus hypothesis (RTs are more ancient than retroviruses, which are the source of RTs used in this study) [25]. An hypothesis related to the evolutionary transition from a primitive RNA-dependent RNA polymerase into the extant DNA-dependent enzymes that facilitated the transition from an RNA-protein to an RNA-DNA-protein world is posed by Lazcano and co-workers in [25]. They envision that this transition of enzyme capabilities happened via an ancestral reverse transcription activity, that was mediated by low specificity nucleic acid polymerases in which novel catalytic properties were acquired by few conformational modifications. Their argument is based on the highly structural conservation of a 14-amino-acid-residue segment consisting of an Asp-Asp pair flanked by hydrophobic amino acids, a same type of domain being found in almost all cellular and viral RNAPs. This is interpreted as the result of adaptation of the ancestral polymerase to different templates and substrates during the transition from RNA to DNA cellular genomes. An enzyme involved in this transition should thus have reverse transcription capabilities, as well as showing a degree of non-specificity towards the nature of its substrate and template. As described earlier in this section, the capabilities of the T7 RNAP and it mutants are reported to include the extension of DNA and RNA primers with rNTPs and dNTPs (with less efficiency and reducing specificity by replacing Mg^{2+} with Mn^{2+}), plus unprimed RNA directed RNA synthesis and unprimed reverse transcription[21]. Our study, together with earlier studies [7, 19], prove that the T7 RNAP exhibits non-specificity towards its template to some degree. This makes the T7 RNAP with its flexible properties an interesting model enzyme from an evolutionary perspective. It also provides an interesting starting point to develop an RNAP that can recognize and transcribe from a promoter sequence on a fully dsRNA template.

5.6. SUMMARY

In this study, we investigated how a nucleic acid amplification cycle (based on 3SR) could be integrated into a gene expression environment in order to combine DNA replication with transcription and translation processes. We confirmed the individual steps

Construct name	Transcription by	For hybrid production of
Sp6p-T7p-mYFPgo-LL-Spinach-polyA	SP6 RNAP	T7D
dT-T7p-mYFPgo-LL-Spinach-rSP6p	SP6 RNAP	T7R
T7p-SP6p-mYFPgo-LL-Spinach-polyA	T7 RNAP	SP6D
dT-SP6p-mYFPgo-LL-Spinach-rT7p	T7 RNAP	SP6R

Table 5.1: DsDNA constructs that are utilized in the process of hybrid template production.

of the nucleic acid amplification cycle that has RNA as an input template and can potentially amplify DNA and RNA molecules in a cyclic manner, for a one-gene construct of ca. 1 kb in length. We investigated the intermediate products of this cycle and found that DNA:RNA hybrid templates can be transcribed by the viral T7 RNAP. Hybrid templates are transcribed by the T7 RNAP with reduced efficiency depending on the nature of the template strand, in the order dsDNA > hybrid DNA_{coding}:$RNA_{template}$ > hybrid RNA_{coding}:$DNA_{template}$. The closely related viral SP6 RNAP was also able to transcribe hybrid templates of both orientation. More aborted products were generated upon transcription from hybrid templates, although this did not prevent the translation of full-length protein in a minimal gene expression system starting from hybrid templates.

5.7. MATERIALS AND METHODS

DNA TEMPLATES

The DNA templates originate from the gene optimized (Eurogentec) mYFPgo-LL-Spinach construct described in Chapter 3, with exception of the templates in the experiment in Figure 5.3 and fig:RT-pure where the DNA templates are variants of the mYFP-LL-Spinach (Figure 5.3B and Figure5.4), or the YFP-Spinach (Figure 5.3C) described in Chapter 2.

Modifications to this template such as the change of promoter, addition of reverse promoters and polyA tails, are introduced by PCR reactions (Phusion enzyme, Finnzymes) with the relevant primers that are listed in the Supplementary Information in Table 5.2. The Sp6p-T7p-YFP-Spinach-polyA template is produced by attaching the SP6 promoter (SP6p) and a polyA-tail to the T7p-YFP-Spinach-T7t (referred to as YFP-Spinach in Chapter 2) with primers 166 ChD and 167 ChD, respectively. The same procedure holds for the Sp6p-T7p-mYFPgo-LL-Spinach-polyA with the mYFPgo-LL-Spinach as PCR template. The dsDNA control for the T7 template (Figure 5.8 and Figure 5.15) contains the same sequence from both ends as the T7D hybrid template and is produced by a PCR reaction with primers 427 ChD (166 ChD without SP6 promoter sequence) and 167 ChD. The dT-T7p-mYFPgo-LL-Spinach-rSP6p (r=reverse) is produced with primers 280 ChD and 279 ChD.

The SP6p-mYFPgo-LL-Spinach-polyA template is produced by replacing the T7 promoter with an SP6 promoter and the T7 terminator with a poly A tail in the mYFPgo-LL-Spinach construct with primers 233 ChD and 190 ChD, respectively. The T7p-SP6p-mYFPgo-LL-Spinach-polyA is subsequently produced with primers 442 ChD and 190 ChD and the dT-SP6p-mYFPgo-LL-Spinach-rT7p is produced with primers 443 ChD and 444 ChD.

An overview of the templates that are used for hybrid construction is given in Table 5.1.

ASYMMETRIC PCR TO OBTAIN SSDNA TEMPLATES

Single-stranded DNA was obtained with asymmetric PCR, a method described in detail in [26]. The ssDNA without SP6 promoter sequence used in the experiment in Figure 5.3B was produced by two consecutive PCRs with the primers 235 ChD and 190 ChD. The first PCR served to get rid of the SP6 promoter sequence, and the second to generate a mix of ssDNA/dsDNA, by diluting the forward primer (235 ChD) 50 times with respect to the reverse primer. This creates a limiting concentration of the forward primer at one point in the PCR process, resulting in the amplification of only the strand that is elongated from the primer present in excess. Staining of gels with ssDNA was done with SybrGold that has a higher affinity to ssDNA than EtBr. The primer 234 ChD contains the SP6 promoter and an overhang to anneal to the ssDNA template created, as used in the experiment in Figure 5.3.

REVERSE TRANSCRIPTION REACTION IN PURE BUFFER AND PURE SYSTEM

The reverse transcription with the AMV RT was performed in a 25-μL reaction with 1 μg input RNA template (containing the T7p-mYFP-LL-Spinach-polyA sequence), 30 units of AMV RT (Promega), 1 mM dNTPs, 0.5 μg poly-$(dT)_{24}$ primer in either the PURE*frex*ΔR system (10 μl feeding solution and 1 μl enzyme solution) or in the PURE buffer (here: 35 mM HEPES, 70 mM potassium glutamate, 9 mM magnesium acetate, pH 7.6). The reaction was incubated for 3 h at 37 °C and 2.5-μl samples were harvested after the indicated time-points and quenched with EDTA (50 mM final). Samples were analyzed on a 1.2% agarose gel stained with EtBr. The backpriming control was performed with 0.7 μg input RNA, 1 mM dNTPs and 15 units of AMV in the 1x AMV buffer from the supplier (Promega) for 3 h at 37 °C.

NUCLEIC ACID CYCLE WITH FLUORESCENT PRIMER INCORPORATION

The cycle starting from the RNA template up to the dsDNA template was performed in the PURE buffer (here: 35 mM HEPES, 70 mM Potassium Glutamate, 9 mM Magneisum Acetate, pH 7.6) in a 20 μl reaction containing 400 ng of RNA (produced by the SP6 RNAP from the SP6p-mYFPgo-LL-Spinach-polyA template), 1 μM of the 5'-FAM-$(dT)_{30}$ primer (263 ChD), 2 μM of the 5'-TAMRA-SP6p primer (ChD 261), 1 mM dNTPs, 1 mM DTT and either 20 units of AMV RT or 1 μl of AccuScript enzyme (Agilent). The RNA and the 5'-FAM-$(dT)_{30}$ primer were heated for 5 min at 65 °C and cooled down for 5 min at room temperature prior assembly of the reaction mixture. After 1 h, 10-μl sample was stored at -20 °C until further use, and 0.25 μl of RNase H (13 units, Promega) was added to the remaining 10 μl reaction mixture and incubated for an additional 2 h. The reaction samples were column purified with the RNeasy MinElute Cleanup kit (Qiagen) according to the manufacturer's protocol, to remove background fluorescence from non-specific binding to macromolecules in the solution that otherwise interferes with gel analysis. Samples were loaded on a 1.1% agarose gel, imaged on a Typhoon scanner in the FAM (excitation: 488 nm emission: 520 BP 40) and TAMRA (excitation: 532 nm emission: 610 BP 30) channels, and afterwards stained with EtBr.

CONSTRUCTION OF HYBRID TEMPLATES

The hybrid templates are generated by a reverse transcription reaction on an RNA template with a hybridized DNA oligo. First, the RNA is transcribed from the DNA templates

in Table 5.1 by the T7 or SP6 RiboMax Large Scale RNA production kit (Promega) and purified with RNeasy MinElute Cleanup kit (Qiagen) according to the manufacturer's protocol, including the DNase I treatment. Secondly, 5 μM of poly-dT(30) oligo is hybridized to 1.5 μg of RNA in RNase-free water and AccuScript buffer (Agilent) by heating the solution for 5 min at 65 °C and cooling down for 5 min at room temperature. Then, DTT (10 mM final), dNTP mix (1 mM final, Promega) and 1 μl AccuScript enzyme (Agilent) are added to reach a final reaction volume of 20 μl and incubated for 3 h at 39 °C. Alternatively, this reverse transcription reaction is performed with the AMV (Promega) according to the manufacturer's protocol. After an additional incubation of 30 min at 37 °C with 0.5 μl RNaseOne (Promega), the reaction samples are column purified (Qiagen, same as above) and eluted in RNase free water. Concentrations of the obtained hybrid templates are measured on the Nanodrop (Thermo Scientific) with absorbance measurements at 280 nm, utilizing the same extinction coefficient as for DNA. Purity of the final products is checked on a 1.0% agarose gel with EtBr. RNA samples are denatured in formaldehyde loading buffer at 65 °C for 5 min and chilled on ice prior loading on the gel. The 1-kb plus DNA ladder (Promega) and ssRNA ladder (NEB) are used to measure the length of the products. To exclude the possibility that traces of DNA or RNA in the hybrid sample can lead to transcription or translation in downstream reactions, control samples are generated and tested following the exact same protocol but omitting the addition of the reverse transcriptase enzyme.

QUANTIFICATION OF TRANSCRIPTION EFFICIENCY IN THE PURE SYSTEM

For transcription and successive translation of the templates, the in vitro gene expression PURE*frex* (GeneFrontier) is utilized. The PURE*frex* system contains the T7 RNAP and when indicated SP6 RNAP (Promega, 80 units) is supplemented to the system. The chromophore DFHBI (20 μM final) is added to the reaction mixture and becomes fluorescent upon binding the Spinach aptamer. For real-time monitoring of RNA and protein production, a 20-μl reaction is transferred to a 15-μl cuvette (Hellma) positioned in a Peltier thermostated four-cell holder maintained at 37 °C, in a fluorescence spectrophotometer (Cary Eclipse, Varian). Acquisition of the Spinach and YFP fluorescence occurs every 30 s or 60 s at excitation/emission wavelengths of 460/502 and 515/528 nm.

VISUALIZING TRANSCRIPTION FROM HYBRID TEMPLATES ON NUCLEIC ACID GELS

To visualize the low amounts of RNA products transcribed from the hybrid template, in vitro transcription assays were performed in the presence of a fluorescently labelled nucleotide. Transcription was performed for 7 h at 37 °C by the T7 or SP6 RiboMax Large Scale RNA production kit (Promega) in 20 μl reactions containing ca. 700 ng of hybrid template, RNase inhibitor (10 units SUPERase.In, Ambion), 1 mM NTPs and 0.05 mM of 5-Propargylamino-CTP-Cy3 (Jena Biosciences). Samples after clean-up were loaded on a 1.0% agarose gel and imaged on a Typhoon scanner with an excitation wavelength of 532 nm and emission filter 580 ± 30nm.

VISUALIZING TRANSLATION PRODUCTS FROM HYBRID TEMPLATES ON PROTEIN GEL

Production of proteins was monitored by the incorporation of a fluorescently labelled amino acid that was supplemented to the PURE*frex* reaction. Protein synthesis was performed in a 10-μl reaction with 150 ng of hybrid templates or 20 ng of SP6-T7p-mYFPgo-

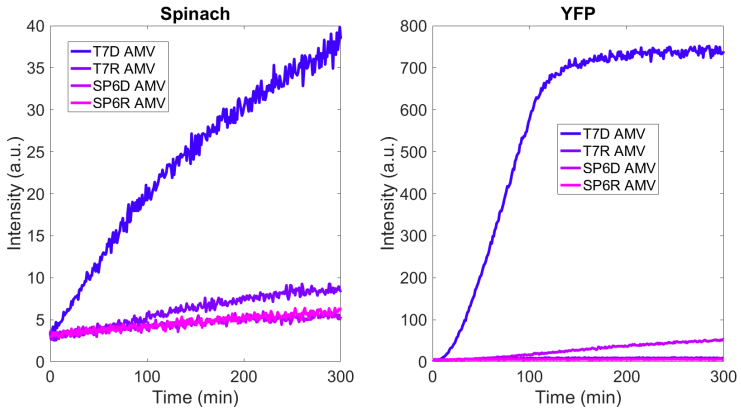


Figure 5.12: The reverse transcription products of the AMV RT were tested for their ability to be transcribed by the RNAP of its promoter signal. (Partly) hybrid templates with a T7 promoter (T7D or T7R) were made according to the scheme in Figure 5.6 (replacing AccuScript with AMV RT). The SP6D and SP6R were made similarly by reversing the SP6 and T7 promoters and omitting the T7 terminator in Figure 5.6A and B, respectively. Transcription was only observed from products were the DNA is the template strand (T7D and SP6D).

5

LL-Spinach-polyA template in the presence of BODIPY-Lys-tRNA$_{Lys}$ (FluoroTect^{TM} GreenLys, Promega) and 6 units SUPERase.In^{TM}, Ambion). The SP6D and SP6R samples are supplemented with 40 units of SP6 RNAP (Promega). Samples were incubated for 5 h at 37 °C, treated with 0.5 μl RNaseONE (Promega) for 30 min at 37 °C, denatured for 2.5 min at 65 °C in the 2x Laemmli loading buffer (Sigma-Aldrich), and analyzed on a 12 % SDS-PAGE using a fluorescence gel imager (Typhoon, Amersham Biosciences).

5.8. ACKNOWLEDGEMENTS

I would like to thank Christophe for the idea to investigate the RT process for DNA replication and Fabrizio Anella and David Dulin for discussions. Great thanks to Ilja Westerlaken who performed control experiments (Figure 5.10)and for interesting discussions. I am grateful to Sabine van Schie, who joined this project in the first stages and performed excellent work, resulting in Figure 5.3C and Figure 5.4 and the idea to produce ssDNA with the asymmetric PCR method, among others. I want to thank Sabine for her experimental accuracy, scientific creativity and enthusiasm that contributed substantially to the progress at the early phase of this project.

5.9. SUPPLEMENTARY INFORMATION

Primer	5'-3' Sequence	label
166 ChD	AAGATTTAGGTGACACTATAGAATACAAGCTTGGG CTGCAGCGCGAAATTAATACGACTCACTATAGGGAGACC	
167 ChD	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAA CCCCTCAAGACCCGTTTAGAGG	
190 ChD	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATGCT AGCCCGGGGATATC	
233 ChD	TTATTTAGGTGACACTATAGAAGGGAGAGACCACA ACGGTTTCCCTCTAG	
234 ChD	TTATTTAGGTGACACTATAGAAGGGAGAGACCAC AACGGTTTC	
235 ChD	GAAGGGAGAGACCACAACGGTTTCC	
261 ChD	same as 234 ChD	5'-TAMRA
263 ChD	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT (dT_{30})	5'-FAM
279 ChD	AAGATTTAGGTGACACTATAGAATACAAGCTTATG CTAGCCCGGGGATATC	
280 ChD	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCGAAAT TAATACGACTCACTATAGGGAGACC	
427 ChD	GAATACAAGCTTGGGCTGCAGCGCGAAATTAATAC GACTCACTATAGGGAGACC	
442 ChD	GCGAAATTAATACGACTCACTATAGGGAGACCTAA TTGCCTATTTAGGTGACACTATAGAAGGG	
443 ChD	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCTA TTTAGGTGACACTATAGAAGGGAGAGACC	
444 ChD	GAAATTAATACGACTCACTATAGGGAGACCTGCTA GCCCGGGGATATCGACGCGAC	

Table 5.2: List of primer sequences. Promoter sequences of SP6 and T7 are underlined.

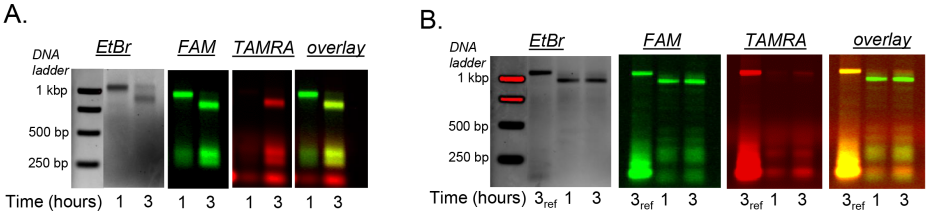


Figure 5.13: Trials of the nucleic acid amplification cycle starting from RNA and with simultaneous addition of the AccuScript RT and the RNase H led to slightly shorter products than the expected full-length dsDNA with either both primers present or mainly the dT primer (FAM). A) This reaction was performed in a 20 μl solution according to the composition described in section 5.7, with the difference that 0.3 μl RNase H was added from the beginning and 300 ng mRNA was used transcribed from the SP6-mYFP-Spinach-polyA template. B) A small fraction of full-length dsDNA that incorporated two primers is visible. This reaction was performed in a 20 μl according to the composition as described in section 5.7, with the difference that 0.5 μl RNase H and 1 μM 261 ChD primer was added from the beginning. As a dsDNA reference the final product after 3 h (3_{ref}) from the AccuScript reaction with subsequent addition of RNase H from Figure 5.5 is displayed.

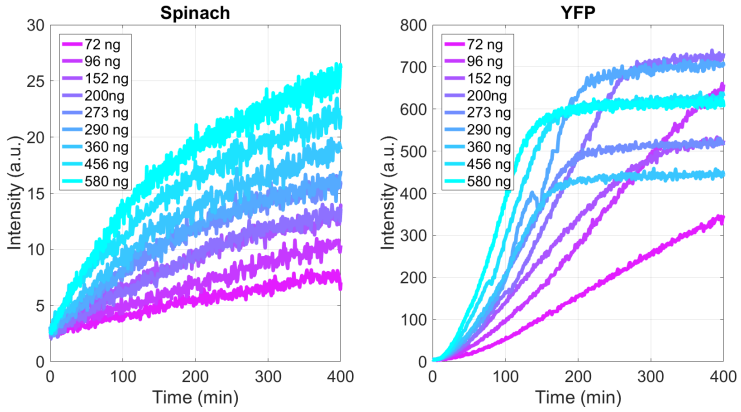


Figure 5.14: Expression kinetics of Spinach and YFP from the **T7 dsDNA** template in the PURE*frex* system.

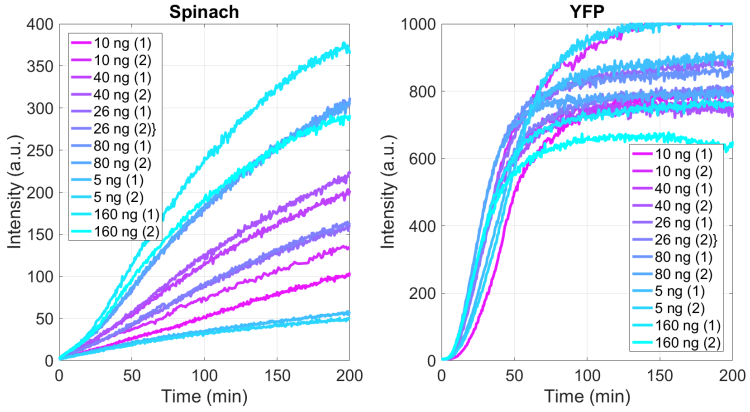


Figure 5.15: Expression kinetics of Spinach and YFP from the **T7D** hybrid template in the PURE*frex* system.

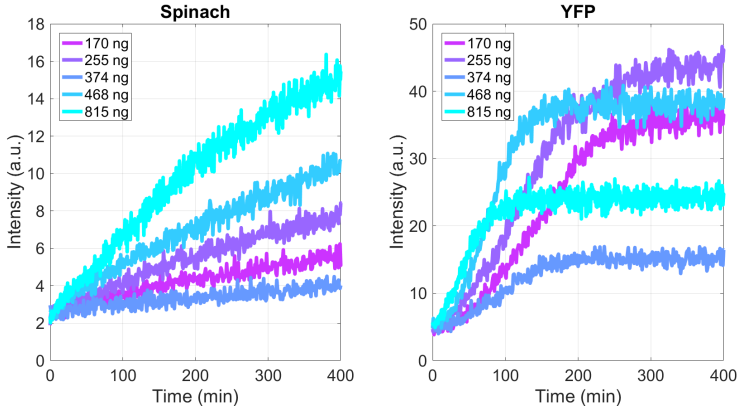


Figure 5.16: Expression kinetics of Spinach and YFP from the **T7R** hybrid template in the PURE*frex* system.

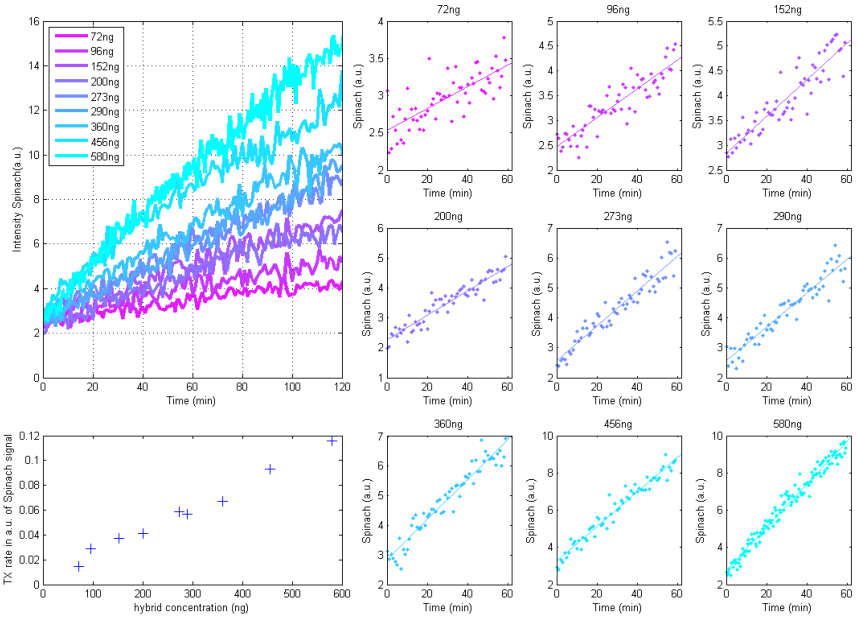


Figure 5.17: Linear fits of the first 60 min of the Spinach transcription signal for all the curves analyzed for the **T7D** templates and displayed in the main panel.

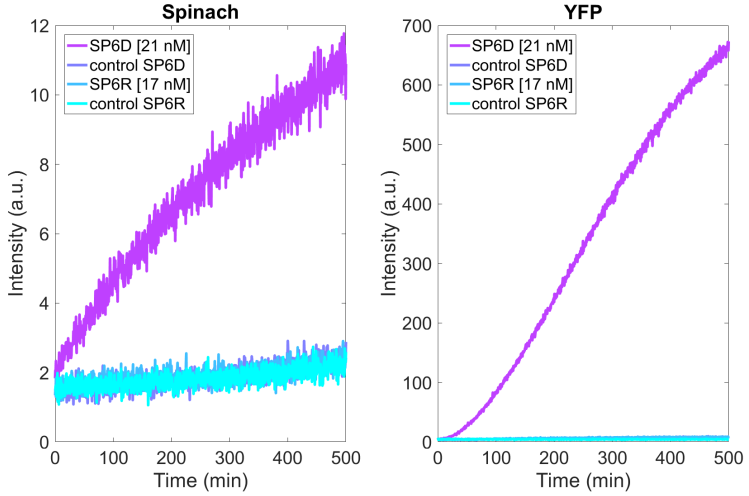


Figure 5.18: The SP6 RNAP exhibits transcription activity on the $DNA_{template}$:RNA_{coding}, but not on the $RNA_{template}$:DNA_{coding} hybrid template (no increase in Spinach signal compared with the control sample). The control sample is produced using the hybrid construct protocol but with omission of the AccuScript enzyme. This control experiment confirms that there is no residual amount of carry over of DNA/RNA that contributes to transcription or translation reactions.

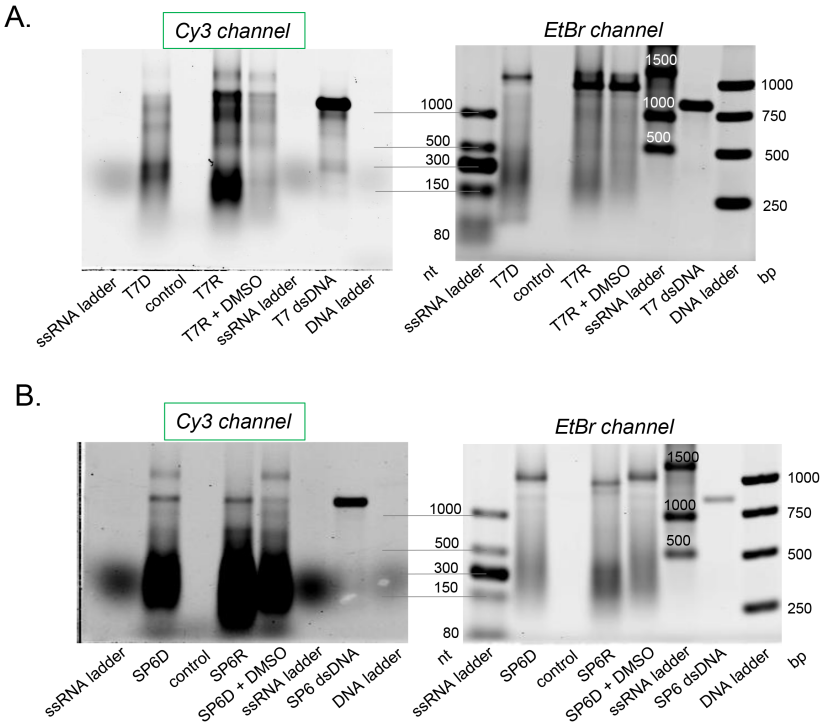


Figure 5.19: Transcription of the hybrid templates with incorporation of a fluorescently labelled Cy3-CTP imaged on RNA agarose gels for the T7 templates (A) and SP6 templates (B). Transcription from both hybrid templates generates a large amount of non-full length products (mostly between 150-500 nt), whereas this is not the case for transcription from a ds DNA template containing the same sequence. DMSO does not attenuate the formation of aborted products. The control samples were produced with the hybrid construct protocol with omission of the Accuscript enzyme. This control experiment confirms that there is no residual amount of carry over of DNA/RNA that contributes to transcription or translation reactions.

REFERENCES

[1] J. C. Guatelli, K. M. Whitfield, D. Y. Kwoh, K. J. Barringer, D. D. Richman, and T. R. Gingeras, *Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication.* Proceedings of the National Academy of Sciences of the United States of America **87**, 7797 (1990).

[2] J. Compton, *Nucleic acid sequence-based amplification.* Nature **350**, 91 (1991).

[3] B. Wlotzka and J. S. McCaskill, *A molecular predator and its prey: coupled isothermal amplification of nucleic acids,* Chemistry & Biology **4**, 25 (1997).

[4] P. R. Moll, J. Duschl, and K. Richter, *Optimized RNA amplification using T7-RNA-polymerase based in vitro transcription.* Analytical biochemistry **334**, 164 (2004).

[5] B. Deiman, P. van Aarle, and P. Sillekens, *Characteristics and applications of nucleic acid sequence-based amplification (NASBA).* Molecular biotechnology **20**, 163 (2002).

[6] N. Ma and W. T. McAllister, *In a Head-on Collision, Two RNA Polymerases Approaching One Another on the Same DNA May Pass by One Another,* Journal of Molecular Biology **391**, 808 (2009).

[7] K. E. McGinness and G. F. Joyce, *Substitution of ribonucleotides in the T7 RNA polymerase promoter element,* Journal of Biological Chemistry **277**, 2987 (2002).

[8] A. Tuiskunen, I. Leparc-Goffart, L. Boubis, V. Monteil, J. Klingstrom, H. J. Tolou, A. Lundkvist, and S. Plumet, *Self-priming of reverse transcriptase impairs strand-specific detection of dengue virus RNA,* Journal of General Virology **91**, 1019 (2010).

[9] R. R. Breaker and G. F. Joyce, *Emergence of a replicating species from an in vitro RNA evolution reaction.* Proceedings of the National Academy of Sciences of the United States of America **91**, 6093 (1994).

[10] M. Gebinoga and F. Oehlenschläger, *Comparison of self-sustained sequence-replication reaction systems.* European journal of biochemistry / FEBS **235**, 256 (1996).

[11] J. J. DeStefano, R. G. Buiser, L. M. Mallaber, T. W. Myers, R. A. Bambara, and P. J. Fay, *Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and moloney murine leukemia viruses are functionally uncoupled,* Journal of Biological Chemistry **266**, 7423 (1991).

[12] A. Bibillo and T. H. Eickbush, *High processivity of the reverse transcriptase from a non-long terminal repeat retrotransposon,* Journal of Biological Chemistry **277**, 34836 (2002).

[13] A. Baranauskas, S. Paliksa, G. Alzbutas, M. Vaitkevicius, J. Lubiene, V. Letukiene, S. Burinskas, G. Sasnauskas, and R. Skirgaila, *Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants,* Protein Engineering, Design and Selection **25**, 657 (2012).

[14] S. Mohr, E. Ghanem, W. Smith, D. Sheeter, Y. Qin, O. King, D. Polioudakis, V. R. Iyer, S. Hunicke-Smith, S. Swamy, S. Kuersten, and A. M. Lambowitz, *Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing.* RNA (New York, N.Y.) , 958 (2013).

[15] P. van Nies, A. S. Canton, Z. Nourian, and C. Danelon, *Monitoring mRNA and protein levels in bulk and in model vesicle-based artificial cells.* Methods in enzymology **550**, 187 (2015).

[16] P. van Nies, Z. Nourian, M. Kok, R. van Wijk, J. Moeskops, I. Westerlaken, J. M. Poolman, R. Eelkema, J. H. van Esch, Y. Kuruma, T. Ueda, and C. Danelon, *Unbiased tracking of the progression of mRNA and protein synthesis in bulk and in liposome-confined reactions.* Chembiochem : a European journal of chemical biology **14**, 1963 (2013).

[17] T. Stögbauer, L. Windhager, R. Zimmer, and J. O. Rädler, *Experiment and mathematical modeling of gene expression dynamics in a cell-free system.* Integrative biology : quantitative biosciences from nano to macro **4**, 494 (2012).

[18] H. Niederholtmeyer, L. Xu, and S. J. Maerkl, *Real-Time mRNA Measurement during an in Vitro Transcription and Translation Reaction Using Binary Probes,* ACS synthetic biology **2**, 411 (2013).

[19] N. Arnaud-Barbe, V. Cheynet-Sauvion, G. Oriol, B. Mandrand, and F. Mallet, *Transcription of RNA templates by T7 RNA polymerase.* Nucleic acids research **26**, 3550 (1998).

[20] M. M. Konarska and P. a. Sharp, *Replication of RNA by the DNA-dependent RNA polymerase of phage T7.* Cell **57**, 423 (1989).

[21] R. Sousa and R. Padilla, *A mutant T7 RNA polymerase as a DNA polymerase.* The EMBO journal **14**, 4609 (1995).

[22] M. L. DePamphilis and S. D. Bell, *Genomes,* in *Genome duplication* (Garland Science, 2011) pp. 4–10.

[23] G. M. Cheetham, D. Jeruzalmi, and T. A. Steitz, *Structural basis for initiation of transcription from an RNA polymerase-promoter complex.* Nature **399**, 80 (1999).

[24] M. Rong, R. K. Durbin, and W. T. McAllister, *Template Strand Switching by T7 RNA Polymerase,* Journal of Biological Chemistry **273**, 10253 (1998).

[25] A. Lazcano, V. Valverde, G. Hernandez, P. Gariglio, G. E. Fox, and J. Oro, *On the early emergence of reverse transcription: theoretical basis and experimental evidence.* Journal of molecular evolution **35**, 524 (1992).

[26] M. Citartan, T. H. Tang, S. C. Tan, C. H. Hoe, R. Saini, J. Tominaga, and S. C. B. Gopinath, *Asymmetric PCR for good quality ssDNA generation towards DNA aptamer production,* Songklanakarin Journal of Science and Technology **34**, 125 (2012).

6

dsDNA REPLICATION BY SYNTHESIZED PROTEINS OF THE PHI29 VIRUS IN THE PURE SYSTEM

We propose the DNA replication mechanism by the bacteriophage phi29 proteins, DNAP (p2), TP (p3), SSB (p5) and DSB (p6), as a potential candidate for the DNA replication module in the minimal cell. We have designed templates encoding for two replication proteins flanked by ca. 200 bp of the right and left end of the phi29 genome containing the origins of replication (oriLR-p2-p3 and oriLR-p5-p6). The replication activity of the PURE system synthesized phi29 DNAP and TP is confirmed by imaging the samples after a purification protocol on a neutral gel and by analyzing the band intensities of the DNA of interest at different time-points. The two synthesized proteins are also capable of amplifying the ca. 20-kb phi29 genome. This minimal replication system generates large amounts of side products in absence of the accessory proteins (p5 and p6). Co-expression of the four replication proteins reduces the side product formation in the amplification reaction of the phi29-genome. The groundwork is laid out for further investigations and suggestions to optimize conditions for DNA replication in the PURE system are discussed.

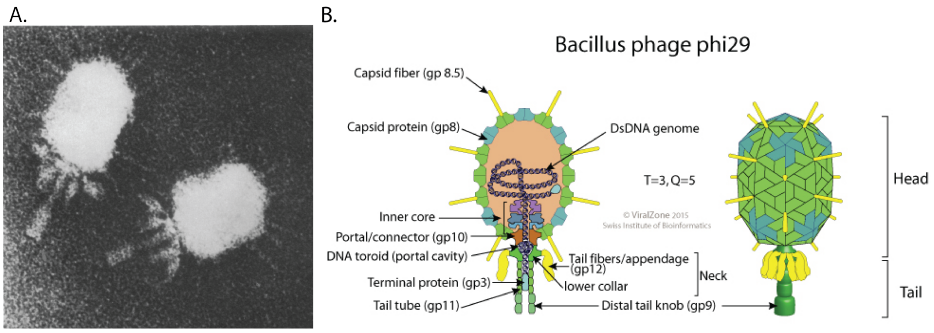


Figure 6.1: A) Micrograph of phi29 virion, adapted from [1], original from [2]. The dimensions of the head are in the order of 50 nm. B) Schematic of the structure and geometry of the phi29 virion, source: Expasy Viralzone [3].

DNA replication is at the core of information transfer, vital to all propagation of life. Nature invented more than one type of replicon, with an increasing order of complexity from viruses to prokaryotes to eukaryotes. Therefore, to opt for a minimal DNA replication system, as discussed in Chapter 4, viral life forms serve as a first source of inspiration. In this chapter, we will further dive into the mechanism of the symmetrical mode of replication that is performed by the phi29 virus. The aim of this research is to implement, analyze and evaluate this type of replicon in the minimal cell framework.

6.1. THE PHI29 VIRUS

The phi29 virus (its virion is shown in Figure 6.1) is a member of the podoviridae family[1] and a lytic phage of *Bacillus subtilis*. *B. subtilis* is a gram-positive bacterium that is naturally a soil organism but also adapted to have an intestinal life cycle after having entered the gut microflora as a spore [4]. The phi29 virus masters dealing effectively with the sporulating bacteria [5]. The common viral strategy of lysogeny is for the phi29 however not an option; the linear double-stranded 19.3 kb genome of the phi29 has terminal proteins covalently attached to its ends. Its strategy is to wait for the right moment of germination and thus inhibits transcription of its genes in the early phase of sporulation by binding the spore initiator proteins produced by its host.

The terminal proteins are the solution for the phi29 virus to the 'end-replication problem' or so called termination paradox as described in the previous chapter. Briefly, any DNA polymerase can only polymerize deoxynucleotides in the 5' to 3' direction and needs a primer in the form of a 3'-hydroxyl group to attach the first nucleotide. If the termination paradox is not solved, it means a few nucleotides at the 5'-end of the genome are not replicated, resulting in loss of information at each generation. We have already seen different solutions to this problem, such as utilizing RNA as a primer or circularization of the genome. Because some amino acids contain hydroxyl groups in their side

[1] Charasteristics: dsDNA genome, icosahedral head, short tail. Another example of this family is the T7 bacteriophage.

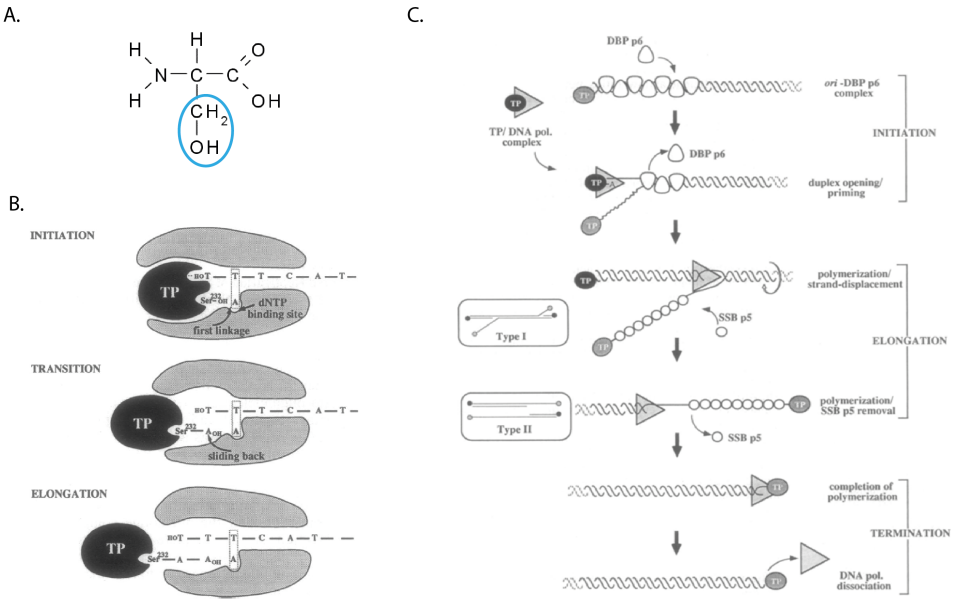


Figure 6.2: DNA replication of the phi29 genome is a protein-primed process and proceeds in a symmetric mode, where both ends serve as an origin of replication. A.) The hydroxyl group (blue circle) in the serine residue in the TP serves as a primer. B.) The sliding back mechanism, adapted from [6]. C.) Besides the TP and the DNAP, two other proteins are necessary for successful DNA replication *in vivo*: a SSB protein that protects the displaced single-stranded DNA and a DSB protein that facilitates the initiation reaction by aiding the opening of the DNA helix. Adapted from [7].

chains, such as the serine, threonine or tyrosine residues, it is possible to prime DNA synthesis with an amino acid inside a protein. This is the strategy of the phi29 phage. The protein-primed DNA synthesis is an elegant resolution in its simplicity and it represents one of the main reasons we are interested in the DNA replication mechanism of the phi29 virus for the minimal cell.

6.2. DNA REPLICATION MECHANISM OF THE PHI29 VIRUS

The phi29 genome encodes for its own DNA replication proteins that are transcribed by the host RNA polymerase in the early phase of infection. The replication of the phi29 genome is a protein-primed process and has a symmetrical mode of replication. The serine residue (Figure 6.2A) at position 232 of the terminal protein (TP, p3) provides the hydroxyl group necessary for incorporation of the first deoxynucleotide catalyzed by the phi29 DNAP (p2). The TP and DNAP form a 1:1 complex in solution, an interaction stimulated by the presence of ammonium ions. The TP-DNAP complex recognizes the origins of replications at the termini of the genome. Part of this origin is a 6-bp terminal inverted repeat of the sequence: 5'-AAAGTAA-3'. Initiation of replication occurs at the second nucleotide, after which the TP-dAMP complex slides back to basepair the

3'-terminal nucleotide of the template as depicted in Figure 6.2B [8]. The DNAP dissociates from the TP after synthesis of ten nucleotides resulting in DNAP-DNA interactions and elongation of the newly created DNA primer till the other end of the genome [9]. As the DNAP possesses a strand displacement activity, the replication complex does not need a helicase to separate the DNA helix. Aside TP and DNAP, two accessory proteins stimulate the DNA replication process, namely a single-stranded DNA binding protein (SSB) and a double-stranded DNA binding protein (DSB). The p6 DSB facilitates the local unwinding of the DNA at the origin. The p5 SSB coats the ssDNA and protects it from degradation until it is complemented into dsDNA. The concerted functioning of the four replication proteins is schematically depicted in Figure 6.2C [7]. Because the initiation of the DNA replication reaction can start at both origins (replicative intermediate Type I in Figure 6.2C), the two DNAPs approach each other on the template and will meet at one point during the replication process. During this collision, both DNAPs remain bound to their template strand resulting in replicative intermediate of type II. The DNAPs dissociate from the DNA once polymerization of the strand is completed. The process of phi29 DNA replication is called symmetric because it involves two replication origins and continuous synthesis of two DNA strands is carried out by two units of a DNAP.

In the remainder of this section, we will review important properties of the DNAP, TP, SSB and DSB proteins and of the developed *in vitro* DNA amplification system based on these proteins. We will then explain the specific aims of our research in the next section.

DNAP, p2

Besides the utilization of a protein as a primer, the main catalytic properties of the phi29 DNAP are the strand displacement ability that is coupled to the polymerization process, the high processivity and proofreading activity [10]. The processivity of the DNAP was found to be more than 70 kb on a primed ssDNA plasmid as template [11]. The strong binding of the phi29 DNA polymerase to single-stranded DNA is probably the reason for its high processivity. The DNAP binds to both ssDNA and dsDNA, though binding to ssDNA is favoured and more resistant to salt [11]. Once bound to a primer-template junction, both the strand displacement activity and processivity are not affected by conditions that increase the stability of the DNA helix, even not at 0 °C [2].

The 3'-5' exonuclease domain of the phi29 DNAP is physically separated from the polymerization domain (C-terminal region) and its activity satisfies the biochemical criteria of the proofreading mechanism[3]. The DNAP prefers degradation of ssDNA over dsDNA, though higher temperatures (37 °C) significantly increase degradation of dsDNA in absence of dNTPs [10].

The fidelity of the polymerization domain alone is between 10^5 and 10^6 and is increased 1-2 orders of magnitude further by the proofreading activity of the DNAP [12]. Because of its high fidelity the phi29 DNAP is widely used in the sequencing and single-cell genomics field. Examples are the TempliPhi™ and GenomiPhi™ DNA Amplification Kits

[2] It is therefore important to stop the DNA polymerization reaction of an experiment chemically, by either depleting the free magnesium ions or inhibiting the DNAP itself.

[3] These cirteria are: 1) release of dNMPs on ssDNA as the optimal substrate, 2) preferential excision of a mispaired primer terminus, 3) physical association with the DNA polymerase, both acting coordinately to enhance the fidelity of DNA synthesis [10].

(Amersham BioSciences), which rely on multiple-primed rolling circle amplification or linear amplification[4].
The ability of the DNAP to perform strand displacement coupled to the polymerization process reduces the amount of molecules in the replication fork is to a minimum. Hence, it is the second main reason we opt for this system as a candidate for genome replication in the minimal cell.

TP, P3

The TP has an elongated three-domain structure, a disordered N-terminal domain that has binding capacity to DNA, an intermediate domain that confers specificity to the interaction with DNAP and a C-terminal domain that contains the priming residue [13]. The parental terminal protein aids in recruiting the TP-DNAP complex to the replication origin [9]. The two parental TPs are essential for replication of the phi29 genome *in vivo*[14]. A phi29 genome containing a TP only at one end was not replicated inside *Bacillus subtillis* protoplasts, which suggests that the strand displaced TP-DNA by itself is not a template for replication and that initiation normally occurs at non-template strand before the strand is fully displaced.
It has been shown that fusion of proteins at the TP N-terminus may not impair its function, which can be exploited for possible applications such as labeling the TP with a fluorescent protein to localize it inside cells [15].

SSB, P5

The SSB is highly abundant in infected cells and protects the displaced ssDNA from degradation by host nucleases, while preventing non-productive binding of replication proteins to the ssDNA. The SSB has a relatively low binding efficiency (K_{eff}=10^5 M^{-1}) and cooperative binding mode, covering 3-4 nucleotides per protein monomer [16]. The cooperative nature of the SSB is responsible for the ability to displace oligonucleotides that are hybridized to a ssDNA template and to eliminate secondary structure of the displaced ssDNA in the replicative intermediates of phi29 genome replication. The helix-destabilizing activity of the SSB can increase the DNA elongation rate of strand displacement reactions by promoting helix-opening in front of the DNAP [16]. Importantly, binding of the SSB to ssDNA has shown to prevent strand switching by the phi29 DNAP in *in vitro* phi29 genome replication and rolling circle amplification experiments [17, 18].

DSB, P6

The p6 protein is the most abundant protein in infected cells, with a concentration around 1 mM [19]. At such concentrations, the p6 associates into oligomers. It binds to dsDNA through the minor groove as one dimer per 24 bp [20]. High affinity binding sites are present at the ends of the genome (for example the sequence 35 to 58 bp from the left end) and although patterned binding starts to become weaker from ca. 200 bp from the end, the p6 proteins can coat the whole viral genome and functions as a histone-like protein [21]. The p6 protein recognizes structural features like bendability and it strongly distorts the complexed DNA, a process that stimulates the incorporation of the

[4] see also Chapter 4

first nucleotides to the TP-dAMP initiation complex [22]. Interestingly, the correct location of the p6 relative to the genome end is required, as a phase change was found to completely abolish the ability of p6 to promote the initiation reaction [20].
Another function of the p6 protein *in vivo* is the repression of some early promoters and, together with protein p4, it is also involved in activation of a late promoter[5].

In vitro DNA AMPLIFICATION SYSTEM

The *in vitro* DNA amplification system based on the four purified replication proteins was developed by the group of Prof. Salas and was first described for amplification of the phi29 genome [7]. For understanding their most important findings and interesting features of this system that will be relevant for our research, key figures are reproduced in Figure 6.3. First of all, it should be noted that the TP and DNAP alone are capable of amplifying the phi29 genome, although with low efficiency and only when starting from high amount of initial template (500 ng). In Figure 6.3A, the graphs on the left show the increase in DNA synthesis for different initial amount of phi29 genome input DNA when the minimal system is supplied with either the p5 and p6 separately, or simultaneously. The complete system described in this paper has the following composition: in a 10 μl reaction, purified proteins are present as 15 ng of preformed TP-DNAP complex, 20 ng free TP, 10 μg p6, 8 μg p5, different amounts of the phi29 genome containing the parental TPs, all in a buffer of 50 mM Tris HCl (pH 7.5), 10 mM $MgCl_2$, 4% glycerol, 1 mM DTT, 0.1 mg/ml with 80 μM of each of the dNTPs. After 1 h incubation at 30 °C the final DNA synthesized is similar for all input DNA. The amplification factor starting with 50 ng of input DNA template is around 13-fold in this experiment. The left graphs in Figure 6.3A show that increasing amounts of p5 or p6 proteins (upon the background of the other protein) leads to a close to linear increase in the amplification yield when starting from 0.5 ng input DNA.
The effect of SSB in the *in vitro* amplification process is further studied in [17]. Here, they found that the absence or low concentrations of SSB in the *in vitro* amplification system led to the generation of short products (Figure 6.3C left), with typical lengths varying between 200 and 6000 bp. Because the short DNAs were reamplified when used as input templates in a fresh amplification reaction, they concluded that the short molecules are *in vitro* replicons derived from the phi29 DNA template ends. The mechanistic explanation for this is that the displaced strand serves as a competing template for the DNAP when it is not bound by SSB. The potential template switching event by the DNAP will continue until the end of the displaced strand, generating a new double stranded replication origin. As schematically depicted in Figure 6.3C, initiation on this origin would lead to a short molecule containing two identical and functional origins of DNA replication; a palindromic molecule that can be replicated further. Because these DNAs are much shorter than the full-length phi29 genome, their duplication time is shorter and they are in direct competition for the replication resources. Duplication of TP-DNA requires the steps of protein-primed initiation and DNA elongation (that is length-dependent) that have different time durations. When the elongation time is shorter than the initiaton time, which is the case for short DNAs, one such molecule can easily outnumber its full-

[5] Is it possible that the coating of the DNA by the p6 might interfere with transcription by the T7RNAP in our designed constructs?

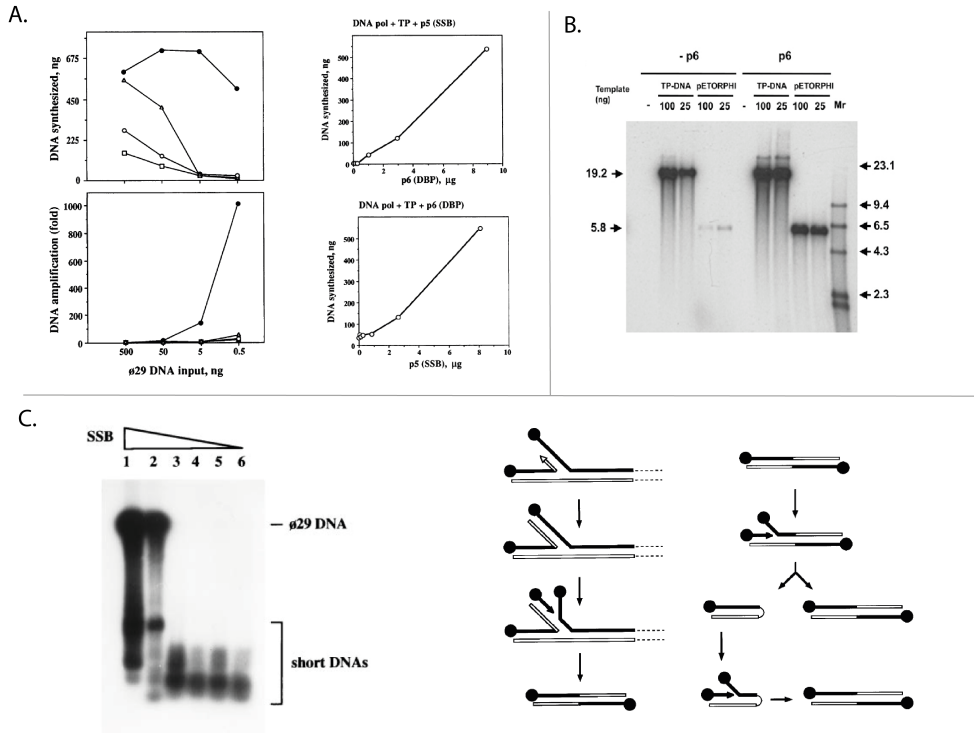


Figure 6.3: The *in vitro* DNA amplification system is based on the four phi29 replication proteins. Important findings of literature are presented A.) Adapted from Figure 3 in [7]. Left: DNA synthesized (above) and DNA amplification (below) *in vitro* as a function of input DNA with different combinations of phi29 DNA replication proteins by the minimal system of TP and DNAP (squares), supplemented with p6 (open circles) or p5 (triangles) individually or simultaneously (black circles). Right: The requirement of proteins p6 and p5 for DNA amplification is shown for an input DNA of 0.5 ng phi29 genome. B.) Adapted from Figure 2 in [15] The heterologous pETORPHI DNA template (containing 191 bp from the left and 194 bp from the right end of the phi29 genome) is an efficient template for DNA synthesis dependent on the presence of p6. C.) Adapted from Figures 1 and 6 in [17]. Left: Effect of SSB on the size of in *in vitro* amplified DNA using 0.5 ng of phi29 genome. Short DNAs accumulate in the reaction with decreasing amount of SSB present (below 2 μg). Right: Proposed model for the generation and replication mechanisms of the short DNA products that are hypothesized to be palindromic TP-DNAs.

length competitors in a short time period (for a detailed analysis see [17]). Interestingly, when increasing the initation time by having only subsaturating concentrations of the initiator dATP molecule present, the generation of short DNAs was severely limited.

The *in vitro* amplification method was extended for its general use by introducing the design and amplification of heterologous DNA templates [15]. By flanking a DNA sequence of interest with ca. 200 bp (the strong binding region of p6 proteins) of the left and right ends of the phi29 genome, in principle any template can be amplified by the phi29 system. In Figure 6.3B, the amplification of the 5.8 kb pETORPHI template is shown to highly depend on the presence of the p6 protein. The minimal amount of template required for amplification was 10 ng for both pETORPHI and the phi29 genome (the protein concentrations in this system are at least 2.5-fold lower, see Materials and Methods for the full composition), an interesting observation since the minimum mass input of DNA templates does not correspond to the same number of DNA molecules. Starting with 25 ng of template, the average amplification factors after 2 h of incubation at 22 °C were 83 ± 19 and 30 ± 7 for the phi29 genome and pETORPHI, respectively. A template containing on both sides 68 bp from the left side of the genome, pETORPHI68L, was also amplified with a yield that was 30% less. This research showed that the terminal protein is not necessary for *in vitro* replication of templates, though a 5' phosphate group is required.

The highly inspiring mechanism of phi29 DNA replication that is performed by only two to four proteins and the promising results found for the *in vitro* developed amplification system are the motivation for the project described in this chapter.

6.3. RESEARCH AIM

The aim of this research is to investigate if the DNA replication system based on the phi29 proteins is compatible with our biosynthesis platform for the minimal cell, the PURE system. We will investigate if the expressed phi29 proteins are functional in this system. Because we know from literature that in some conditions the minimal replication system consisting of the DNAP and TP was able to perform DNA replication, our first aim is to test if the expressed DNAP and TP can replicate their own encoding DNA template, thus completing a full cycle of the central dogma. As the near future of the minimal cell research requires co-expression of a multitude of genes, another aim is to replicate a longer DNA template, such as the ca. 20-kb phi29 genome. A third aim is to optimize conditions to achieve a high yield of DNA replication in the PURE system. This will be a challenging task since there are many parameters involved. By implementing the phi29 DNA replication machinery in an *in vitro* transcription and translation system we hope to learn new facets of the working mechanisms of the DNA replication by the phi29 proteins when confronted to other genome processing reactions.

For more details about the minimal cell context of this research we refer the reader to Chapter 1, and for dicussions on related research we refer to Chapter 4 [23, 24].

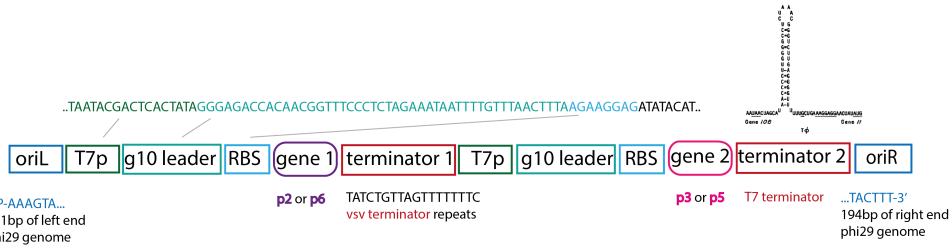


Figure 6.4: Schematic picture of the elements in the designed DNA constructs for replication. The hairpin formation of the T7 terminator is adapted from [26].

6.4. DESIGN OF THE DNA TEMPLATES ENCODING FOR THE REPLICATION PROTEINS

An important aim in this project is to replicate the double stranded DNA that encodes for its own replication proteins. Our first goal was therefore to construct a DNA template that contains the genes for the terminal protein and the DNA polymerase, plus the origins of replication. We placed both genes under their own promoter, to tentatively achieve similar levels of transcription and subsequent translation. The schematic picture of the double gene construct is shown in Figure 6.4. A design with multiple gene-operons is another strategy, which would require adequate tuning and understanding of the mRNA folding and strengths of the ribosome binding sites. Because a recent study of tricistronic gene expression in the PURE system reported a biased expression towards high protein production of the first gene, we did not opt for this approach [25].

The two origins of replication and the gene sequences of the phi29 genome are retreived from the NCBI Genbank Reference Sequence NC0110481. To ensure high protein expression in the PURE system, coding and non-coding signal sequences are optimized. The sequences for the proteins are gene optimized for expression in *E. coli* by the OptimumGene$_{TM}$ algorithm of GenScript, which includes optimization of e.g. codon usage bias, GC content and mRNA secondary structure. We chose to perform gene optimization for *E. coli* because the PURE system is based on the *E. coli* translation system including the pool of tRNAs [6].

The sequence from T7 promoter to START codon in all our constructs is the same as in gene 10 from the T7 genome [7]. This T7 promoter is the strongest *in vivo* consensus promoter of the class III genes of the T7 virus, which represents the genes for virion structure and assembly, all requiring high expression levels. The major protein expressed after infection of T7 in *E. coli* is the capsid protein, gene 10. Expression of foreign genes with the g10-leader sequences are around 40-fold greater than the consensus *E. coli* ribosome binding site(RBS)[8] [27]. The upstream sequence of the RBS forms a secondary structure that serves as an affinity ligand for the small subunit ribosomal proteins. For

[6] supplier Roche extracts the tRNA from E. coli MRE 600.
[7] To be exact: nucleotide position 22887 to 22966 in GenBank entry NC001604.1 .
[8] Or Shine-Delgarno sequence (SD).

this reasons, the gene 10 sequence of the T7 genome is the basis for the widely used pET, pRSET and piVEX expression vectors in *E. coli*-based (cell-free) systems [28]. Inserting genes at the START codon of the gene 10[9] is equivalent to inserting genes with the Nde I restriction enzyme into the pET-3 a, b and c vectors or utilizing Nco I sites in the pET-3d vector. The commonly used pRSET vectors contain part of the g10-leader after the START codon, resulting in addition of amino acids at the N-terminus of the synthesized protein. Because the N-terminal of most of the replication proteins serve a protein or DNA interacting function, this would not be desirable.

To produce mRNA encoding a single gene, termination signals for the T7 RNAP are placed after the STOP codon of the genes. In this design we employ two different terminator sequences, the natural T7 terminator and an artificial terminator. The T7 genome has only one T7 RNAP terminator, Tphi, that is positioned between gene 10B and gene 11 to ensure production of large amounts of the capsid protein. The termination is not completely efficient, as read-through is necessary for transcription of gene 11 and 12 that do not contain their own promoters [26]. The sequence of the terminator forms a stable stem-loop structure followed by a run of U's, destabilizing the RNA:DNA hybrid and resulting in termination. Apart from this class, there is a second class of terminators that does not generate an evident structure in the RNA but can pause the T7 RNAP. The 7-bp sequence ATCTGTT (5' to 3' in the coding strand) is responsible for the pause, and release of the RNAP from the template is favored by a stretch of U's in the 6 to 8bp downstream [29]. This sequence was first identified to terminate the T7 RNAP in a study where they aimed to *in vitro* transcribe the cloned human preproparathyroid hormone (PTH) gene, and also when transcribing the ssRNA of the vesicular stomatitis virus (vsv). We utilized the class II terminators as described in two papers of Prof. Foster's group [30, 31]. The vsv-terminator is only 18bp (TATCTGTTAGTTTTTTTC) and has by itself a termination efficiency of 53-62% *in vitro* [30], which is close to the reported termination efficiency of Tphi of 70%. A tandem repeat of the vsv-terminator, separated by a spacer of 16 bp, even reached 90% termination effiency of T7 RNAP transcription *in vitro*[31]. It should be possible to reach complete termination if more repeats are inserted, which will be important for more complex multiple gene constructs. The 52-bp tandem repeat of the vsv-terminator is in our design implemented between the two genes. We chose to have the two different terminators for functional reasons regarding the assembly PCR, that is described in the Materials and Methods.

In Table 6.1, an overview of the constructs used in this research is reported, of which most of them are displayed on the DNA gel in Figure 6.5. The constructs containing origins have a 5'-phosphate group introduced during the PCR. The constructs that do not have origins of replication are otherwise of the same design. The complete sequences are available upon request and will be published in the near future.

A second aim of this study is to copy a long DNA template, to evaluate the performance of the phi29 DNA replication mechanism to amplify genome with lengths relevant for a minimal cell. The genome of the phi29 virus is the ideal choice for this pur-

[9] The sequence of interest is ordered at the company GenScript, circumventing cloning steps.

Table 6.1: DNA constructs of the phi29 proteins

Construct	Length (bp)	Origins of replication	Comments
Terminal protein (TP, p3)	932	No	MW of protein: 31 kDa
DNA Polymerase (DNAP, p2)	2k	No	MW of protein: 66 kDa
p5	508	No	MW of protein: 13.3 kDa
p6	447	No	MW of protein: 12 kDa
oriLRp6	835	OriR194, OriL191	T7 terminator (T7t)
oriLR-p2-p3	3212	OriR194, OriL191	tandem-vsv and T7t
oriL68-p2-p3	2959	OriL68 at both ends	tandem-vsv and T7t
oriLR-p6-p5	1382	OriR194, OriL191	tandem-vsv and T7t
Phi29-genome	19282	Yes	parental TPs at 5'-ends

pose, and it was kindly provided to us by our collaborators Prof. Margarita Salas and Dr. Mario Mencia from Madrid. The phi29 genome is visualized on agarose gels in Figure 6.6. We observe that the TP-phi29-genome separates into two distinct bands on a neutral agarose gels, of which one of them remains close to the well. After heat denaturation treatment of the genome, only one band of ssDNA is visible. This indicates that the upper band is the phi29 genome in a different state or folded structure. On a denaturing agarose gel two distinct bands are likewise observed when loading high amounts of DNA (more than 190ng). Our interpretation is that partially denatured and interwoven ssDNA aggregate and then remain near the well (personal communication Salas and Mencio).

6.5. RESULTS

6.5.1. TRANSCRIPTION FROM DNA TEMPLATES

To examine the transcription efficiency of both genes from a double-gene construct we performed an in vitro transcription reaction in the PURE system without ribosomes. The catalytic units of the ribosome are RNAs, the rRNAs, that have lengths of 1.5 knt and 2.9 knt. Ribosomes were omitted in the PURE system when we aim to detect transcripts of similar size. Earlier research in our lab showed that transcription is not affected when omitting the ribosome solution [32]. In Figure 6.7 transcripts originating from the p2, p3 and the combined oriL68-p2-p3 template are analyzed on gel (right). A small amount of read-through RNA product from an RNAP that started at the first T7 promoter and did not terminate at the vsv-repeats, is present. A rough estimation based on band intensity analysis of the separate RNA species and the assumption that band intensity scales linearly with the total ng amount of RNA present, gives 20% read-through, and thus 80% termination efficiency of the vsv-repeat. Note that on this gel the resolution does not allow us to make a distinction between transcripts that were aborted at the T7 terminator or at the end of the template (when the RNAP falls off the template). To determine the efficiency of the T7 terminator in our experimental system, we analyzed the transcript of the oriLRp6 template on the left gel in Figure 6.7. The transcription by the T7 RNAP in the *in vitro* T7 Ribomax Large Scale RNA production (Promega) system and in the PURE system with ribosomes were compared. If the T7 RNAP reads through the T7 terminator, it would transcribe the extra sequence of the oriR, increasing the length of the RNA from

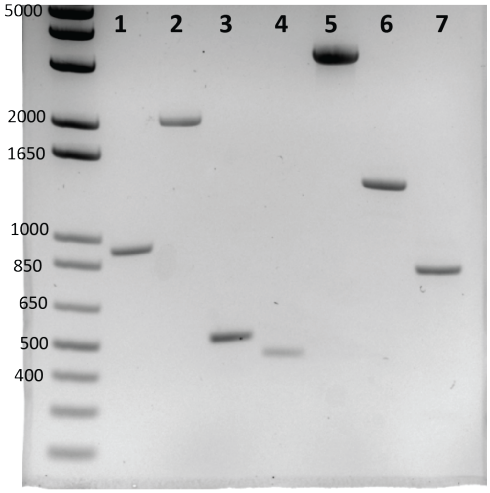


Figure 6.5: Characterization of linear DNA constructs containing the genes for the phi29 DNA replication proteins on a neutral 1.2% agarose gel. The constructs of separate genes without origin of replication are in lanes 1-4: TP (932 bp), DNAP (ca. 2 kbp), p5 (508 bp), p6 (447 bp). The constructs in lanes 5-7 contain 191 bp of the left side and 194 bp of the right side of the phi29 genome (oriLR), described in [15] and in between the DNAP and TP genes (oriLR-p2-p3, 3212 bp, lane5), the p6 and p5 genes (oriLR-p6-p5, 1382bp), or only the p6 gene (oriLRp6, 835bp). The right length of the constructs is confirmed by the reference DNA ladder in the left-most lane. Other constructs used in this study but not shown here include the oriL68-p2-p3 and the oriL24-mYFPco-LL-Spinach-T7t-oriR24.

around 400bp to 600 bp. Based on band intensity analysis, the termination efficiency of the T7 terminator in the PURE system environment is around 70% and seems to be higher than in the T7 IVT kit [10]. With this demonstration we aim to make the reader aware of the incomplete termination at termination signals for the T7 RNAP, without going further into robust numbers and statistics. As a proof of principle, we identified that the vsv-repeats show similar high termination efficiency in our experimental environment as reported in other reaction media [30, 31]. Another interesting observation is that more RNA molecules of the TP than of the DNAP gene are transcribed, although the two genes are under the control of the same promoters. We hypothesize that this might be a consequence of the position of the gene on the template, where the TP is the second gene. RNAPs released from the DNA are already in close proximity to the promoter of the second gene, increasing the chance to initiate transcription there. A strategy to avoid bias in protein expression could be to add one or two extra T7 promoters before the first gene at the 5'-terminus which has proven to be a successful strategy in the *in vivo* translation of pentameric constructs [31].

[10] The estimated termination efficiency from this gel in the IVT kit is 56%

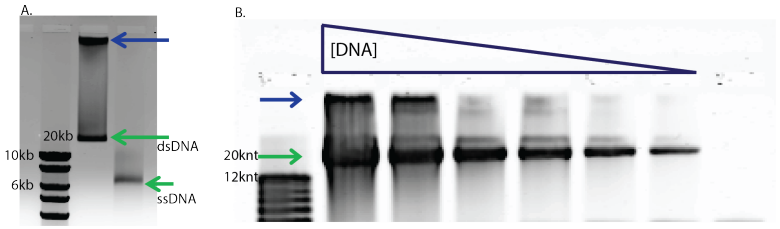


Figure 6.6: The phi29 genome with TP covalently attached to its 5'-ends visualized on agarose gels. A) neutral agarose gel with 50 ng of TP-phi29-DNA, in presence of 0.1% SDS, without (1st lane) and with 2.5 min heat-denaturation at 95 °C (2nd lane). B) the TP-phi29-DNA on a 0.7% alkaline agarose gel with decreasing concentrations from left (300 ng) to right (10 ng).

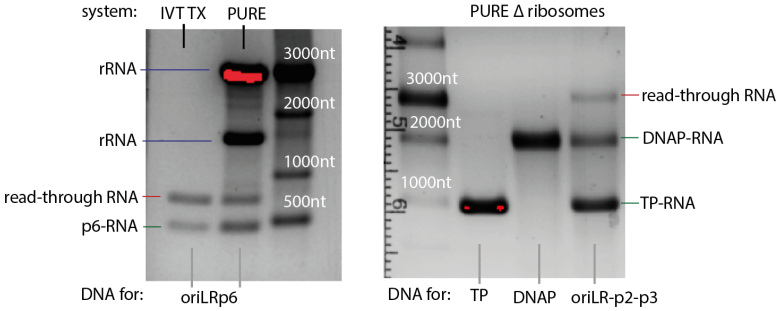


Figure 6.7: Transcription of the oriLRp6 (left) and the p2, p3 and oriLR-p2-p3 DNA (right) constructs imaged on agarose RNA gels shows that the T7 RNAP partly reads through the termination signals of the T7 terminator and the vsv-repeats.

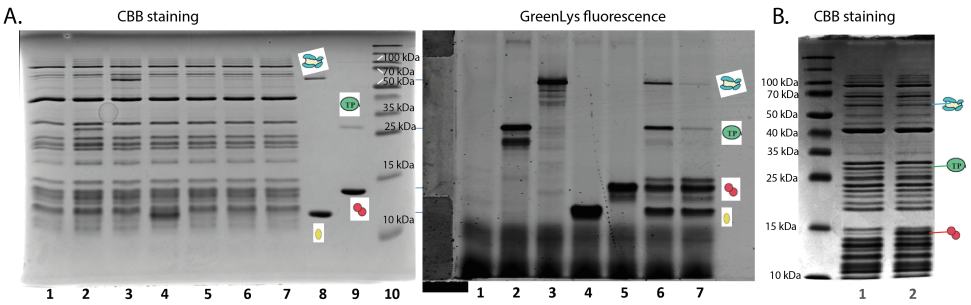


Figure 6.8: Expression of phi29 DNA replication proteins in the PURE system after 3h of incubation at 30 °C is visualized on a polyacrylamide protein gel. A.)The 15% protein gel is imaged in two conditions: after staining with CBB (left), displaying all the PURE system proteins present in the sample, and analyzed on a fluorescence scanner to visualize the incorporated labeled amino acid, detecting only the synthesized proteins in the sample (right). Lane 1: reference PURE system without DNA, expression of monocistronic genes: TP (lane 2), DNAP (lane 3), p5 (lane 4), p6 (lane 5), expression of the double gene constructs oriLR-p2-p3 and oriLR-p6-p5 in equimolar (lane 6) and experimental conditions (lane 7, excess of oriLR-p6-p5). Purified proteins serve as a reference of the MW in lane 8 (180 ng p2 and 2 ug p5) and lane 9 (180 ng p3 and 2 ug p6). Lane 10 contains a protein ladder with indicated MW. B.) On a 12% protein gel the expression of p6 is clearly visible with CBB staining by comparing the samples of co-expression from oriLR-p2-p3 in lane 1 with the co-expression pattern of the 3 proteins from the p6 and oriLR-p2-p3 templates in lane 2.

6.5.2. EXPRESSION OF THE P2, P3, P5 AND P6 DNA REPLICATION PROTEINS IN THE PURE SYSTEM

We verified that the full-length proteins are expressed from their DNA templates in the PURE system. For the detection of synthesized proteins a fluorescently labelled lysine was supplemented to the PURE system, which allows us to visualize protein products on a fluorescence scanner without the PURE system background. We also performed Coomassie Blue staining of the gels to compare the size of the translation products with that of the purified proteins. The correct molecular weight (MW) for the p2, p3 and p5 proteins were confirmed in the protein gel of Figure 6.8. The observed MW of the p6 protein was higher than expected, though it migrates at the same height as the purified p6 protein. In Figure 6.8B the presence of the p6 band on the CBB stained gel is more visible. As an indication for the amount of protein production we compare the band intensities of the purified p2 and p3 proteins with the expressed p2 and p3 proteins, and find a lower limit of 0.5 μM of p2 and 1.2 μM of p3 proteins produced when expressed separately under these experimental conditions. Succesful co-expression of the four full-length phi29 DNA replication proteins in the PURE system is hereby confirmed. The next step is to investigate their activity.

6.5.3. SYNTHESIS OF A TRANSCRIPTIONALLY ACTIVE DNA TEMPLATE BY THE PHI29 DNAP

To verify the activity of the DNAP, we designed an assay based on the read-out of fluorescent reporters for RNA and YFP, which are only produced if the active DNAP has elongated a primer-template junction (PTJ) to create dsDNA from ssDNA. This was a straightforward assay considering the history of our lab (see Chapters 2 and 3). The ssDNA for

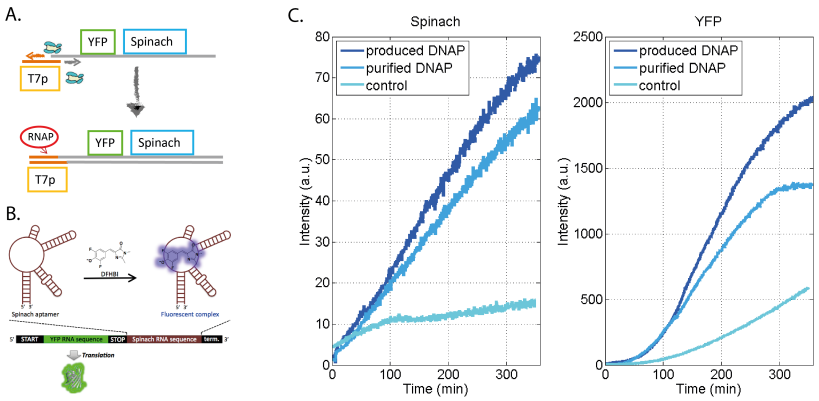


Figure 6.9: Activity assay of synthesized DNAP in the PURE system by the read-out of fluorescent reporter products. A.) Schematic of steps when the synthesized DNAP is active in this assay B.) Spinach fluorescence detection of mRNA is based on the binding of DFHBI to an aptamer at the 3'-untranslated region of the YFP-Spinach mRNA. C.) The capability of synthesized DNAP to elongate a PTJ is compared with purified phi29 DNAP (purDNAP, NEB) activity. The RNA production (left) confirms the ability of the DNAPs to polymerize dNTPs efficiently into a transcriptionally active dsDNA product. The corresponding YFP fluorescence kinetics are shown on the right panel.

this assay is synthesized by an asymmetric PCR and encodes the mYFPco-LL-Spi-T7t sequence without the T7 promoter. A ssDNA oligo containing the T7 promoter sequence is then hybridized to the coding ssDNA (Figure 6.9). Both the ssDNA-oligo complex and the DNA encoding for the phi29 DNAP are supplemented to the PURE system together with dNTPs and the DFHBI chromophore. The consecutive reactions that will take place then are:
– DNAP is synthesized by the PURE system components from its DNA template
– DNAP binds to the primer-template junction of the hybridized T7 promoter oligo and the ssDNA
– DNAP polymerizes dNTPs to create a ds-promoter and ds-template
– T7RNAP can now initiate transcription from the ds-promoter to produce mRNA that encodes for the YFP and the Spinach aptamer
– the aptamer binds DFHBI to form the fluorescent Spinach complex (read-out) and the mRNA is translated in YFP (secondary read-out)

In Figure 6.9C it is confirmed that the purified DNAP as well as the synthesized DNAP can indeed elongate a PTJ. Note that there is almost no time-delay before the Spinach signal - and thus transcription - starts, which we attribute to an efficient expression of the DNAP within the few minutes it takes to transfer the sample to the measurement instrument at room temperature. The small production of RNA and protein in the control sample that does not contain DNAP but only ssDNA, likely comes from small traces of transcriptionally active dsDNA that are carried over from the asymmetric PCR. This DNAP activity assay is complemented with other conventional assays described in the

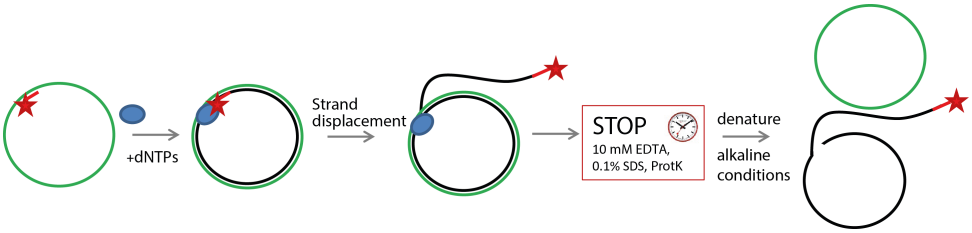


Figure 6.10: Schematic of a RCA by a DNAP with strand displacement activity. A (fluorescent) primer is hybridized to a ssDNA plasmid (here, ssM13) to create a PTJ where the DNAP can bind and initiate polymerization of dNTPs. Once it has completed one circle, the DNAP displaces the original primer-strand and continues with polymerizing dNTPs until the reaction is quenched. The DNA is denatured in alkaline conditions before imaging on an alkaline agarose gel.

next paragraphs [11].

6.5.4. ROLLING CIRCLE AMPLIFICATION BY THE PHI29 DNAP

A rolling circle amplification (RCA) reaction is a primer extension assay on a ss circular DNA, as described in Figure 6.10. In this assay, samples of the amplification reaction by the synthesized DNAP are quenched at different time-points and the length of the synthesized DNA is analyzed on a denaturing agarose gel as displayed in Figure 6.11A. It can provide insights into three important properties of DNAPs[12] : strand displacement activity, processivity and average polymerization rates. DNAPs that do not possess a strand displacement activity will only complete one circle and thus produce template of the same size as the input ssDNA. A template challenger assay, in which DNAPs that initiated on a fluorescent primer and fall off their templates will start on a new PTJ having a non-fluorescently labeled primer, provides information on the processivity of the studied DNAP. In principle, it should also be possible to deduce the average rate of polymerization by the DNAP, by comparing the lengths of synthesized products with a reference DNA ladder. The first RCA of the phi29 DNAP described in literature revealed the strand displacement activity and the high processivity, namely more than 70kb on the singly primed M13 DNA [11]. Our motivation to perform this assay is to confirm activity of the synthesized DNAP in the PURE system at different temperatures. As we observe in Figure 6.11B, the synthesized DNAP is active when produced at 22°C and 30°C, whereby the latter condition favors a higher DNA production. In the next chapter we will describe a few more RCA assays aiming to investigate how the activity of the purified phi29 DNAP is altered from ideal conditions compared to the environment of the PURE system which contains millimolar concentrations of rNTPs that compete with dNTPs for the active site of the polymerase.

[11] At lower temperatures than 37 °C, this assay was not convincing to confirm activity, likely due to temperature effects on ssDNA folding.
[12] in presence of dNTPs. In the absence of dNTPs a 3'-5'exonuclease activity could be observed

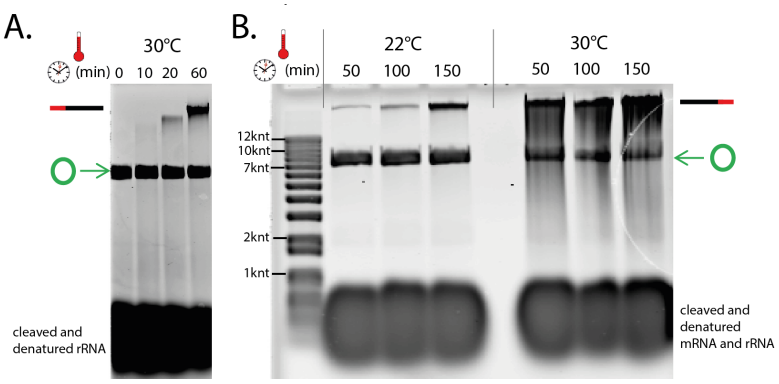


Figure 6.11: RCA of the ssM13 by the phi29 DNAP in the PURE system visualized on alkaline agarose gels. A) an example of RCA by purified DNAP at 30°C at different time points. At t = 0min only the ssM13 (7400 nt) is visible. At 60 min the elongated primer, far exceeding the length of the ssM13 is clearly visible. B.) RCA by synthesized DNAP from its DNA template at 3 time points and two temperatures. Comparison of the amount of synthesized DNA product at 30°C and 22°C, shows that the combination of DNAP expression and elongation reactions is faster at 30°C.

6.5.5. ACTIVITY ASSAYS BASED ON INCORPORATION OF FLUORESCENTLY LABELED NUCLEOTIDES

The traditional technique to study dNTP incorporation by DNAPs is to supplement the dNTP solution with a small percentage of radioactive labeled dNTP, like the alpha-^{32}PdATP. In the literature about DNA replication by the phi29 proteins, quantification of the synthesized DNA is measured as the total amount of incorporated dNTPs, calculated from the amount of radioactivity of the excluded volume after filtration through a spin column [7]. As our department did not have an easy access to a radio-isotope facility, an alternative was to replace the radioactively labelled dNTPs with fluorescently labelled nucleotides. The disadvantage of this method is that the introduced fluorescent label is bulky and might hinder or alter the incorporation efficiency by the DNAP. For example, the proofreading activity may excise a modified nucleoside after incorporation, slowing down the polymerization. Our approach is to utilize fluorescently labelled nucleotides in our initial experiments to confirm activity of the replication proteins and to scan different ranges of parameters (DNA concentrations, dNTP concentrations, temperature, etc.). When replication efficiency of the system of interest is high enough so that the reaction products can be observed on a gel with intercalating dyes, we omit the fluorescently labeled nucleotides to avoid any negative bias in the DNA replication reaction. The phi29 DNAP is commonly used in combination with fluorescent nucleotides for sequencing techniques using exclusively terminal phosphate-labeled nucleotides, demonstrating processive enzymatic DNA synthesis with thousands of consecutive nucleotide incorporation events [33, 34]. To have a fluorescent molecule incorporated in the newly synthesized strand, the nucleobase or an C molecule of the dNTP should be labelled, because the terminal phosphate labels would be cleaved off upon polymerization by the DNAP. We have utilized two fluorescently labeled nucleotides in this study,

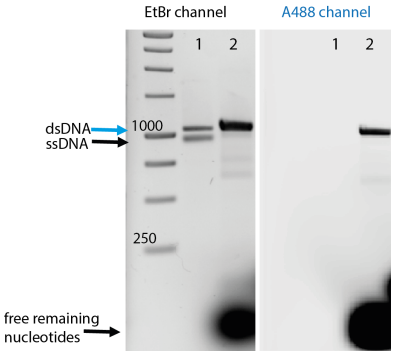


Figure 6.12: Primer extension assay with incorporation of Alexa488-dUTP by the purified phi29 DNAP. The purified ssDNA (lower band) and expected dsDNA endproduct (upper band) are the starting products of the experiment (lane 1). After hybridization of a primer to the ssDNA, the phi29 DNAP and dNTPs are added to the solution to initiate DNA synthesis. The dsDNA final product is imaged in lane 2. In the Alexa488 channel the incorporation of the modified nucleotide in the newly synthesized DNA is confirmed.

the Chromatide Alexa Fluor 488-5-dUTP (Alexa488-dUTP)[13] (Life technologies) and the Fluorescein-12-dUTP (ThermoFischer).
The purified phi29 DNAP can incorporate the Alexa488-dUTP in a primer extension assay (similar to the first step in Figure 6.9A, in the phi29 buffer instead of the PURE system) with a ratio of non-labeled dTTP : labeled dUTP of 3:1 (Figure 6.12). We kept this ratio of nucleotides for the experiments described in the PURE system.

6.5.6. INCORPORATION OF FLUORESCENTLY LABELED NUCLEOTIDES BY THE SYNTHESIZED DNAP IN A PROTEIN-PRIMED REACTION

Having confirmed the activity of the synthesized DNAP in the PURE system, the next aim is to investigate if the synthesized TP and DNAP can form a complex to initiate DNA replication on a dsDNA template. For this purpose the oriL68-p2-p3 template was expressed in the PURE system supplemented with 20 mM NH_4SO_2, dNTPs and a fluorescent dUTP. In Figure 6.13 we indeed observe that the synthesized proteins are active, as indicated by the fluorescence band at the height of the dsDNA template. Note that the observed fluorescence does not provide us with the information whether one strand or two strands are synthesized by the DNAP. Three different concentrations of input template and two differently labelled nucleotides were compared. The amount of incorporated fluorescent-dUTPs increases upon increasing the template input in the range of 10-140 ng (Figure 6.13). Both modified nucleotides produce high enough intensities on the gel to be visible. For economical reasons, the remainder of the experiments are performed with the Fluorescein-dUTPs.

[13] The Alexa488-dUTP is modified at the C-5 position of the uridine, with a net length of the spacers of 5 atoms.

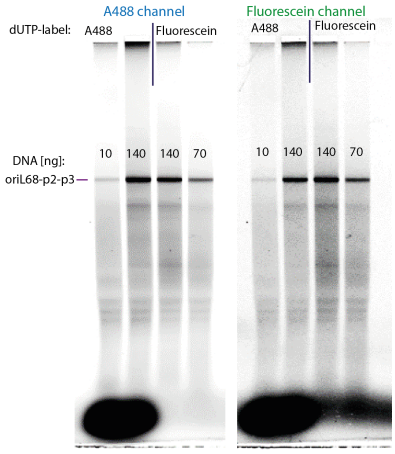


Figure 6.13: Incorporation of different fluorescently labeled nucleotides by the synthesized phi29 DNAP in a protein-primed DNA replication reaction.

6.5.7. DUPLICATION OF THE DSDNA TEMPLATE THAT ENCODES FOR THE DNAP AND TP BY ITS SYNTHESIZED PROTEINS

To study the efficiency of dsDNA replication by the synthesized DNAP and TP of its own template, we performed gene expression of the oriLR-p2-p3 in the PURE system omitting the fluorescent nucleotides. The reaction samples of different time points are analyzed on both an alkaline gel and on a neutral gel which are displayed in Figure 6.14.

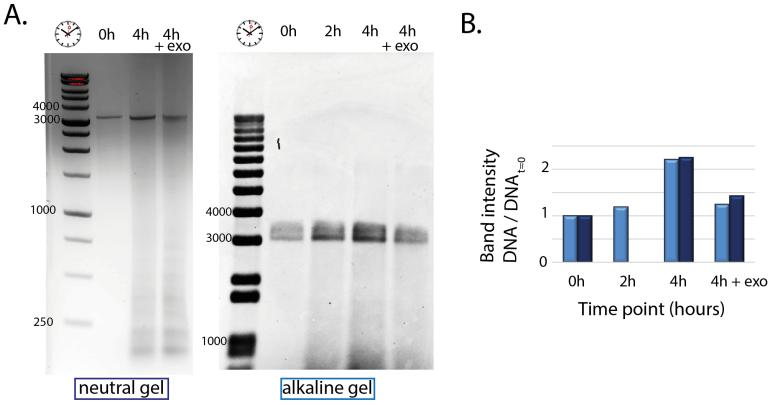


Figure 6.14: The PURE system synthesized DNAP and TP, expressed from the oriLR-p2-p3 construct, duplicate their own dsDNA template. A.) Two independent reactions are visualized on neutral (left) and denaturing (right) agarose gels. The final 4h time point is treated with lamda-exonuclease, that processively degrades DNA strands when they are not protected by a terminal protein at the 5'-end. The remaining dsDNA after this treatment is a complementary proof of DNA amplification. B.) Quantification of the intensity of the bands of the oriLR-p2-p3 DNA show a 2-fold amplification after 4h on both alkaline (light blue) and neutral gel (dark blue).

The advantage of analysis on alkaline gels is that RNA, present in the PURE system as rRNA, mRNA and tRNA, does not have to be removed by RNase treatment and column purification from the samples because it will get partially cleaved in the alkaline gel conditions. It will therefore not interfere with the DNA bands of interest. By reducing the amount of steps in the protocol this analysis is more robust, as less human mistakes can be introduced. However, the disadvantage is that the gel shows only ssDNA species and it is therefore not possible to make a distinction between the scenario of merely protein-primed strand displacement reaction events by the DNAP, or true dsDNA amplification requiring initiation at the two ends of the same DNA template. The advantage of analysis on a neutral gel is that an observed increase of the dsDNA band intensity directly proves DNA replication[14]. The experimental difficulty lies in the need for the complete removal of all RNA species prior to imaging on gel. For this reason it is also possible to visualize any by-products the reaction may produce, that would otherwise be blurred by small RNA molecules in the gel. The protocols are described in detail in the Materials and Methods section.

Analysis of the samples on gels showed a two-fold replication of the dsDNA template after 4 hours (Figure 6.14B). Moreover, we exploit the fact that the replication products have a covalently attached terminal protein to the 5'-ends to perform a control experiment. Treatment of the final time point sample with the processive lambda exonuclease should only lead to degradation of DNA that are not protected at either 5'-end by a terminal protein. This means that all the initial, non-replicated, DNA strands are degraded. On the alkaline gel, the result should be that the band intensity corresponding to the starting amount of DNA and the band intensity of DNA after lambda-exonuclease treatment band would sum up to give the band intensity of the untreated DNA at the final timepoint[15]. The interpretation of the neutral gel is more complex and depends on the presence of dsDNA having 0, 1 or 2 terminal proteins covalently attached. The amount of dsDNA that has two terminal proteins can be calculated from the band intensity on the gel. Combining the analysis of neutral and denaturing gels can shed light on the DNA replication mechanism on these templates in the near future, after obtaining more repeat experiments and statistics.
Apart from the duplication of the template of interest, we also observe small distinct side products below 500 bp. These side products do not disappear upon lambda-exonuclease treatment, suggesting that these are true replication products. Similar observations have been made for the replication of the phi29 genome in *in vitro* and *in vivo* environments [17, 35]. The side products are symmetrical dsDNAs with TPs at its termini and originate from palindromic duplication of one origin. The formation of these products is not fully understood but they depend on the specific amplification conditions and can in general be reduced by the presence of phi29 SSB or by limiting the amount of TP-DNAP complex (personal communication Salas and Mencia).

[14] This is to my best knowledge the only direct proof of dsDNA replication of the DNA of interest.
[15] Assuming the lambda-exonuclease treatment is sufficient to degrade all input DNA.

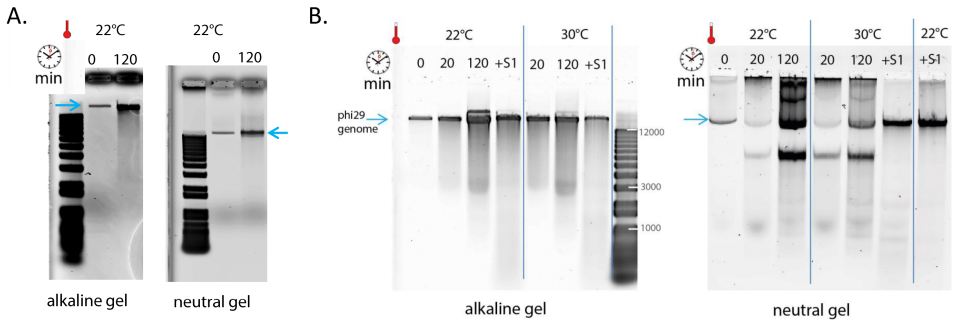


Figure 6.15: Replication of the 20-kb TP containing phi29 genome with the purified p2, p3, p5 and p6 proteins is analysed on alkaline (left) and neutral (right) agarose gels. Two experiments are shown in which amplification of the phi29 genome is visible with EtBr staining. Analysis on neutral gels shows that ssDNA molecules are generated in some experiments (in B, not in A). The amplification yield after 2h in these experiments estimated from band intensity analysis (3-16 fold) is an order of magnitude lower than reported in literature (90-fold) [15].

6.5.8. PROTEIN-PRIMED AMPLIFICATION OF THE PHI29 GENOME BY THE PURIFIED AND EXPRESSED P2, P3, P5 AND P6 DNA REPLICATION PROTEINS

As described in the introduction, an *in vitro* amplification system based on the purified p2, p3, p5 and p6 proteins was developed and optimized by the group of Prof. Salas, by investigating the amplification of the phi29 genome in different conditions. We aim to translate this knowledge to understand and trouble-shoot the DNA replication mechanism in the minimal cell framework. The PURE system environment is arguably more complex than an optimized buffer containing only the proteins necessary for DNA replication. It is however also evidently less complex than the crowded bacterial cytoplasm of a *Bacillus subtilis*. To value if the synthesized replication proteins perform the DNA replication reaction as 'expected', requires a comparison with a reference DNA replication reaction by the purified proteins in the same environment.
Furst, it was confirmed that the purified proteins were active in our hands. In Figure 6.15 the amplification of the phi29-genome by the purified p2, p3, p5 and p6 proteins, in the same conditions as reported in [15], is analyzed. The following observations were made:
• The replication reaction was more efficient at 22 °C than at 30 °C.
• On the neutral agarose gel the lower band of DNA and the bands close to and inside the well disappeared upon S1 nuclease treatment prior loading on the gel.
This leads us to conclude that
• A large amount of ssDNA, visible as the lower band in the neutral gel, is generated. It probably originates from the strand displacement reaction at only one end of the genome.
• A fraction of ssDNA aggregates in the well.
• The alkaline gel does not distinguish between the amplified ssDNA or dsDNA products, whereas this distinction is visible on a neutral agarose gel.

Next, we investigated the amplification of the phi29 genome by the purified proteins

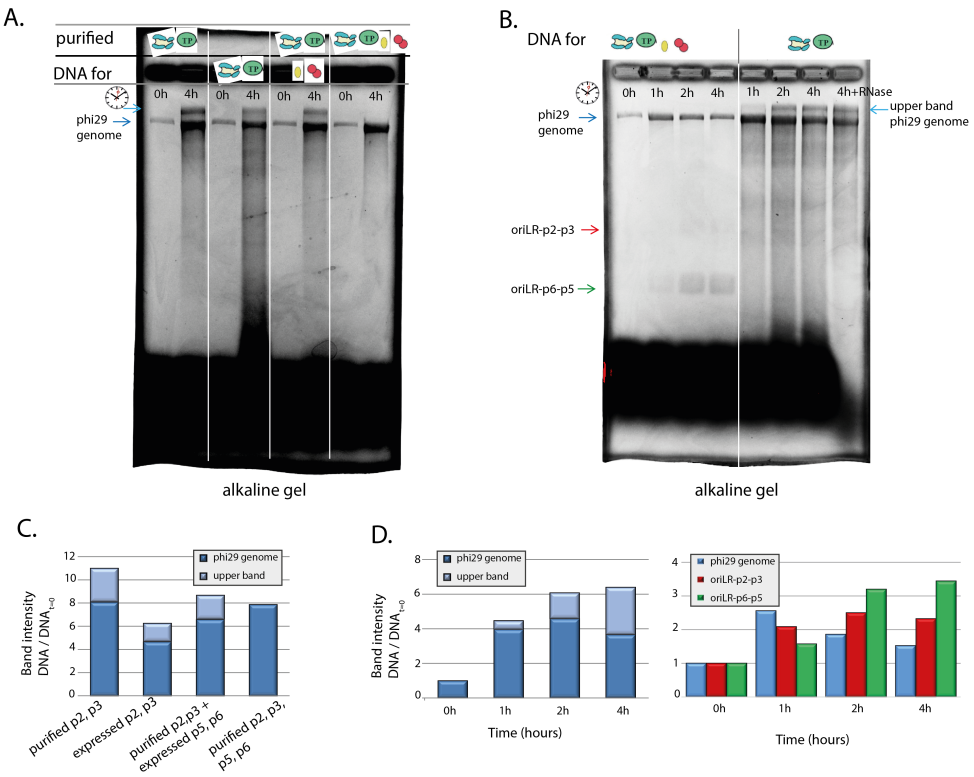


Figure 6.16: Replication of the phi29-genome in the PURE system by the purified and/or expressed phi29 DNA replication proteins visualized on an alkaline agarose gels. A) Amplification of the phi29-genome is performed by purified or expressed phi29 DNA replication proteins, or by a combination of the two and quantified in (C). B) Amplification of the phi29-genome by the expressed replication proteins, quantified in (D).

in the PURE system. For this purpose the PURE system was supplemented with ca. 90 ng of the phi29 genome, 40 ng of the purified p2 and p3, 2 μg p6, and 7.5 μg p5 protein, and analyzed on an alkaline gel the amount of replication product of after 4 h reaction (Figure 6.16A). Similarly, amplification of the phi29 genome by only the DNAP (p2) and TP (p3), purified or expressed, and amplification by the purified p2 and p3 combined with the expressed p5 and p6 proteins were analyzed. The amplification yields in these experiments vary between 5 and 11-fold after 4 h, as displayed in Figure 6.16C. These positive results confirm that a DNA replication system based on the phi29 proteins is compatible with the PURE system environment and making it an attractive candidate as a DNA replication mechanism in the minimal cell.

Interestingly, the highest amplification yield in these experiments is obtained with only the purified p2 and p3 proteins. This suggest that under these conditions, with a high starting amount of phi29 genome, the p5 and p6 proteins do not increase the yield. However, their presence reduces the amount of smear observed in the gels and thus prevents the formation of side products. Note that this smear is not present at time-point zero and we can therefore safely assume this is not caused by any of the rRNA or proteins from the PURE system. A larger amount of smear was detected when the DNAP and TP are expressed from their templates than in the experiment with purified p2 and p3 proteins. In the earlier situation, the amount of p2 and p3 proteins is not limited to the input concentration (p2: 30 nM and p3: 65 nM) and will be higher due to continuous protein expression from the DNA template over the time course of the experiment. As stated in literature, a small increase in the amount of TP-DNA polymerase complex with respect to the one allowing optimal amplification of full-length phi29 DNA leads to the generation of immature elongation products [7]. The higher protein concentrations might therefore be responsible for the large generation of side products. The maximum observed amplification of 11-fold is equivalent to ca. 4 nM of phi29-genome in the final reaction sample. This implies that even in the case of purified p2 and p3 protein, still an excess amount of free p3 protein (ca. 57 nM) and p2 is present in the sample.

Now that it is confirmed that the expressed p2 and p3 replication proteins can amplify the phi29-genome, we can study the effect of the co-expression of the p5 and p6 proteins. In Figure 6.16B, the amplification products of the phi29 genome in the PURE system over time by either the complete set of the four expressed phi29 replication proteins or by the DNAP and TP only was analyzed. The phi29 genome was amplified 4-6 fold by the DNAP and TP alone, and 1.5-2 fold by the DNAP, TP, SSB and DSB proteins. Also here it was observed clearly that more side products were formed in the absence of the p5 and p6 protein. Because we wondered if the formation of these shorter replication products could be reduced upon lowering the total amount of expressed DNAPs and TPs in the reaction, a sample of the reaction was treated with RNase after 40 min to stop the translation of proteins. After 4 h, less shorter products were indeed observed, though the reduction is not as convincing as in the case when the p5 and p6 proteins were expressed[16].

[16] At this moment we cannot completely rule out that some smear is caused by uncleaved mRNA. Note that the RNase treatment also successfully removes the lower blob of RNA present in the alkaline gel.

The amplification yield of the phi29-genome is lower in the experiment when the p5 and p6 proteins are co-expressed with the p2 ,p3 proteins, as quantified in Figure 6.16D. This might result from a lower expression of the p2 and p3 proteins, and their distribution to the other oriLR-p2-p3 and oriLR-p5-p6 templates. A 2-3 fold band intensity increase for these templates was observed, indicating that multiple templates containing origins of replication can be amplified in a single reaction.

6.6. DISCUSSION

In this chapter, we have investigated for the first time the phi29 DNA replication mechanism in a minimal gene expression system and provide the *proof-of-principle* that the ensemble of the synthesized p2, p3, p5 and p6 proteins can replicate a dsDNA template with a length of about 20 kb, namely the phi29 genome. We have observed that the expressed p2 and p3 proteins alone are capable of amplifying the phi29 genome and of duplicating a heterologous DNA template containing the same origins as the phi29 genome but lacking the terminal protein. The amplification of the phi29 genome (more than 4-fold starting from 90 ng) of this minimal system is higher than reported for *in vitro* experiments described in literature, where an input of 50 ng of DNA was amplified only two-fold [7]. However, no significant stimulation of the DNA amplification was observed in our system with addition of purified or expressed p5 or p6 proteins. We do not have a robust explanation for this. On a positive note, the presence of the p5 and p6 proteins reduced the formation of side products and therefore the quality of the total amount of newly synthesized DNA.

Despite the fact that the amplification of DNA templates by the expressed proteins is, at the moment, lower than we had naively expected, this is a good starting point to optimize DNA replication in the minimal cell framework. What amplification rate would be satisfying? The answer of this question depends on the context. Duplication of the genome *in vivo* is necessary before division into daughter cells take place. A duplication rate of 20 min, as achieved by fast growing bacteria, is in our case not realistic: the measured elongation rate of the phi29 DNAP is between 10-60 nt/s, which is between 1 and 2 orders of magnitude slower than the rate of fork movement in *E. coli* [17, 36, 37]. The time required to duplicate a 20-kb genome by phi29 DNAP is between 8 and 35 min (taking into account the initation time of 2 min) [17]. However, since nothing close to a minimal cell exists at the moment, this context imposes yet no strict requirements and leaves room for imagination. Our aim is to understand how DNA replication by the expressed proteins in the PURE system reaction can be optimized in terms of yield and quality of DNA products. Once an optimum is reached, it allows us to proceed to the next step in the minimal cell project that is to combine different modules, like DNA replication and lipid biosynthesis in the same environment and study how they affect each other. In this section, we will further discuss the ongoing and future research directions into the DNA replication by the phi29 proteins.

OPTIMIZATION OF TEMPERATURE

• What is the optimal temperature for DNA replication by the phi 29 proteins?
We have learned from DNA replication experiments with purified proteins that the yield is higher at 22°C than at 30°C. Though, the question of choosing an optimal temperature goes hand in hand with the same question for other necessary processes that are performed in the artificial cell-like environment. When proteins are expressed in the PURE system, a lower temperature will slow down the processes of the translation reaction, thus it takes more time to reach the same production of proteins. A lower expression temperature might also lead to a positive effect, since some proteins fold better at lower temperatures without the need of chaperones (that are not present in the PURE system) [23]. Replication of the phi29 genome by the four expressed proteins also works at 22°C (Figure 6.17A).
In this experiment it is evident that no fluorescent nucleotides are incorporated into the final product at 20 min, which indicates that DNA replication of the full genome takes more than 20 min when the replication proteins are expressed from their DNA templates in the PURE system.

We also observed that the broadening of the oriLR-p2-p3 and oriLR-p5-p6 templates in the alkaline gel does only happen inside the PURE system and not when the templates are dissolved in water. We therefore suppose that the broadening of the bands is not due to Joule heating but a buffer dependent process and intrinsic to the protocol of analysis for PURE system samples. We note that the integrated intensity of the separate bands does not significantly change upon the broadening, see the band quantization in Figure 6.17B.

AMPLIFICATION YIELD DEPENDENT ON DNA CONCENTRATIONS

• What is the minimal amount of DNA template necessary to initiate the amplification reaction?
In the *in vitro* amplification systems based on the purified phi29 proteins, the minimal amount of input template was determined to be 10 ng for both the phi29 genome and the heterologous DNA template containing the phi29 origins [15]. In earlier studies, amplification of 0.5 ng of input phi29 genome was achieved when high amounts of p6 protein were present [11]. In Figure 6.13 we observe incorporation of fluorescent nucleotides into newly synthesized DNA when starting with an input of 10 ng of oriL68-p2-p3 template. In some of our experiments, we have observed that in a reaction with two DNA templates containing the phi29 origins, the DNA template that is present in the highest absolute amount is the DNA that is the most amplified. Interestingly, this was also the case when the oriLR-p2-p3 template and the phi29 genome were present in absolute amounts of 108 ng (2.7 nM) and 48 ng (0.2 nM), respectively. In Figure 6.18 fluorescent nucleotides incorporated in newly synthesized oriLR-p2-p3 DNA was observed, not in the full-length phi 29 genome. This is in contrast to the situation where only the phi29 genome (48 ng) is amplified when low amounts (14 ng) of oriLR-p2-p3 are present in the sample. This observation is surprising because the *in vitro* amplification factor of the phi29 genome is found to be around three times higher than that of the heterologous DNA template containing the phi29 origins as described [15]. The TP that is covalently attached to the 5'-end of the phi29 genome, decreases the amount of template molecules

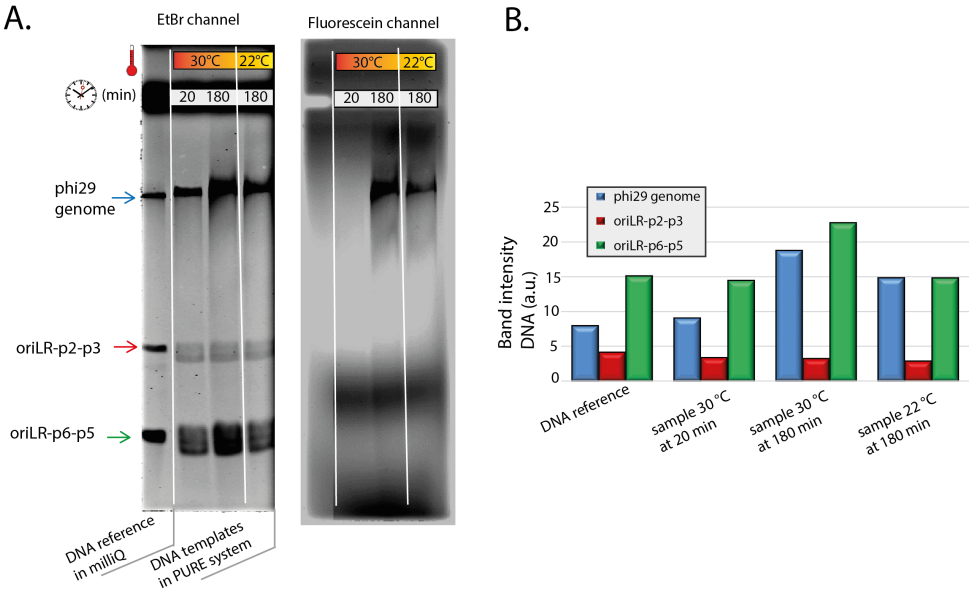


Figure 6.17: Replication of the phi29-genome in the PURE system by its expressed proteins at 22°C and 30°.

needed for amplification. It was therefore not expected that the phi29 genome was not amplified at all.

In Figure 6.19A it is confirmed that low amounts of oriLR-p2-p3 templates (10 ng) are not amplified in the PURE system when other templates containing the oriLR sequence are present in excess. When the oriLR-p6 (50 ng) template, or the oriLR-p6-p5 (83 ng) are simultaneously expressed together with low amounts of oriLR-p2-p3 template, fluorescent nucleotide incorporation is not visible for the latter DNA. To our surprise an excess of the p6 template (27 ng) without origins was used as a template for initiating the DNA replication reaction. The 5'-end sequences of this template are AAAAAA (primer 91 ChD in Table 6.2) and GCGAAA (primer 73 ChD), therefore initiation at one end of the template can occur as it fulfills the requirement of a triplet sequence. Because only one end of the p6 template contains phi29 origin signal, the observed amplification on the ethidium bromide gel remains unexplained.

Another observation we made is that some DNA templates are amplified more efficiently than others, for example the oriLR-p6 versus the oriLR-p2-p3 (data not shown). This could be tested by supplementing the PURE system lacking the T7 RNAP with the template of interest and the purified p2, p3, p5 and p6 proteins followed by DNA gel band intensity analysis. We hypothesize that amplification of templates can be length dependent (simply because shorter templates are synthesized faster, and maybe because the DNAP-TP complex once bound to the DNA is closer to an origin). We therefore propose to find the right initial template conditions in each specific case necessary to achieve a

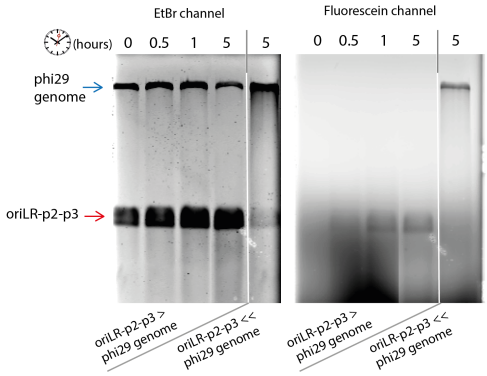


Figure 6.18: Amplification of DNA molecules by the expressed p2, p3 proteins is dependent on the amount of input templates. In the experiment where the oriLR-p2-p3 is in excess of the phi29 genome, the oriLR-p2-p3 is used as template for replication and amplified 2-fold, whereas it is not amplified when the phi29 genome is present in excess.

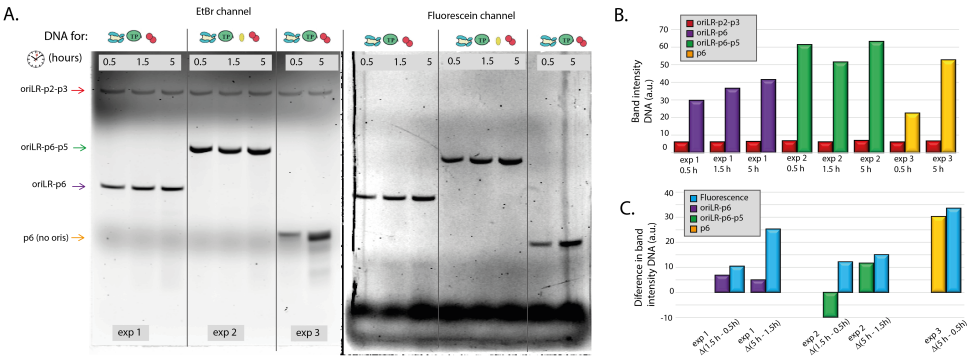


Figure 6.19: Amplification of templates depends on their amount of input DNA. A) DNA replication experiments in the presence of fluorescent nucleotides were visualized on a neutral agarose gel in the EtBr (left) and fluorescein channels (right). The oriLR-p2-p3 template is not amplified, but either the oriLR-p6, the oriLR-p6-p5, or the p6 no oris template, that were present in higher amounts, were utilized as template for nucleotide incorporation. B) Quantification of band intensities in the EtBr channel. C) Differences in EtBr band intensities are not always directly correlated with increases in the fluorescein channel.

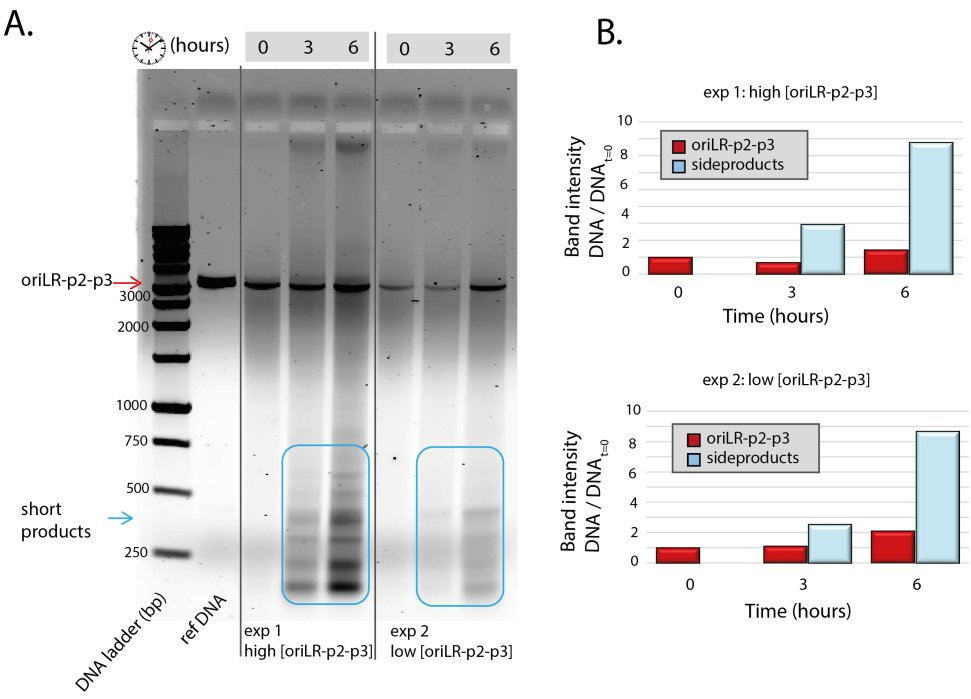


Figure 6.20: Short products are created during DNA replication. A) DNA replication reaction samples visualized on neutral agarose gel show the formation of small products over time. B) Quantification of side products by band intensity analysis.

desired duplication rate.

INEFFICIENT REPLICATION DUE TO THE FORMATION OF SHORT PRODUCTS

In Figure 6.14 the generation of side products during replication of the oriLR-p2-p3 template by its own encoded DNAP and TP was visible in the lower part of the neutral gel. These products are likely replicons and thus compete with the full-length template for the DNA replication resources. These shorter replicons indeed accumulate over the time course of the experiment (Figure 6.20). The amount of DNA generated as small products below 750 bp is quantified in Figure 6.20B, relative to the input DNA of the oriLR-p2-p3 template in two different experiments. In this experiment 4.6 nM (90 ng) or 1.9 nM (37 ng) of the oriLR-p2-p3 template was expressed together with 1.9 nM (5 ng) of p6 template. A similar 8-fold amount of side-products is generated starting with either a high or low amount of DNA template. This means that the formation of (absolute amount of) short products is more abundant when starting with higher DNA concentrations. For a rough estimate, we assumed that a side product has an average size of 300 bp and its nature is double-stranded. After 6 h, around 80-fold more short molecules than input DNA is present in the reaction sample. This means that in the experiment with the high oriLR-p2-p3 concentration, around 0.4 μM templates are present that contain origins of

replication, with 0.8 μM of TP covalently attached to their ends [17]. It is thus of utmost importance to prevent the formation of these side-products to prevent depletion of the TP for the replication of the template of interest.
The main question we have to address is:
• Is the presence of the p5 protein alone sufficient to prevent the formation of side products?
In the same line of thought, we would like to continue our investigations to answer the following questions:
• Can we omit the p6 protein in our system since we have observed reasonable amplification with only the TP and DNAP proteins expressed?
• What is an optimal stoichiometry of proteins to reach a desired amplification rate and good quality DNA?
• How can we tune the expression of the proteins when all genes are integrated into one DNA template?
For the latter two questions, we would need to have a better understanding of 1) the amount of proteins that are expressed in the PURE system at given conditions, 2) how co-expression alters the amount of the individual expressed proteins and 3) how the competition of different levels of mRNA transcription (as a consequence of promoter strength or gene position) and mRNA translation signals (RBS strengths) resonates in final expression yields.

OPTIMIZATION OF *in vitro* TRANSCRIPTION AND TRANSLATION ENVIRONMENT FOR DNA REPLICATION

The *in vitro* transcription translation system in which the DNA replication is performed poses a boundary to the optimization of the process, as some reactions or reaction components can hinder DNA replication but are necessary for gene expression.
• Can the PURE system be optimized, without compromising protein expression yields, for DNA replication?
One of the ideas here is to lower the amount of NTPs in the solution as they compete with the dNTPs for the active site of the DNA polymerase, slowing down the elongation process. This idea will be discussed further in the next chapter.
We believe that utilizing liposomes with semi-permeable membranes as micro-reactors for the combined transcription, translation and DNA replication reactions can aid in optimizing the concerted functioning of these processes. An idea would be to have membrane pores that serve as transport channels for NTPs and dNTPs from the outside solution to reach the inside of the liposome. We hope to learn:
• What is the effect of confining the reaction inside a cell-sized liposome?
Related to the PURE system context, we also wonder:
• What happens when a progressing phi29 DNAP encounters and collides with a T7 RNAP or another roadblock on the DNA template?
This has already been beautifully adressed for the *in vivo* situation in an *in vitro* assay studying head-on and co-directional collisions of the multi-subunit *Bacillus subtilis*

[17] This might explain the observation that when the final time-point sample of a DNA replication experiment in the PURE system is loaded on a protein gel, in the best case only very faint bands of the TP protein are visible compared to control experiments where no dNTPs are added (data not shown).

RNAP and the phi29 DNAP [37, 38]. In the next chapter we describe the *in vitro* assay we designed to address this question for single-subunit viral DNAP and RNAPs.

IMPROVING EXPERIMENTAL PROCEDURES

More experiments will have to be conducted to confirm the replication efficiency on the different templates. In the Materials and Methods section we explain concisely what the current protocol is and what problems we encountered before establishing it. Note that replication efficiencies are always determined by the band intensity increase on the ethidium bromide stained DNA gels. We experienced that the relation between the increase in band intensity measured in the fluorescein and ethidium bromide channels on neutral agarose gels does not follow a consistent trend, see Figure 6.19C. A possible explanation is that reactions where only one strand is replicated and displaces the native strand, only an increase in fluorescein signal at the dsDNA band is expected, as no netto dsDNA is produced.

As a final experiment to demonstrate that the increase in band intensity on the gel represents an accurate copy of a dsDNA template, we plan to provide an answer to the question:

• Can we prove experimentally that the replicated DNA is used for further transcription and translation?

We will first define a combination of templates that can faithfully replicate full-length oriLR-DNA without the formation of side products. Then, in order to isolate the replicated DNA molecules from the initial DNA molecules present in the reaction sample, lambda-exonuclease treatment will be performed as in Figure 6.14. The purified sample free from RNA and proteins is subsequently used as input for a new PURE system reaction (or for a sequencing reaction). Thereafter we can either analyze RNA transcription or protein production from these templates on gel, or purifiy and analyze the DNA again to observe another round of replication. A control sample will not be supplemented with dNTPs, to confirm the complete degradation of the input template by the lambda-exonclease.

6.6.1. FUTURE WORK

Several ideas for future research were discussed and we will touch upon a few of them in Chapter 7. Lastly, the results presented here, that show that DNA replication proteins can copy a template encoding their own genes, opens the door for evolution experiments (using an error-prone DNAP mutant).

6.7. CONCLUDING REMARKS

In summary, we propose the DNA replication mechanism by the bacteriophage phi29 proteins, DNAP (p2), TP (p3), SSB (p5) and DSB (p6), as a potential candidate for the DNA replication module in the minimal cell. We have designed templates encoding for two replication proteins flanked by ca. 200 bp of the right and left end of the phi29 genome containing the origins of replication (oriLR-p2-p3 and oriLR-p5-p6). The replication activity of the PURE system synthesized phi29 DNAP and TP is confirmed by imaging the

samples after a purification protocol on a neutral gel and by analyzing the band intensities of the DNA of interest at different time-points. The two synthesized proteins are also capable of amplifying the ca. 20-kb phi29 genome. This minimal replication system generates large amounts of side products in absence of the accessory proteins (p5 and p6). Co-expression of the four replication proteins reduces the side product formation in the amplification reaction of the phi29-genome. The groundwork is laid out for further investigations and suggestions to optimize conditions for DNA replication in the PURE system are discussed.

6.8. MATERIALS AND METHODS

DNA CONSTRUCT PREPARATION

The origin of the phi29 gene sequences is described in the main text. Sequences of interest are ordered at the GenScript® company and cloned by them in the pUC57 vector with EcoRV cloning sites. The full sequences for oriLRp6 and oriLR-p6-p5 are directly ordered at the company, whereas the TP and DNAP genes were ordered separately at the start of the project. The TP gene and the DNAP gene with vsv-terminator were assembled into a fusion DNAP-TP construct by assembly PCR. In Table 6.2 the list and sequences of PCR and sequencing primers used in this project are given and their descriptions is reported in Table 6.3. Other specific primers are mentioned in their respective Materials and Methods section. Regular PCR reactions were performed with 1-10 ng of plasmid or linear DNA as template, 1 unit of Phusion polymerase (Finnzymes) in HF buffer containing 0.2 mM dNTPs, 0.2 μM forward and 0.2 μM reverse primers in a final volume of 50 μL. After an initial heating step for 30 s at 98 °C, the PCR consists of 30 cycles of 10 s melting the DNA at 98°C, followed by hybridization of the primers for 15 s at 60 °C, and elongation by the DNAP at 72 °C for 30 s per kb template. After the 30 cycles the temperature remains 5 min constant at 72 °C to allow the DNAP to complete all remaining polymerization reactions. The PCR-generated linear DNA fragments are purified with the PCR Clean-up kit from Promega according to the manufacturer's protocol. The concentration of the purified DNA was measured on the NanoDrop (Thermo Scientific) and the purity of the DNA products checked on a TAE 0.7-1.1% agarose gel using 100 ng of DNA and ethidium bromide staining. The BenchTop 1kb DNA Ladder from Promega or the 1kb Plus DNA ladder from ThermoFischer Scientific was used to confirm the correct size of the dsDNA templates.

PURE SYSTEM

The cell-free reaction solutions of PURE*frex* were purchased from GeneFrontier Corporation (Japan). The kit comes in three vials: enzyme mixture (T7 RNA polymerase, translation factors, energy recycling system, etc.), buffer (feeding) mixture (amino acids, NTPs, tRNAs, etc.), and ribosome solution. The reaction mixture for a 20-μl PURE*frex* reaction is described in Table 6.4. Unless indicated otherwise in the text, the default volume of a reaction is 20 μl. Specific conditions and other supplements are indicated at the individual reaction descriptions. Reactions are incubated in a nuclease-free PCR tube (VWR) in a ThermalCycler (C1000 Touch, Biorad) at a default temperature of 30 °C. We specifically choose to utilize this PURE system and not the PURExpress, since DNA degradation starting from the ends of the templates has been observed in the PURExpress system [39].

TRANSCRIPTION OF DNA TEMPLATES

Trancription of DNA templates was performed with the RiboMAX™ Large Scale RNA production kit (Promega), congruous with the recommended protocol, or with the PURE*frex*ΔR kit. All RNA purification was done with RNeasy MinElute Cleanup kit (Qiagen) following the manufacturer's protocol. RNA concentrations were determined using a Nanodrop (Thermo Scientific) with absorbance measurements performed at 280 nm. After the purification step the RNA samples were imaged on a 1.2% agarose gel containing EtBr and

Primer	Sequence 5'-3'
73 ChD	GCGAAATTAATACGACTCACTATAGGGAGACC
91 ChD	AAAAAACCCCTCAAGACCCGTTTAGAGG
173 ChD	CACACAGGAAACAGCTATGAC
181 ChD	CAAAAAACCCCTCAAGACCCGTTTAGAGG
363 ChD	AAAGTAAGCCCCCACCCTCACATGATAGCGAAATTAATACGACTCACTATAGGGAGACC
364 ChD	AAAGTAGGGTACAGCGACAACATACACAAAAAACCCCTCAAGACCCGTTTAGAGG
365 ChD	CAGTCACGACGTTGTAAAACGAC
422 ChD	GGTCTCCCTATAGTGAGTCGTATTAGCAGTCGACGGGCCCGGGATCCGAT
423 ChD	ATCGGATCCCGGGCCCGTCGACTGCTAATACGACTCACTATAGGGAGACC
430 ChD	CCATACAGGCTGTTCAGCATC
431 ChD	GTGTCATCGACCAGCACAAC
432 ChD	CCATGGATTCTTCCAGGGTG
433 ChD	CACCCTGGAAGAATCCATGG
434 ChD	AAAGTAAGCCCCCACCCTCACATGATAG
435 ChD	AAAGTAGGGTACAGCGACAACATACAC
491 ChD	AAAGTAAGCCCCCACCCTCACATG
492 ChD	AAAGTAGGGTACAGCGACAACATACAC

Table 6.2: Primer sequences.

Primer	FW/RV	Purpose	Comments
73 ChD	FW	includes T7 promoter	7 nt before T7 promoter
91 ChD	RV	includes T7 terminator	5'-AAA is potential ori
173 ChD	RV	isolate DNAP gene from pUC57	
181 ChD	RV	includes T7 terminator	5'-CAA, no possible origin
363 ChD	FW	extend T7 promoter with minimal ori	oriL24, 5'-phosphorylated
364 ChD	RV	extend T7 terminator with minimal ori	oriR27, 5'-phosphorylated
365 ChD	FW	isolate DNAP gene from pUC57	
422 ChD	FW	assembly PCR DNAP-TP	
423 ChD	RV	assembly PCR DNAP-TP	
430 ChD	RV	for sequencing of DNAP-TP	
431 ChD	RV	for sequencing of DNAP-TP	
432 ChD	RV	for sequencing of DNAP-TP	
433 ChD	FW	for sequencing of DNAP-TP	
434 ChD	FW	assembly PCR	5'-phosphorylated
435 ChD	RV	assembly PCR	5'-phosphorylated
491 ChD	FW	oriL24, for all oriL68 and oriL191	5'-phosphorylated
492 ChD	RV	oriR25, for all oriR194	5'-phosphorylated

Table 6.3: Primer list.

Component	final concentration	Volume
Buffer solution (B)		10 μl
Enzyme solution (E)		1 μl
Ribosomes (R)	1.2 μM	1 μl
DNA constructs	10-150 ng	x μl
Ammonium Sulfate	20 mM	2 μl
RNase free H_2O (milliQ)		x μl
dNTP mix	0.3 mM	0.3 μl
Total		20 μl

Table 6.4: Reaction mixture for a 20uL PUREfrex reaction

the band intensities were analyzed with ImageLab.

PROTEIN EXPRESSION VISUALIZATION

To visualize protein expression the standard PURE*frex* reaction mixture was supplemented with 0.5 μL BODIPY-Lys-tRNA$_{Lys}$ (FluoroTect^{TM} GreenLys, Promega) that labels translation products. Around 3 nM of the DNA templates of TP, DNAP, p5 and p6 are separately expressed in a reaction volume of 20 μL for 3.5 h at 30 °C. Co-expression of the four proteins is performed starting with 0.7 nM of both oriLR-p2-p3 and oriLR-p5-p6 templates. The oriLR-p2-p3 and oriLR-p5-p6 are also expressed with a ratio of 0.35 nM : 4.7 nM that is more realistic to the conditions of co-expression in DNA replication reactions. Co-expression of the DNAP, TP and p6 proteins was initiated from 4.6 nM oriLR-p2-p3 and 3.8 nM p6 template. Samples were treated with RNase A (0.1 mg/ml final concentration) for 30 min at 30 °C to degrade the remaining labeled tRNA. Then, 5 μL of the reaction is mixed with 2x Laemmli Sample buffer and 10 mM DTT final, denatured for 2.5 min at 65 °C and analyzed on a 12 or 15 % SDS Page gel. Fluorescence detection of the labeled protein was performed on a fluorescence gel imager (Typhoon, Amersham Biosciences) and visualization of the PURE proteins and purified proteins is analyzed by Coomassie Brilliant Blue (Promega) after-staining on a GelImager. The purified proteins

ACTIVITY ASSAY WITH SS-YFP-SPINACH-DNA

The ssDNA of the YFP-Spinach template is produced in two steps. First, the promoter sequence is removed from the initial mYFPco-LL-Spinach-T7t as described in Chapter 3, by a PCR reaction with FW primer 185 ChD: 5'-GAGACCACAACGGTTTCCCTCTAG-3' and RV primer 190 ChD:
5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATGCTAGCCCGGGGATATC-3' (replaces the T7 terminator with a poly-A tail). This purified DNA is then used as input template for an asymmetric PCR, where the FW primer is diluted by a factor of 50. The result is a mixture of ssDNA and dsDNA, both without the T7 promoter sequence, thus ΔT7p-mYFPco-LL-Spinach-polyA. The standard PURE*frex* reaction is supplemented with 0.2 mM dNTP mix (Promega), 0.5 μL of Superase inhibitor (10 units final; SUPERase.In, Ambion) 20 μM DFHBI (final concentration), 200 ng ss/dsDNA of ΔT7p-mYFPco-LL-Spinach-polyA, 1 μM of Primer 361 ChD:
5'-GCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAG-3' (the oligonu-

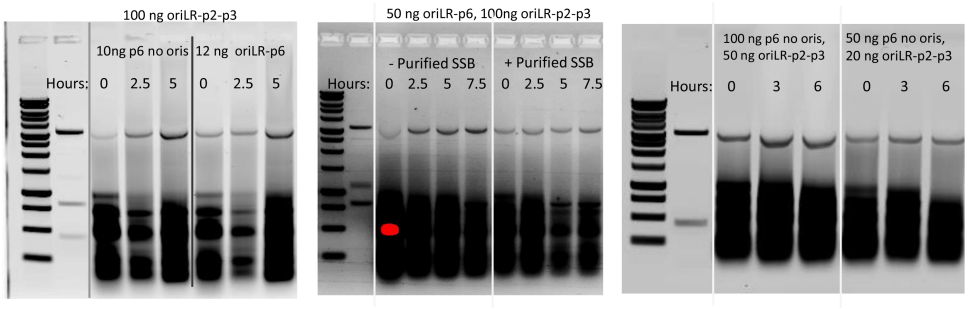


Figure 6.21: Examples of inconclusive experiments that were visualized on agarose gels stained with EtBr and contained large smears. The smear that is attributed to incomplete RNA digestion renders the quantification of these gels unreliable, and shorter DNA templates are not visible. Due to the loss of DNA in the first timepoint in the most left gel, it is hard to conclude whether the increase of band intensity actually comes from DNA replication or is due to the bias introduced by the protocol.

cleotide that contains the promoter sequences and an overhang that hybridizes to the ssDNA) and either 90 ng of DNAP-DNA, or 5 units of purified phi29 DNAP (New England Biolabs), or no DNAP enzyme in the control reaction. The reaction mixtures are transferred in a 15-μL cuvette (Hellma) that was mounted in the temperature-controlled holder of a fluorescence spectrophotometer (Cary Eclipse from Varian) held at 37 °C and the fluorescence was recorded every 30 s using the following excitation/emission wavelengths: Spinach, 460/502 nm; YFP, 515/528 nm.

TREATMENT OF PURE SYSTEM SAMPLES PRIOR LOADING TO GELS

The protocol to analyze PURE reaction samples on neutral agarose gels involves a protein and RNA removal step. Without complete RNA removal, gel analysis becomes less reliable (Figure 6.21) as we have encountered the following problems that interfere with DNA band intensity analysis:
- Variable loss of DNA during column purification, probably due to overloading or aggregation of large RNA molecules
- Smear on gels that restrains visibility of short DNA templates
- DNA that does not enter the gel well due to overloading of the well or incomplete removal of proteins (Figure 6.13)
- Fluorescent nucleotide incorporation in RNA molecules

The optimized protocol we used in Figures 6.14 and Figure 6.19 is, to add to 5 μl reaction sample:
- 0.5 μl RNase A (0.4mg/ml final) plus 0.5 μL RNase ONE^{TM} Ribonuclease (2.5-5 units), incubation for 30 min at 30 °C.
- 3 μl of STOP solution (final concentrations: 10 mM EDTA, 0.1% SDS) plus addition of 0.5 μl of 0.1 mg/ml Proteinase K solution, for 1 h at room temperature
- samples are stored at 4 °C or -20 °C until further use
- all time point samples are treated simultaneously with the RNeasy MinElute Cleanup kit (Qiagen) following the manufacturer's protocol finalizing with an elution step in 14 μL milliQ

The samples were loaded with a 6X DNA loading buffer (Promega) on a 1.1% agarose gel containing ethidium bromide and run in TAE buffer for 90 minutes at 90 V.
The sample treatment for analysis on alkaline agarose gels consists only of quenching the reaction with the STOP solution and Proteinase K. Other additions of enzymes need to be performed before this step. The lambda exonuclease treatment in Figure 6.14 consists of addition of 0.3 μl Lambda Exonuclease (1.5 units, New England Biolabs) for 30 min at 30 °C.
All PURE system experiments containing the phi29 genome were analyzed on alkaline gels because the RNA column purification does not recover the long genome. Following the same protocol omitting the column purification step prevents the DNA running into the gel. An alternative to the column purification can be to use buffer exchange columns with a cut-off of 100 kDa or lower, and after centrifugation retrieve the supernatant that still contains the DNA templates but not the small peptides, oligonucleotides and monomers.

ANALYSIS ON ALKALINE AGAROSE GELS

Alkaline conditions separate DNA strands, allowing to separate unfolded ssDNA according to their respective sizes. The guanine and thymine bases lose a hydrogen and become negatively charged, preventing association in the normal DNA base-pairing [40]. RNA is hydrolized in alkaline gels. The alkaline agarose gels are prepared according to the protocol described in [41] that is largely based on the protocol developed by the group of Prof. Studier to analyse restriction fragments of the T7 DNA [40]. In short, 0.7 g agarose is dissolved in 100 mL (for a 0.7% gel) in a sodium-chloride buffer (30 mM NaCl, 2 mM EDTA, pH 7.5) by boiling it in the microwave. The solution is poured into the gel tray and let solidified for 30 min. The gel is then immersed into the alkaline electrophoresis buffer (30 mM NaOH, 2 mM EDTA) for at least 1 h. A 6X alkaline loading buffer (180 mM NaOH, 6 mM EDTA, 18% Ficoll 400) was added to the samples that were heated for 5 min at 70 °C, then chilled on ice for 3 min prior loading onto the gel. For visibility of the samples during the loading process, a small fraction (ca. 0.1%) of bromophenol blue is added to the alkaline loading buffer, although bromcresol green is recommended (as this dye does not lose color in alkali environments). It is important to complex the free Mg^{2+} by EDTA in the sample, otherwise the DNA can precipitate in the alkaline gel conditions. Gels are run at 35 V - 45 V for 4-5 h in the alkaline buffer. Afterwards the gel is immersed for 30 min in 200 ml of 0.5 M Tris-HCl buffer, pH 7.5, and stained with ethidium bromide.

ROLLING CIRCLE AMPLIFICATION ASSAYS

The M13mp18ss (New England Biolabs) circular ssDNA plasmid was used as template for the RCA reactions in combination with the hybridization primer 5'-GTTTTCCCAGTC ACGAC-3' (with 5'-Cy5 label: 332 ChD, or without label: 331 ChD). To pre-hybridize the primer to the template, The M13mp18ss and primer were mixed in nuclease-free water in molar ratio 1:5, preheated for 5 min at 65 °C and cooled down at room temperature for 5 min. The final 25-μl solution contains 1X phi29 DNA Polymerase Reaction Buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 4 mM DTT, pH 7.5 at 25 °C), 20 units phi29 DNAP, 0.3 mM dNTP mix, 0.02 μM template and 0.1 μM primer. When the activity of expressed DNAP was tested, the same protocol applies with omission of the phi29 buffer and instead of the purified phi29 DNAP enzyme, 45 ng of the DNAP DNA

template is added to the PURE*frex* reaction. 5 μl samples are taken out at the indicated time-points and quenched by 2.5 μl of the STOP solution (final concentrations: 10 mM EDTA, 0.1% SDS) and by addition of 0.5 μl of 0.1 mg/ml Proteinase K solution (Promega). The samples were loaded on a 0.7% alkaline agarose gel and after-stained with ethidium bromide.

AMPLIFICATION OF THE PHI29 GENOME WITH PURIFIED OR EXPRESSED PROTEINS

The purified proteins and phi-29 genome that we received from our collaborators Margarita Salas and Mario Mencio have the following stock concentrations and storage buffers:
- p2, 320 ng/μl in 50 mM Tris pH 7.5, 0.5M NaCl, 1 mM EDTA, 7 mM 2-Mercaptoethanol (BME), 50% glycerol.
- p3, 400ng/μl in 25 mM Tris pH 7.5, 0.5M NaCl, 1 mM EDTA, 7 mM BME, 0.025% Tween20, 50% glycerol.
- p5, 10 mg/ml in 50 mM Tris pH 7.5, 60 mM ammonium sulfate, 1 mM EDTA, 7 mM (BME), 50% glycerol.
- p6, 10 mg/ml in 50 mM Tris pH 7.5, 0.1 M ammonium sulfate, 1 mM EDTA, 7 mM BME, 50% glycerol.
- phi29 genome 190 ng/μL in 50 mM Tris pH 7.5, 0.2M NaCl, 1 mM EDTA, 7 mM BME, 0.05% Tween20, 50% glycerol.

The proteins were aliquoted and stored at -80 °C. To dilute the p2, and p3 proteins into working concentrations they are diluted in the dilution buffer (25 mM Tris pH 7.5, 0.1 M NaCl, 0.05% Tween-20) prior use. The amplification assay that describes amplification of heterologous DNA with the phi29 proteins in [15] has the following incubation mixture: in 25 μl, buffer of [50 mM Tris HCl (pH 7.5), 10 mM $MgCl_2$, 5% glycerol, 1 mM DTT, 0.1 mg/ml BSA] plus ammonium sulfate to 20 mM final concentration, 100 μM each of the four dNTPs (plus [α-^{32}P]dATP (1 μCi)), 1.5 μg p6, 7.5 μg p5, 10 ng p2 and 20 ng p3. For the experiments in Figure 6.15 we made an initial solution of 200 μl with equivalent concentrations for a 25 μl reaction of [50 mM Tris HCl (pH 7.5), 10 mM $MgCl_2$, 5% glycerol], 1 mM DT, 20 mM ammonium sulfate, 12.5 ng p2, 25 ng p3, 7.5μg p5, 1.5 μg p6 and 25 ng phi29-genome. At the different time-points, 16-μL samples were stopped by addition of 0.1% SDS and 10 mM EDTA final concentrations, and with 0.5 μl of 0.1 mg/ml Proteinase K. The samples were split equally prior to loading on the neutral and alkaline gels. S1 nuclease treatment was performed on 8 μL samples in 1X S1 nuclease buffer and 20-100 units of S1 nuclease (Promega) for 30 min at 37 °C. In the experiments where the phi29 genome was amplified inside the PURE system, as in Figure 6.16, the amounts of purified protein that were supplemented to the standard PURE*frex* reaction were: 40 ng of p2, 40 ng of p3, 2 μg p6, 7.5 μg p5 protein. The starting amount of phi29 genome is 95 ng in these reactions. The DNAP and TP are expressed from 14 ng oriLR-p2-p3 template and the p5 and p6 proteins from 25 ng oriLR-p5-p6 template. In the experiment where the phi29 genome is amplified in the PURE system by only the expressed proteins (Figure **??**B), 9 ng of oriLR-p2-p3 in combination with 80 ng oriLR-p5-p6 template is used (molar ratio of 1:20). The RNase treatment is performed by addition of 0.3 μL of RNaseONE to 6 μL of sample after 40 min. The experiment in Figure 6.17 the input DNA templates are 50 ng of phi29 genome, 14 ng oriLR-p2-p3 and 80 ng oriLR-p5-p6 (per 20 μL). The reference sample contains the equivalent amount of templates in milliQ.

6.9. ACKNOWLEDGMENTS

I would like to thank Christophe for supporting me in starting this project and for his help in setting up the fruitful collaboration we now have with Prof. Margarita Salas and dr. Mario Mencia from Universidad Autonoma De Madrid (Centro de Biologia Molecular 'Severo Ochoa'). I am very happy and grateful to Margarita Salas and all her current and previous group members, as their work and papers were the inspiration for starting this research. Thanks to Margarita and Mario in particular, for providing us with the purified phi29 DNA replication proteins and phi29 genome, and for stimulating discussions and answers to questions. Thanks to Ilja, for contributing to the improvement of the protocol, performing control experiments and for interesting discussions.

REFERENCES

[1] H. E. Hemphill and H. R. Whiteley, *Bacteriophages of Bacillus subtilis.* Bacteriological reviews **39**, 257 (1975).

[2] D. Anderson, D. Hickham, and B. Reilly, *Structure of Bacillus subtilis bacteriophage phi29 and the length of phi29 deoxyribonucleic acid.* Journal of Bacteriology **91**, 2081 (1966).

[3] C. Hulo, E. De Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios, and P. Le Mercier, *ViralZone: A knowledge resource to understand virus diversity,* Nucleic Acids Research **39** (2011), 10.1093/nar/gkq901.

[4] N. K. M. Tam, N. Q. Uyen, H. A. Hong, H. Le, T. T. Hoa, C. R. Serra, A. O. Henriques, and S. M. Cutting, *The intestinal ife cycle of Bacillus subtilis and close relatives,* Journal of Bacteriology **188**, 2692 (2006).

[5] F. Rohwer, M. Youle, H. Maughan, and N. Hisakawa, *Life in our Phage World* (Wholon, San Diego, 2015) pp. 234–236.

[6] J. Mendez, L. Blanco, A. Esteban, A. Bernad, and M. Salast, *Initiation of phi 29 DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication.* Proceedings of the National Academy of Sciences **89**, 9579 (1992).

[7] L. Blanco, J. M. Lazaro, M. de Vega, A. Bonnin, and M. Salas, *Terminal protein-primed DNA amplification.* Proceedings of the National Academy of Sciences **91**, 12198 (1994).

[8] J. Mendez, L. Blanco, J. A. Esteban, A. Bernad, and M. Salas, *Initiation of phi 29 DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication.* Proceedings of the National Academy of Sciences **89**, 9579 (1992).

[9] J. Méndez, L. Blanco, and M. Salas, *Protein-primed DNA replication: A transition between two modes of priming by a unique DNA polymerase,* EMBO Journal **16**, 2519 (1997).

[10] C. Garmendia, A. Bernad, J. A. Esteban, L. Blanco, and M. Salas, *The bacteriophage phi 29 DNA polymerase, a proofreading enzyme.* The Journal of biological chemistry **267**, 2594 (1992).

[11] L. Blanco, A. Bernads, J. M. Lazaro, G. Martin, C. Garmendia, and M. Salas, *Highly Efficient DNA Synthesis by the Phage phi 29 DNA Polymerase,* The Journal of biological chemistry **264**, 8935 (1989).

[12] J. a. Esteban, M. Salas, and L. Blanco, *Fidelity of phi29 DNA Polymerase,* The Journal of Biological Chemistry **268**, 2719 (1993).

[13] I. Holguera, D. Muñoz-Espín, and M. Salas, *Dissecting the role of the phi29 terminal protein DNA binding residues in viral DNA replication.* Nucleic acids research , 1 (2015).

[14] C. Escarmis, D. Guirao, and M. Salas, *Replication of Recombinant phi29 DNA Molecules in Bacillus subtilis Protoplasts,* Virology , 152 (1988).

[15] M. Mencía, P. Gella, A. Camacho, M. de Vega, and M. Salas, *Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage Phi29.* Proceedings of the National Academy of Sciences of the United States of America **108**, 18655 (2011).

[16] M. S. Soengas, C. Gutiérrez, and M. Salas, *Helix-destabilizing activity of phi 29 single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication.* Journal of molecular biology **253**, 517 (1995).

[17] J. Esteban, L. Blanco, L. Villar, and M. Salas, *In vitro evolution of terminal protein-containing genomes,* Proceedings of the National Academy of Sciences of the United States of America **94**, 2921 (1997).

[18] C. Ducani, G. Bernardinelli, and B. Hogberg, *Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA,* Nucleic Acids Research , 1 (2014).

[19] A. M. Abril, M. Salas, J. M. Andreu, J. M. Hermoso, and G. Rivas, *Phage phi29 Protein p6 Is in a Monomer-Dimer Equilibrium That Shifts to Higher Association States at the Millimolar Concentrations Found in Vivo,* Biochemistry **36**, 11901 (1997).

[20] M. Serrano, M. Salas, and J. M. Hermoso, *A novel nucleoprotein complex at a replication origin,* Science **248**, 1012 (1990).

[21] M. Alcorlo, M. Jiménez, A. Ortega, J. M. Hermoso, M. Salas, A. P. Minton, and G. Rivas, *Analytical Ultracentrifugation Studies of Phage ϕ{symbol}29 Protein p6 Binding to DNA,* Journal of Molecular Biology **385**, 1616 (2009).

[22] L. Blanco, J. Gutierrez, J. M. Lazaro, A. Bernad, M. Salas, C. D. B. M. Csic-uam, U. Autnoma, and C. Blanco, *Replication of phage phi29DNA in vitro: role of the viral protein p6 in initiation and elongation,* Nucleic acids research **V**, 4923 (1986).

[23] K. Fujiwara, T. Katayama, and S.-I. M. Nomura, *Cooperative working of bacterial chromosome replication proteins generated by a reconstituted protein expression system.* Nucleic acids research , 1 (2013).

[24] Y. Sakatani, N. Ichihashi, Y. Kazuta, and T. Yomo, *A transcription and translation-coupled DNA replication system using rolling-circle replication,* Scientific Reports **5**, 10404 (2015).

[25] F. Chizzolini, M. Forlin, D. Cecchi, and S. S. Mansy, *Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength.* ACS synthetic biology (2013), 10.1021/sb4000977.

[26] J. J. Dunn and F. W. Studier, *Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements.* Journal of molecular biology **166**, 477 (1983).

[27] P. O. Olins, C. S. Devine, S. H. Rangwala, and K. S. Kavka, *The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in Escherichia coli,* Gene **73**, 227 (1988).

[28] J. W. D. F. William Studier; Alan H. Rosenberg; John J. Dunn, *Use of T7 RNA Polymerase to Direct Expression of Cloned Genes,* Methods in Enzymology **33**, 3 (1990).

[29] D. L. Lyakhov, B. He, X. Zhang, F. W. Studier, J. J. Dunn, and W. T. McAllister, *Pausing and termination by bacteriophage T7 RNA polymerase.* Journal of molecular biology **280**, 201 (1998).

[30] L. Du, R. Gao, and A. C. Forster, *Engineering multigene expression in vitro and in vivo with small terminators for T7 RNA polymerase.* Biotechnology and bioengineering **104**, 1189 (2009).

[31] L. Du, S. Villarreal, and A. C. Forster, *Multigene expression in vivo: supremacy of large versus small terminators for T7 RNA polymerase.* Biotechnology and bioengineering **109**, 1043 (2012).

[32] P. van Nies, A. S. Canton, Z. Nourian, and C. Danelon, *Monitoring mRNA and protein levels in bulk and in model vesicle-based artificial cells.* Methods in enzymology **550**, 187 (2015).

[33] J. Korlach, A. Bibillo, J. Wegener, P. Peluso, T. T. Pham, I. Park, S. Clark, G. a. Otto, and S. W. Turner, *Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides.* Nucleosides, nucleotides & nucleic acids **27**, 1072 (2008).

[34] S. Kumar, A. Sood, J. Wegener, P. J. Finn, S. Nampalli, J. R. Nelson, A. Sekher, P. Mitsis, J. Macklin, and C. W. Fuller, *Terminal Phosphate Labeled Nucleotides: Synthesis, Applications, and Linker Effect on Incorporation By Dna Polymerases,* Nucleosides, Nucleotides and Nucleic Acids **24**, 401 (2005).

[35] V. Murthy, W. J. J. Meijer, L. Blanco, and M. Salas, *DNA polymerase template switching at specific sites on the phi29 genome causes the in vivo accumulation of subgenomic phi29 DNA molecules,* Molecular Microbiology **29**, 787 (1998).

[36] L. Blanco and M. Salas, *Replication of phage phi 29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length phi 29 DNA.* Proceedings of the National Academy of Sciences **82**, 6404 (1985).

[37] M. Elías-Arnanz and M. Salas, *Bacteriophage ϕ29 DNA replication arrest caused by codirectional collisions with the transcription machinery,* EMBO Journal **16**, 5775 (1997).

[38] M. Elías-Arnanz and M. Salas, *Resolution of head-on collisions between the transcription machinery and bacteriophage Φ29 DNA polymerase is dependent on RNA polymerase translocation,* EMBO Journal **18**, 5675 (1999).

[39] H. Niederholtmeyer, L. Xu, and S. J. Maerkl, *Real-Time mRNA Measurement during an in Vitro Transcription and Translation Reaction Using Binary Probes,* ACS synthetic biology **2**, 411 (2013).

[40] M. W. McDonell, M. N. Simon, and F. W. Studier, *Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels.* Journal of molecular biology **110**, 119 (1977).

[41] Thermo Fischer Scientific, *Alkaline agarose gel electrophoresis.* (2006).

7

Future Outlook and Conclusion

This chapter touches upon a few future research directions based on preliminary data related to the work presented in this thesis. First, ways to alter and influence gene expression are discussed. Thereafter, we discuss how to improve the DNA replication reaction in the minimal gene expression system. As a first proof of principle, we performed an experiment with coupled DNA replication, transcription and translation inside liposomes, where the DNA template coding for the phi29 DNAP and TP proteins is amplified. Concluding remarks will introduce the final pages of this thesis.

To achieve the biochemical processes of one or more modules of the minimal cell (Figure 1.2), the regulation of co-expression of multiple proteins is a necessity. In the next section, few examples from literature will be discussed that can help achieving more complex regulation of gene expression in the near future. We will then focus on implementations that can benefit the DNA replication process by the expressed phi29 proteins.

7.1. INFLUENCING GENE EXPRESSION

Expression levels in cells are regulated by many different mechanisms. In the minimal cell framework that utilizes cell-free gene expression systems, the regulation of gene expression can likewise be tuned by different pathways. First, regulation can occur on the transcription level: different RNA polymerase have a wide range of affinities for specific promoters and response to terminators. Moreover, genes under control of promoter sequences containing operator sites can be down regulated by DNA binding proteins. In the PURE system, the T7 RNA polymerase is naturally present, though other viral single subunit RNAPs like the T3 and SP6 RNAP are also compatible with this environment. The T7 promoter is a 23-bp highly conserved sequence from the -17 to +6 relative to the start of the RNA, and is found in its 13 variants in the T7 genome [1]. The transcription levels by the T7 RNAP in the PURE system could be tuned substantially (0.1-1.0 relative transcription compared to the conserved class III promoter) by mutating 1-4 nucleotides in the recognition sequence before the start of transcription, or by mutating one base at a time in the +1 to +6 region [2]. Whereas viral RNAPs can thus transcribe starting from different, related, promoter sequences and thereby vary transcription levels, the variety in regulation is still far less compared to that of the multi-subunit bacterial RNAPs and its different competing sigma factors that bind to distinct promoter sequences. Recently, the T7 RNAP has been engineered such that it can be divided into four fragments that have to be co-expressed to function [3]. One of these fragments is analogous to the sigma factor, and the authors built four different sigma fragments and their orthogonal cognate promoters. This can be a promising tool for the future of the minimal cell project, when the repertoire of different viral RNAPs and promoters is not sufficient to reach the required level of complexity for gene regulation. Another tool is the repression of transcription, that has been characterized in literature for both the T7 and T3 RNAPs where TetR and LacI bind to repressible promoters. This has led to the implementation of inducible feedback loops in the PURE system [4–7]. Noteworthy, the *E. coli* RNAP, when supplemented with the transcription elongation factors GreA and GreB, is capable of efficient expression from natural and synthetic sigma 70 *E. coli* promoters in the PURE system [8].

Instead of focusing on the regulation of transcription, gene expression can likewise be influenced by altering the translation process. Examples that can regulate the translation processes are riboregulators [9], varying strengths of the ribosome binding site [10–12] and the positioning of genes within an operon [2].

In the following sections we shortly discuss how transcription regulation can benefit our ongoing projects in the near future, and what the influence of temperature is on gene expression.

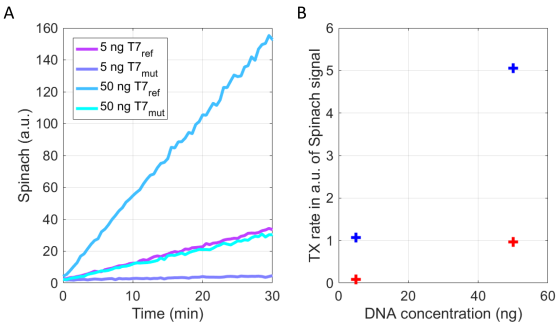


Figure 7.1: Transcription rates in the PURE system can be reduced by changing the T7 promoter sequence. A. Transcription kinetics (fluorescent Spinach signal) from DNA templates containing a regular T7 promoter ($T7_{ref}$) or a mutated T7 promoter ($T7_{mut}$) for two different DNA concentrations. B. Transcription rates of the $T7_{ref}$ (blue) and $T7_{mut}$ (red) curves shown in A plotted for comparison.

7.1.1. REDUCING TRANSCRIPTION RATES BY MUTATING THE T7 PROMOTER SEQUENCE

With an eye on the phi29 DNA replication mechanism that we aim to implement robustly in the PURE system, we have tested a T7 promoter that has one point mutation (T instead of A at position -6) with respect to the consensus promoter and should reduce transcription levels significantly (around two-fold as reported in [2]). This could be a first step towards tuning transcription levels and subsequently protein expression when the four replication proteins are embedded on the same linear DNA template[1]. In the experiment shown in Figure 7.1, transcription from the mYFPco-LL-Spinach template containing the reference T7 promoter was compared with the transcription from the mYFPco-LL-Spinach template containing the mutated T7 promoter, for two different DNA concentrations. It is clear that the transcription rate of the template with the mutated promoter is reduced (5- and 13-fold, for the starting DNA concentrations of 5 ng and 50 ng, respectively) significantly. Further investigations are necessary to understand how the varying levels of transcripts of multiple genes correlate to their corresponding protein expression levels in the PURE system. Because the phi29 DNA replication system based on the p2, p3, p5 and p6 proteins requires low amounts of p2 and p3 but high amounts of p5 and p6 proteins for efficient DNA replication, a starting point for co-expression of the four phi29 replication proteins from the same linear template could be the implementation of this weaker T7 promoter upstream the p2 and p3 genes.

7.1.2. ALTERING GENE EXPRESSION BY VARYING THE TEMPERATURE

In this thesis, protein expression in the PURE system was performed at different temperatures. For the DNA replication project, the temperature was lowered from the standard 37 °C to 30 °C because there was an indication that the DNA replication reaction is more efficient at this temperature. Other reasons to express proteins at lower temperatures

[1] In this thesis, the p2 and p3 proteins were expressed from an oriLR-p2-p3 template and the p5 and p6 proteins separately from their own oriLR-p6-p5 template. We could thus tune expression levels to a certain degree by varying the DNA concentrations of the two templates.

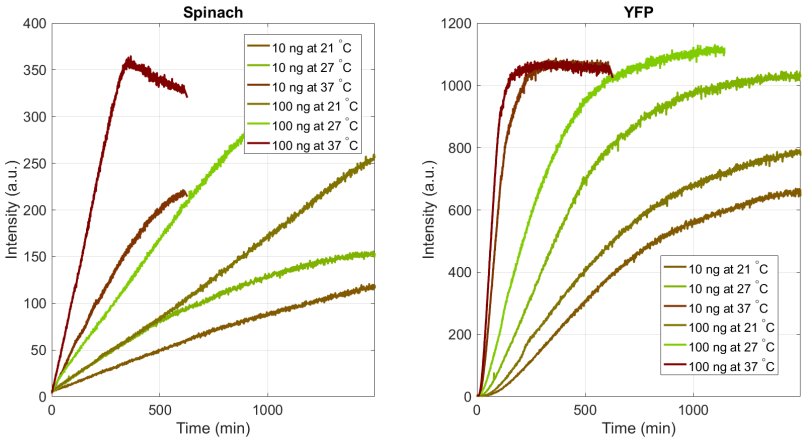


Figure 7.2: The influence of temperature on the apparent expression kinetics of the mYFP-LL-Spinach DNA template in the PURE system shows the general trend that lower temperatures lead to a longer protein expression time but compromise with a small reduction in the final protein yield.

than 37 °C are, as reported in literature, to prevent the expression of non-functional, aggregation prone proteins [13, 14] and to restore the coupling between transcription and translation by reducing the transcription speed of the T7 RNAP [15]. We have tested the influence of temperature on the expression of YFP in the PURE system, as shown in Figure 7.2. Expression at lower temperatures increases the duration of the expression time drastically, but reduces slightly the final yield of YFP expression (Figure 7.3 and Figure 7.17 and Tables 7.1,7.2 and 7.3). Practically this means that the extra time to express the same amount of proteins should be taken into account when designing experiments at temperatures lower than 37 °C. The realization of the prolonged expression time, however, provides a promising possibility in optimizing conditions for the minimal cell framework. Further investigations (modeling and analysis) are necessary to obtain a better understanding about which translation processes and rates are affected most by the change in temperature.

7.2. INFLUENCING DNA REPLICATION PROCESSES

An important aim described in this thesis is to implement DNA replication in the minimal gene expression system for the minimal cell framework. The optimization of the DNA replication reaction based on the synthesized phi29 proteins is an ongoing research topic in our lab. We here discuss a few possible ways to improve the efficiency of DNA replication in the near future.

7.2.1. IMPROVING DNTP/RNTP RATIO IN THE EXPRESSION ENVIRONMENT

As discussed in section 4.3.2, the presence of ribonucleotides (rNTPs) slows down the polymerization reaction of the DNAP, as the rNTPs compete with the dNTPs for the active site of the polymerase. A decrease in polymerization rate of the phi29 DNAP in the

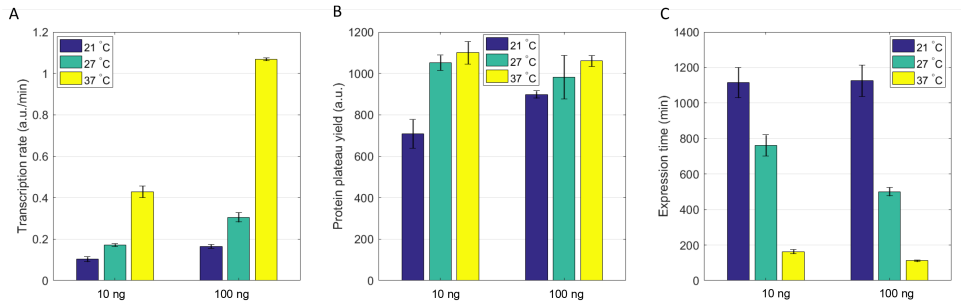


Figure 7.3: The influence of temperature on transcription (rate) and translation processes (final yield and expression time) in the PURE system is quantified for the expression of the mYFP-LL-Spinach DNA template at two concentrations. The protein curves are fitted to equation 3.5 as described in section 3.5. Mean and SEM values are plotted, for detailed data see the Tables in the SI. A) Transcription rates (determined by a linear fit through the Spinach fluorescence during the first 2h-time frame) are ca. four times higher at 37 °C than at 21 °C (Table 7.1). B) The final protein yields increase (slightly) with increasing temperature in the given range (Table 7.3). C) The protein expression time increases drastically (more than 6-fold) when the temperature is lowered to 21 °C compared to that of 37 °C (Table 7.2).

presence of rNTPs has been shown in literature for the *in vitro* phi29 DNA amplification system [16] and for the PURE system [17]. In Figure 7.4 we aimed to observe and quantify the decrease in polymerization rate of the phi29 DNAP due to the presence of rNTPs at PURE system concentrations (see also Table 4.1). For this purpose, an RCA reaction was performed by the phi29 DNAP on the primed ssM13 DNA in the presence or absence of rNTPs, quenched at several time-points, and imaged on an alkaline gel. The DNAP elongates a Cy5-labelled primer that was visualized with a fluorescence gel scanner. In Figure 7.4, it is clearly visible that the polymerization reaction proceeds slower in the presence of rNTPs, as the leading front takes 20 min in this reaction to reach a similar height as the leading front after 5 min in the polymerization reaction in absence of rNTPs. An estimate of the polymerization rate can be calculated from the time-point where the elongated strand first reaches the length (7.4 knt) of the ssM13 template. In the absence of ribonucleotides this happens already after (before) 2.5 min of incubation and the polymerization speed is thus estimated around 50 nt/s in these experimental conditions. In the presence of the ribonucleotides (8 mM total), the estimated elongation rate is around 4-fold lower (12 nt/s). The take home message is that the polymerization reaction by the DNAP, and thus the DNA replication efficiency, is slowed down in the presence of large amounts of rNTPs naturally present in the PURE system. The DNA replication reaction could thus benefit from optimizing the dNTPs/rNTPs ratio in the PURE system, as already proposed in [17]. However, a reasonable amount of rNTPs is always necessary for transcription-translation reactions, so an optimal balance should be found to compromise all occurring reactions. Nevertheless, a two-fold reduction of total rNTP concentration in the PURE system could be made based on the rNTP concentration present in *E. coli* cells (see Table 4.1). Further studies should investigate whether transcription-translation reactions do not suffer from a reduction of the proposed rNTP concentrations, and what the effect is on the polymerization speed of the DNAP.

Considering the high ratio of rNTP/dNTPs present, another question arises: does the phi29 DNAP incorporate rNMPs in the DNA template? This question has been investigated elegantly for a different DNAP, the *E. coli* Pol III, where the authors studied the effect of nucleotide pool imbalance and the incorporation rate of rNTPs [18]. This study found that the polymerization reaction is slowed down around 3-fold in the presence of 3 mM rATP alone, and that the frequency of rNMP incorporation is one per 2.3 kb during chromosome replication. The reported rNTP incorporation rate is deduced from a reduction in DNA intensity on gel when the sample was treated in alkaline conditions that cleave polymers at the position of an rNMP. We have investigated if cleavage of the elongated ssDNA in the RCA reaction (due to cleavage of at the site of an incorporated rNMP) could have biased our interpretation of the experiment described in 7.4 and other experiments analyzed in alkaline conditions. In Figure 7.5, we split an RCA reaction in four different samples that were treated with different loading buffers (LB), prior loading on gel. No difference in band intensities was observed between the four conditions, which indicates that the current protocol of alkaline treatment does not bias the amount of synthesized DNA on gel [2]. When one wants to investigate the incorporation of rNMP by the phi29 DNAP, it is thus recommended to use stronger alkaline conditions (0.3 mM NaOH) according to Yao. et al [18], or higher sample treatment temperatures. Another approach is to study the effect of rNTPs and the frequency of rNTP incorporation with single-molecule techniques such as magnetic tweezers, which have a much higher temporal resolution to detect error incorporations than bulk studies [19].

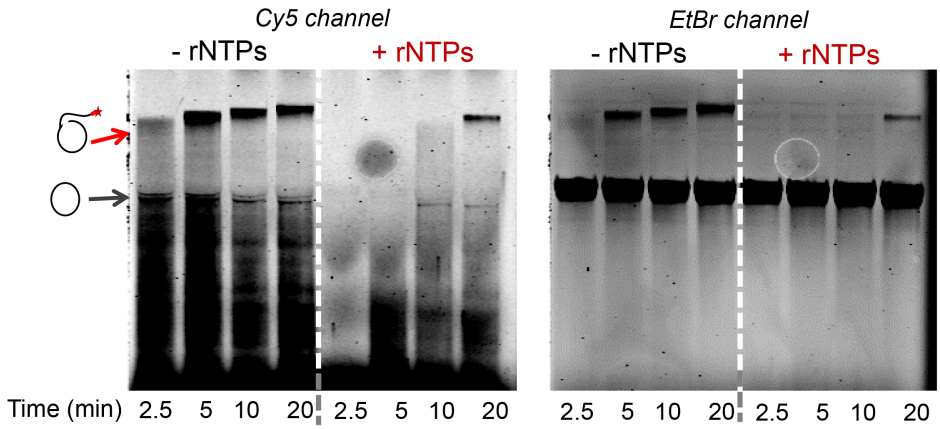


Figure 7.4: The RCA reaction performed by the phi29 DNAP on the (Cy5-)primed ssM13 template proceeds slower when ribonucleotides are present in the solution.

7.2.2. ALTERING DNA TEMPLATES FOR REPLICATION

Besides optimizing external conditions, adjusting the DNA templates of interest could also enhance the efficiency of the replication reaction. For example, altering the termini

[2] Apart from cleavage events that occur during the running of the alkaline gel.

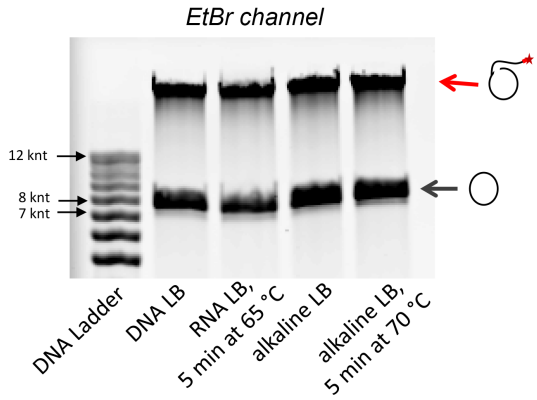


Figure 7.5: An RCA reaction performed by the phi29 DNAP is loaded on an alkaline gel with different buffers to compare the final sample quality. No difference is observed between the different treatments, which indicates that the ssDNA did not undergo cleavage under these specific conditions. Final loading buffer (LB) conditions are, DNA LB: 5% glycerol, RNA LB: 48% formamide, alkaline LB: 30 mM NaOH.

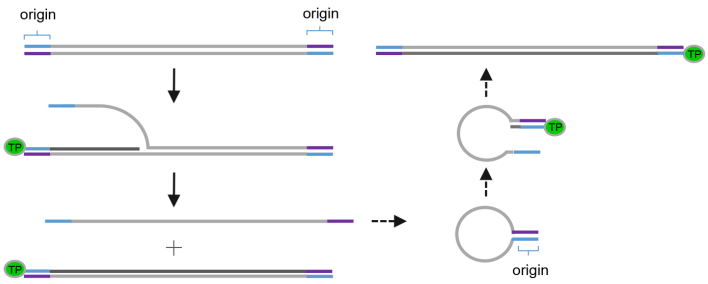


Figure 7.6: When DNA replication is only initiated at one end of the template by a TP-DNAP complex, single-strand displacement will result in the generation of ssDNAs. If the origins of replication are long (ca. 100 bp in adenoviruses) inverted terminal repeats, the ends of the ssDNA can fold upon itself creating a ds-handle and thus serve as a new initiation site. Figure is based on [21], DNAP molecule is not shown for simplicity.

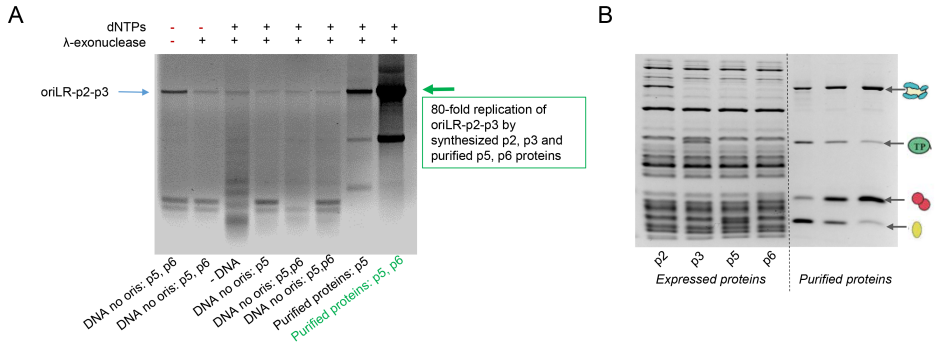


Figure 7.7: The amplification yield is dependent on the amount of p5 and p6 proteins present. A) The amplification of the oriLR-p2-p3 in the PURE*frex* was compared for expression of the p5 and p6 templates from their DNA templates to the addition of purified proteins, as described in Section 7.6. Samples were visualized on a neutral agarose gel after a lambda-exonuclease treatment and purification protocol as described in Section 7.6. The control in the first lanes does not contain dNTPs and had no lambda exonuclease treatment. Clear amplification of the oriLR-p2-p3 template is only visible for the samples where purified proteins were added. This implies that the amount of proteins produced in the PURE system under these co-expression conditions is not high enough to achieve good amplification yields. B) The amounts of synthesized phi29 DNA replication proteins in the PURE*frex* system (4 h expression at 30 °C) were estimated by band intensity analysis on a coommassie stained protein gel (12% polyacrylamide). As reference, the purified phi 29 proteins were loaded at three different known concentrations. The proteins were expressed separately from their own DNA templates to determine an upper limit of protein expression levels. Based on this (one time) gel, the estimate of the four synthesized proteins are 1.0 μM, 3.6 μM, 5.2 μM, 1.7 μM for the p2, p3, p5 and p6 proteins, respectively. Compared to the used amounts of proteins in the *in vitro* phi29 DNA amplification system [20], this is around 60 times higher for the p2 and p3, and 5 times lower for the p5 and p6 proteins.

sequences of the template could prevent the accumulation of ssDNA molecules upon the strand displacement reaction when the DNA replication initiation occurs only at one end of the template (see Figure 6.15). When the ends of the DNA are inverted terminal repeats, the released ssDNA can hybridize upon itself creating a double stranded origin sequence. The new dsDNA handle efficiently serves as an origin for a DNA replication reaction, resulting in another dsDNA molecule as shown in Figure 7.6. Thus, instead of having a distinct oriL and oriR sequence at the ends of the linear DNA template, that only have a 6-bp inverted terminal repeat, one could construct a DNA template having the oriL (or oriR) at both ends.

Comparing the activity of the synthesized DNAP and TP proteins on the TP-DNA phi29 genome and on the oriLR-p2p3 DNA (Figures 6.16A and 6.14), suggested that the presence of the terminal protein indeed stimulates the DNA replication reaction significantly in the absence of the p6 protein [20]. Considering that sufficiently high expression of both p5 and p6 proteins is hardly reachable in the minimal gene expression platform (SI Figure 7.7), the stimulating role of the parental TP might fulfil a crucial role in improving DNA replication efficiency. As a possible solution to reach efficient DNA replication without utilizing most of the resources to express high amounts of replication proteins, we thus suggest to introduce DNA templates with terminal proteins already covalently

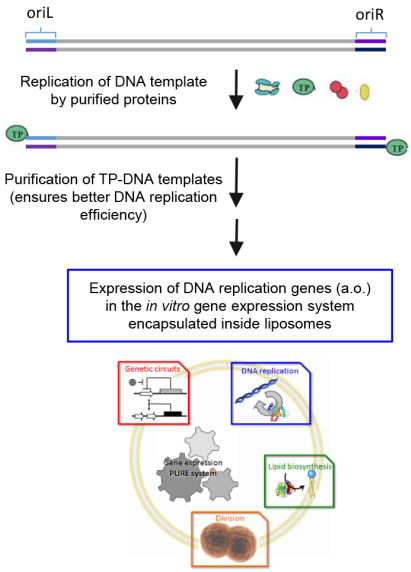


Figure 7.8: A strategy for improving the yield of dsDNA replication in the PURE system is to start with TP-DNA molecules. The terminal protein could be covalently attached to the DNA of interest by first replicating the DNA with the *in vitro* phi29 DNA replication system. The TP-DNA should be retained by a purification protocol that maintains the proper folding and functioning of the TP, such that it can stimulate the initiation reaction of the TP-DNAP complex.

attached to its 5'-ends (Figure 7.8). The DNA replication mechanism based on the phi29 virus might then be reduced to only 3 proteins, the DNAP, TP and the SSB protein.

7.3. DNA TEMPLATE CODING FOR THE PHI29 DNAP AND TP PROTEINS IS AMPLIFIED INSIDE LIPOSOMES

The next step after the identification of the phi29 DNA replication mechanism as a promising candidate, is to test if the DNA replication reaction is compatible with the encapsulation of the system into phospholid vesicles that serve as the core architecture for the minimal cell. We thus examined whether the DNA coding for the phi29 DNAP and TP proteins (oriLR-p2-p3) could be amplified inside liposomes. We first demonstrated that the phi29 DNAP could be synthesized in an active state inside liposome. For this, we performed the DNAP activity assay described in Section 6.5.3 and Figure 6.9, inside liposomes. The gene for phi29 DNAP (p2) was co-encapsulated with an oligo-primed ssDNA that can only serve as a template for transcription and translation, and produce the Spinach and YFP reporters, if the synthesized phi29 DNAP elongates the ssDNA into a transcriptionally active DNA template. Figure 7.9 shows that a significant number of liposomes exhibit YFP signal demonstrating that the vesicle-confined transcription-translation reaction produces active phi29 DNAP. In a control experiment lacking the phi29 DNAP gene, only very few liposomes displayed YFP fluorescence (not shown), pre-

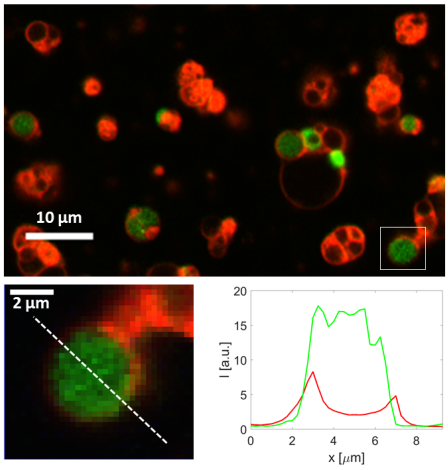


Figure 7.9: Elongation of primed ssDNA into transcriptionally active double-stranded YFP-Spinach DNA by the synthesized phi29 DNAP was confirmed inside liposoms. The top image is a fluorescence micrograph of immobilized liposomes (red membrane dye) with expressed YFP protein displayed in green. A blow-up of one YFP-expressing liposome (framed in the top image) is shown at the bottom, along with the corresponding line intensity profiles of the membrane dye and YFP signals.

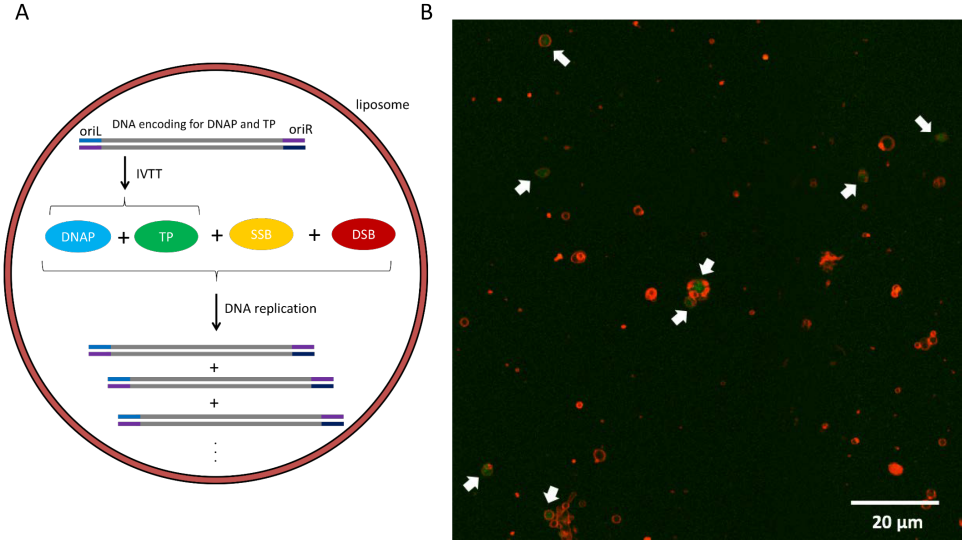


Figure 7.10: Coupled DNA replication, transcription and translation with positive feedback inside liposomes. A) Schematic of the reaction scheme. B) Fluorescence micrograph of immobilized liposomes (red membrane dye). The white arrows point toward liposomes exhibiting higher fluorescence signal of the DNA staining dye acridine orange (green), which corresponds to an accumulation of the phi29 DNAP- and TP-coding template.

sumably resulting from the expression of residual dsDNA.
For the replication reaction of the oriLR-p2-p3 inside liposomes, the purified SSB and DSB were supplemented to the reaction to increase the yield of DNA replication (see also Figure 7.7). Figure 7.10 shows that expression of the oriLR-p2-p3 template leads to increased DNA amounts inside liposomes as reflected by the higher fluorescence intensity of the intercalating dye. A phi29 DNAP- and TP-coding template devoid of the origin of replication sequences fails to produce liposomes with highly fluorescent lumen (not shown), demonstrating that DNA amplification is conditional to the presence of an origin of replication and that the initial amount of DNA is not sufficient to generate liposomes with higher fluorescence than the background signal.

7.4. STUDYING COLLISIONS BETWEEN THE PHI29 DNAP AND ROADBLOCK PROTEINS ON THE SAME DNA TEMPLATE

In Chapter 6, the question *What happens when a progressing phi29 DNAP encounters and collides with a T7 RNAP or another roadblock on the DNA template?* was raised. An example of what can happen in a collision of the phi29 DNAP and a DNA roadblock is given in Figure 6.3C. Here, the DNAP switches to the displaced strand, which eventually leads to the formation of small palindromic DNAs [22]. Template switching by the phi29 DNAP has been as well described for an RCA reaction on a plasmid using the phi29 DNAP, especially in the absence of SSB proteins [23]. Evidence for this behaviour was also found in *in vivo* studies, where the observation of existing short replicated DNA molecules was attributed to the consequences of the collision of the phi29 DNAP with a stationary DNA binding protein roadblock [24]. *In vitro* studies were performed by the group of Prof. Salas to investigate co-directional and head-on collisions of the *B. Subtilis* transcription machinery with the the phi29 DNAP. The phi29 system is interesting in this respect because the DNA replication is symmetrical and starts at both ends of the DNA template; evolution cannot have optimized the phi29 genome organization to favour either one direction of collision between transcription and DNA replication machineries. This is in contrast to most prokaryotic and eukaryotic systems where the most heavily transcribed genes are oriented in the direction of the leading strand of the replication fork (preventing the more detrimental head-on collisions [25]). The head-on collision between two DNAPs during the replication of the phi29 genome is unavoidable and is resolved by the displacement of the DNAPs non-template strand (analogous to the situation in Figure 7.12C.2). The head-on collision of a stalled ternary *B. Subtilis* RNAP complex completely blocks the progression of the phi29 DNAP, but both proteins remain bound to their template strand and the DNAP is able to resume elongation when the RNAP is allowed to translocate again [26]. The outcome of the co-directional collision is that the RNAP slows down the DNAP that cannot pass [27].
Research on collisions of replication forks and transcription units took off more than 25 years ago, when electron micrograph images reveiled for the first time that RNAPs were dislodged during replication in *E. coli* [25]. Shortly after, studies on head-on collisions between the T4 DNA replication fork and the *E. coli* RNAP showed that in presence of the helicase, the replication fork could bypass the RNAP after a short delay and both would

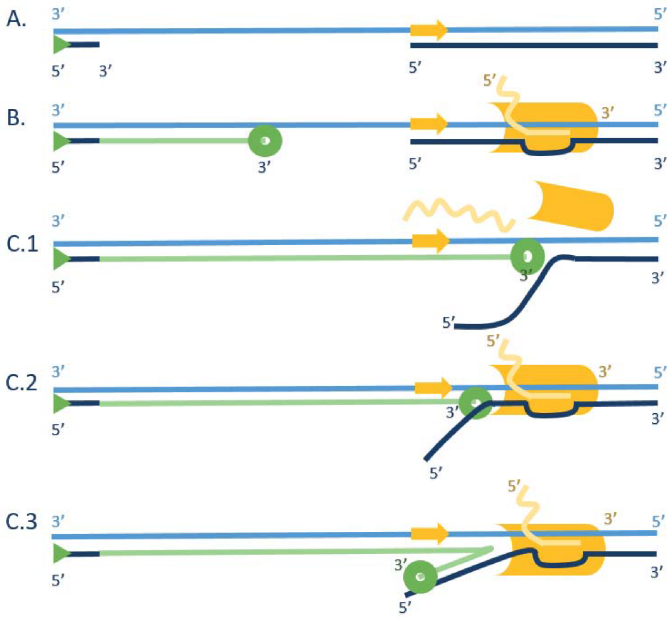


Figure 7.11: Schematic overview of possible outcomes of a co-directional collision between a DNAP (green donut) and an RNAP (yellow fish). Note that the same outcomes can occur when the roadblock is stationary. Relevant examples in literature are given in between brackets. A) Template scheme with a fluorescently labeled primer (green triangle) that serves as a PTJ for the DNAP. The yellow arrow indicates the directionality of transcription by the RNAP starting from the promoter sequence. B) DNAP elongates the primer and RNAP starts transcription; both enzymes make use of the same template strand. Different collision scenarios will follow for DNAPs capable of strand displacement. C.1) The DNAP kicks off the RNAP from the template ([25, 34]) and resumes elongation till the end of the molecule. C.2) The DNAP stalls behind the RNAP ([27, 34]). C.3) The DNAP switches to the displaced strand ([24]).

7

remain competent in resuming chain elongation [28]. Today, this field remains interesting and not only mechanisms to resolve collisions [29] as well as the consequences of collisions, such as collision-induced mutagenesis [30], are topics that receive attention in literature. In the context of the minimal cell framework, we are specifically interested to understand what happens when a phi29 DNAP encounters a (moving) T7 RNAP or a (stationary) LacI roadblock. The LacI protein, when bound to the lac operator, can repress transcription and is used for the implementation of feedback loops and genetic networks. To study these collisions and their possible outcomes, we designed an *in vitro* collision assay.

The idea of this assay is to be able to distinguish the different outcomes of a collision based on the length of a fluorescent primer that is elongated by a DNAP. The possible outcomes of a co-directional collision between a DNAP and RNAP are schematically depicted in Figure 7.11 and for a head-on collision in Figure 7.12. The assay is based on the construct design displayed in Figure 7.13 and should be able to distinguish between a *kick*, *stall*, or *switch* scenario. Due to the nature of the DNA template that is

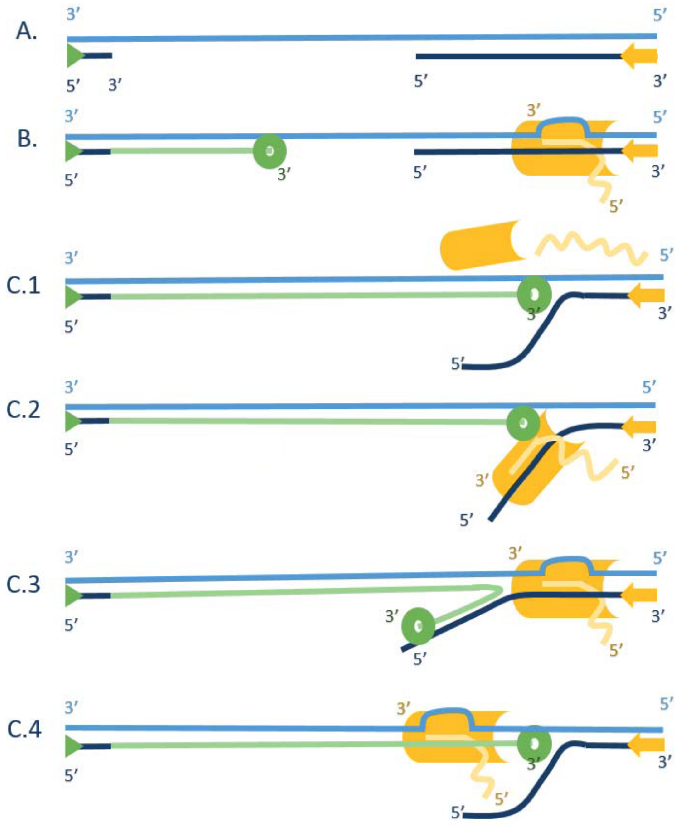


Figure 7.12: Schematic overview of possible outcomes of a head-on collision between a DNAP (green donut) and an RNAP (yellow fish). Relevant examples in literature are given in between brackets. Note that outcomes C.2 and C.4 are specific to moving roadblocks, and outcomes C.1 and C.3 are essentially the same as in Figure 7.11. A) Template scheme with a fluorescently labeled primer (green triangle) that serves as a PTJ for the DNAP. The yellow arrow indicates the directionality of transcription by the RNAP starting from the promoter sequence. B) DNAP elongates the primer and RNAP starts transcription; both enzymes make use of opposite template strands. C.1) The DNAP kicks off the RNAP, leading to aborted transcription products ([25]). For moving roadblocks such as the RNAP, it is in theory also possible that the DNAP is kicked off. C.2) The DNAP displaces its non-template strand, while the RNAP remains bound to its template strand and can resume transcription ([26, 33]. C.3) The DNAP switches to the displaced strand. C.4) The DNAP passes the RNAP and the RNAP switches to the newly synthesized DNA strand ([28]).

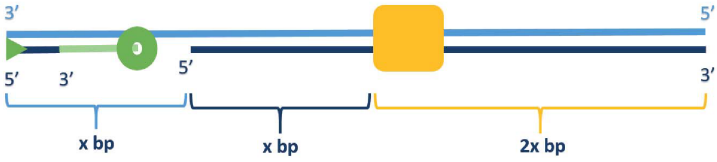


Figure 7.13: The assay construct is a DNA template that is partly double stranded and partly single stranded. The long strand (light blue) has a length of $4x$, whereas the shorter strand has a length of $3x$, where x can be any number in theory. The length of elongated fluorescently labelled primer (green triangle) reports on four different DNAP behaviours when encountering a roadblock (yellow rectangle, here a stationary roadblock). The lengths of produced DNA and interpretations are as follows: x: DNAP without strand displacement activity. $2x$: DNAP **stall**s in front of the roadblock. $3x$: DNAP **switch**es it template and elongates the primer until the end of the displaced strand. $4x$: DNAP **kick**s off the roadblock from the template and is competent to resume elongation after the collision.

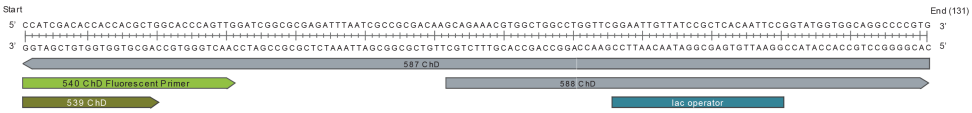


Figure 7.14: LacO construct to investigate collisions between the phi29 DNAP and the stationary LacI roadblock. The sequence of the lac operator is indicated in blue. Primers 587, 588 and 540 ChD were hybridized as described in Section 7.6. The lengths of produced DNA and interpretations of the collision outcomes are as follows (see also Figure 7.13): stall - ca. 85 nt, switch - ca. 109 nt, kick - 131 nt.

partially single-stranded and partially double-stranded, this assay demands a DNAP capable of strand displacement activity for a collision to occur. This assay can be extended for other DNAPs, for instance by introducing a non-complementary 5'-overhang at the double stranded DNA (distance x in Figure 7.13) that can load an accessory helicase enzyme. The main advantage of this assay is that it is simple to set up as it requires only the DNAP and the protein roadblock of interest. Choosing the right conditions (based on affinities of the roadbloack for the DNA template), can ensure that each template has a roadblock bound, and thus each DNAP that elongates a PTJ will lead to one event where it encounters a roadblock.

We designed one template that can study the collision between a phi29 DNAP and a LacI roadblock (Figure 7.14) and one that can study a head-on collision between the phi29 DNAP and a T7 RNAP (Figure 7.15). As a first step, we showed that the DNA oligos were hybridized and formed the hybrid ss-ds DNA constructs in Figure 7.16A. That the hybrid constructs were formed is evident from the disappearance of the shorter 31-nt and 71-nt oligos that thus present at the 130 nt hybrid band. Note that the mobility of the 130-nt ssDNA oligo (585 ChD) is similar to the mobility of the hybridized 130-nt templates. This is attributed to the fact that the relative mobility of ssDNA and dsDNA is similar for these small constructs on a polyacrylamide gel [31]. The binding of the LacI protein to the LacO construct was confirmed in Figure 7.16B. That transcription occurred from the T7 construct by the T7 RNAP was confirmed in Figure 7.16C.
This preliminary data showed that the current protocol has to be adjusted to investigate the length of the produced DNA, as the hybrid construct (see Figure 7.16C, lane 1) was

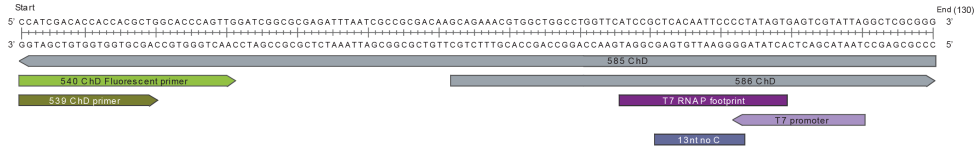


Figure 7.15: T7 construct to investigate head-on collisions between the phi29 DNAP and a T7 RNAP roadblock. Primers 585, 586 and 540 ChD were hybridized as described in Section 7.6. T7 promoter and T7 RNAP footprint are indicated [33]. A stable ternary complex of the RNAP-DNA, is achieved when the T7 RNAP is in elongation phase (no more aborted initiations); by supplying only three nucleotides (ATP, GTP, UTP), this complex is formed with a 13-nt transcript. The lengths of produced DNA and interpretations of the collision outcomes are as follows (see also Figures 7.13 and 7.12): stall - ca. 85 nt, switch - ca. 109 nt, kick or pass - 130 nt. A distinction between kick or pass can be made by investigating whether transcripts of RNA are produced upon adding the fourth missing nucleotide, CTP.

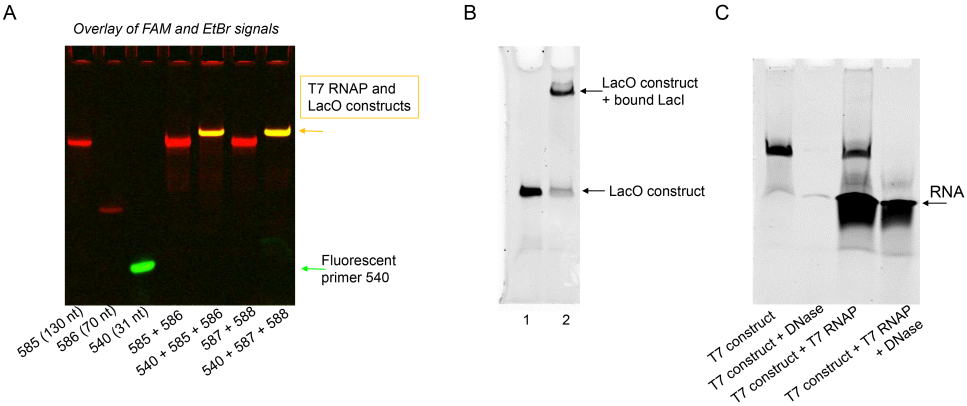


Figure 7.16: The DNA templates for studying collisions of the phi29 DNAP and roadblocks were constructed and the binding of roadblocks confirmed. A) The T7 (primers 585, 586 and FAM-540 ChD) and LacO (primers 587, 588 and FAM-540 ChD) were successfully hybridized and visualized on a 10% polyacrylamide gel by a fluorescence scanner. An overlay of the FAM (green) and EtBr (red) channels is shown. Yellow color indicates that the final constructs contain the fluorescent 540 ChD primer. B) Binding of LacI to the LacO construct was confirmed by a gel shift assay: the construct with bound LacI (lane 2) has a lower mobility than the naked DNA constuct (lane 1) and thus appears higher in the gel. Based on band intensity analysis, 85% of LacO constructs had LacI bound to it. C) The transcribability of the T7 construct was confirmed on a 10% polyacrylamide gel containing 7 M urea. The samples where the T7 RNAP and NTPs were added to the construct show a clear extra band that corresponds to RNA transcripts. Note that the hybrid construct is not denatured in lane 1.

not denatured with the current protocol. It is recommended to run gels in the presence of formaldehyde (or glyoxal) to denature the dsDNA hybrid irreversibly. Next steps include experiment with primer elongation by the phi29 DNAP on the LacO construct with or without LacI. Further studies could also include a DNA template with multiple lacO sites, such that a lacI tetramer forces the DNA to form loops that are harder to dissociate. The steps for the experiments of the T7 RNAP - phi29 DNAP collision are as follows (see also Figure 7.15: First, make a stable ternary DNA-T7 RNAP complex by halting the T7 RNAP right after it reaches elongation phase, this is achieved by adding only 3 of the 4 rNTPs. Second, add the phi29 DNAP and dNTPs to the reaction such that elongation of the fluorescent primer can start. To test if the stalled T7 RNAP can resume transcription after the collision with the phi29 DNAP, the 4th nucleotide should be added. To prevent free RNAPs in the solution to perform transcription (even if the original T7 RNAP was kicked off the template, transcription can start by free RNAPs that bind anew to the promoter sequence), an inhibitor molecule such as heparin [32] should be added right after the addition of the phi29 DNAP.

For the minimal cell framework, it would be ideal if we would find that neither the phi29 DNAP, nor the T7 RNAP, will loose its connection with its template strand in a head-on collision and can resume the polymerization reaction. This would be similar to a collision of two phi29 DNAPs or the collision of a T7 RNAP and a T3 RNAP [33]. When a head-on collision between the phi29 DNAP and the T7 RNAP is not detrimental to transcription or DNA replication, the compatibility of these enzymes strengthens the possible successful implementation of the phi29 DNA replication system for the minimal cell.

7.5. CONCLUSION

We set out in this thesis to gain new insights on a minimal gene expression system, its expression in liposomes, and to identify a promising candidate for genome replication in the minimal cell. We have learned about the quantitative dynamics of expression of a single gene, the YFP-Spinach, in bulk and in liposomes. We found that protein expression in liposomes reaches in some cases significantly higher levels than in bulk reactions. The liposome, apart from being the core architecture of the minimal cell framework, thus brings the advantage of high protein expression levels. These are necessary when multiple genes are expressed that will perform a variety of biochemical functions. For example, the minimal gene expression system we have utilized as a platform alone already consists of more than 9 μM of proteins, excluding the ribosome (present in concentrations of about 1 μM and consisting of 55 proteins). To make this transcription-translation system autonomous will be a challenging task for the future. We have focussed on extending the transcription-translation system with a DNA replication mechanism. The reaction where a minimal set of DNA replication proteins (phi29 DNAP and TP) is synthesized from and catalyses the synthesis of its own DNA template can be seen as an autocatalytic DNA replication cycle, thus recapitulating the central dogma of biology in liposome compartments. The results presented in this thesis open the door to implement evolution in the framework of a DNA-based minimal cell. In analogy to *in vivo* cellular mechanisms, one can envisage that random mutations occurring during

DNA replication through activity of the wild-type or an error-prone phi29 DNAP would lead to a gain of function of the corresponding protein or network of proteins, thus providing an evolutionary advantage for that specific minimal cell.

7.6. MATERIALS AND METHODS

RCA IN PRESENCE OF RNTPS

RCAs were performed by the commercial phi29 DNAP (20 units, NEB) and accompanying buffer as previously described in Section 6.8. The (+ rNTP) sample contained 3 mM of both rATP and rGTP and 1 mM of both rCTP and rUTP. The dNTP concentration was 0.3 mM of each dNTP.

AMPLIFICATION OF DSDNA TEMPLATE BY EXPRESSED AND PURIFIED PHI29 PROTEINS

In Figure 7.7, the DNA template for replication in the PURE*frex* reaction is the oriLR-p2-p3, present in all eight samples. To bias expression towards the p5 or p6 proteins, the first two hours of reaction were incubated at 30 °C with a lower amount of oriLR-p2-p3 DNA template resulting in a ratio of 40 ng p5 (no oris), and/or 40 ng p6 (no oris) and only 17 ng of oriLR-p2-p3 in 20 μl reaction. After two hours, 123 ng of oriLR-p2-p3 template was added to all samples (140 ng final) to have enough input DNA template for the replication reaction. The samples were incubated an additional 4 h at 30 °C, prior to lambda exonuclease treatment (1 μl to 10 μl, incubation 40 min at 30 °C) and the purification protocol as described in Section 6.8. The samples were imaged on a 1.1% agarose gel stained with EtBr.

ELONGATION OF PRIMED SSDNA BY SYNTHESIZED PHI29 DNAP IN LIPOSOME-CONFINED REACTIONS

The ssDNA ΔT7p-eYFP-co-LL-Spinach template was prepared by the assymetric PCR with primers 185 and 190 ChD as described in Section 6.8. Liposomes were formed by spontaneous swelling of a lipid film from lipid-coated beads. The lipid-coated beads were prepared with the following protocol: A mixture containing 1 mg DOPC, 0.25 mg DOPG, 0.06 mg DSPE-PEG(2000)-biotin (all from Avanti Polar Lipids) and 0.125 mg DHPE-Texas Red (Invitrogen), dissolved in chloroform, was prepared. This mixture was deposited drop-by-drop on 1 g of glass beads (212 μm -300 μm, acid washed, Sigma-Aldrich) that were preheated to 60 °C. After lipid deposition, residual chloroform was removed by 2 h of desiccation. Lipid-coated beads were stored under argon to protect lipids from oxidative damage. The swelling solution for the experiment and the control consisted of a 19-μl standard PURE*frex* 2.0 reaction that was supplemented with 200 μM dNTP mixture (Promega), 10 units SUPERase inhibitor (Thermo Fisher), 100 μM PTH2 and 222 ng ss/dsDNA mixture of ΔT7p-eYFPco-LL-Spinach-polyA DNA that was pre-hybridized with 1 μM primer ChD 361 as described in section 6.8. The solution was split in two, and one aliquot was supplemented with 33 ng phi29 DNAP DNA, whereas an equal volume of RNase free water was supplemented to the aliquot of the negative control. To both 10-μl reactions, 7.5 mg of lipid-coated beads was added. Liposomes were formed by spontaneous swelling of the lipid film for 2 h on ice, during which both samples were subjected to regular tumbling. For immobilization of the liposomes, custom-made glass reaction chambers were coated with a mixture of BSA and BSA-biotin, and subsequently with

Neutravidin to provide an immobilization surface for the biotinylated liposomes. 2 μl of the liposome solution was pipetted carefully (with a cut tip to avoid damaging the liposomes) into the reaction chamber. After 1 min of immobilization, 5 μl of feeding solution (a 1:1 mixture of PURE*frex* 2.0 Solution I and a buffer of 50 mM HEPES, 100 mM potassium glutamate, pH 7.6), supplemented with 33 μM DFHBI and 0.25 mg/ml Proteinase K, was added to the liposome solution in the reaction chamber. The reaction was incubated overnight at 37 °C. Imaging of the samples was performed with a Nikon A1R laser scanning confocal microscope. During image acquisition, the sample temperature was maintained at 37 °C. Acquisition was performed with the following excitation/emission wavelengths: 457/482 nm (Spinach), 514/540 nm (YFP), and 561/595 nm (Texas Red). Two independent experiments were conducted.

AMPLIFICATION OF DSDNA TEMPLATE BY SYNTHESIZED PHI29 DNAP AND TP PROTEINS IN LIPOSOME-CONFINED REACTIONS

The liposomes experiments were prepared with the same procedure as described in the previous section. The control p2-p3 template devoid of origin sequences was prepared by a PCR reaction with the primers 73 ChD and 181 ChD (see Section 6.8). A pooled reaction of 16-μl standard PURE*frex* 2.0 reaction, supplemented with 250 μM dNTPs each, 10 units SUPERase inhibitor, 100 μM PTH2, 20 μM of purified p5 and 8 μM of purified p6, was prepared. The solution was split in two and supplemented with 100 ng DNA of either oriLR-p2-p3 or p2-p3 (control), plus RNase free water to reach a final volume of 10 μl for each reaction. To these aliquots, 7.5 mg of lipid-coated beads were added and liposomes were formed as described before. About 2 μL of the liposome-containing solution was transferred to nuclease free PCR tubes. Next, 5 μL of feeding solution (a 1:1 mixture of PURE*frex* 2.0 Solution I and a buffer of 50 mM HEPES, 100 mM potassium glutamate, pH 7.6), supplemented with 0.25 mg/ml Proteinase K, was added. The reaction mixture was incubated for 3 h at 30 °C. Surface functionalization and liposome immobilization were performed as described in the previous section. To stain and detect dsDNA, the extraliposomal solution was supplemented with acridine orange (AO, Sigma-Aldrich; final concentration 9.5 μM). We chose AO because (i) it is a low-molecular weight nucleic acid staining dye (compared e.g. to YOYO-1 and SYTOX-green), which is an important factor for translocating across the liposome membrane, and (ii) binding to dsDNA can be discriminated from binding to ssDNA and RNA through AO spectral properties, thus reducing background fluorescence from PURE system components (rRNA, tRNA) and produced mRNA. After 30 min of incubation at 37 °C, imaging of the samples was performed with a Nikon A1R laser scanning confocal microscope using the following excitation/emission wavelengths: 488/525 nm (AO, settings specific to detect formation of complex with dsDNA), and 561/595 nm (Texas Red). Two independent experiments were conducted.

PREPARATION COLLISION CONSTRUCTS

DNA primers 540, 585-588 ChD were ordered at Ella Biotech. Primer 540 ChD is labelled with a 5'-FAM fluorescent dye. Hybridisation of the primers was performed in the PURE buffer (50 mM HEPES, 100 mM potassium glutamate, 13 mM magnesium acetate, pH = 7.6) by heating the samples at 95 °C and cooling it down to 14 °C in steps of 1 °C every 4

min. Samples are loaded on a 8% PAGE gel, after-stained with EtBr and imaged on a Typhoon scanner in the FAM (excitation: 488 nm, emission: 520 BP 40) and EtBr channels (excitation: 532 nm, emission: 610 BP 30). LacI was purified in house by former colleague Jacqueline Enzlin. The purified lacI had a stock concentration of 9 μM and was stored in 10% glycerol, 25 mM Tris pH 8.0, 200 mM NaCl and 5 mM β-mercaptoethanol, at - 80 °C.

ROADBLOCK BINDING TO COLLISION CONSTRUCTS

For the experiment in Figure 7.16B, 0.33 μM of hybrid LacO construct was mixed with 3 μM LacI in the PURE buffer and incubated for 5 min at room temperature, prior loading on gel with a 6x DNA loading buffer (Promega). For the experiment in Figure 7.16C, 0.25 μM of T7 construct was mixed in a 20 μl with 1 mM of each rNTPs, 1 mM DTT and 2 μl of T7 polymerase (RiboMAX^{TM} Large Scale RNA production kit, Promega) in the PURE buffer. The reaction mixture was incubated for 30 min at 37 °C and half of the sample was treated with 1 μl DNase I (Promega) for 15 min at 37 °C, after which both samples (- DNase, + DNase) were treated at room temperature for 15 min with 1 μl of 0.1 mg / ml Proteinase K solution. As a control, the same reactions were performed with a solution of T7 construct without the addition of T7 RNAP. Samples were heated for 5 min at 65 °C in a RNA loading buffer (NEB, containing 50 % formamide) and analyzed on a 10% polyacrylamide gel containing 7M urea.

7.7. ACKNOWLEDGEMENTS

I would like to thank Emma Gerritse for preliminary characterization of the influence of temperature on gene expression, and for her contribution to the Figures and experiments described in Section 7.4. Great thanks to Ilja Westerlaken, for experimental support and interest in the phi29 projects, to Duco Blanken for performing the liposome experiments and to Christophe for bringing all the efforts together. Thanks to Margarita Salas and Mario Mencia for providing us with the purified phi29 proteins.

7.8. SUPPLEMENTARY INFORMATION

Table 7.1: Summary of fitted values of the transcription rates at different temperatures.

T	input DNA	#N curves	mean TX rate	SEM (a.u./min)	relative to 37 °C
21 °C	10 ng	5	0.10 (a.u. / min)	0.01	0.24
27 °C	10 ng	4	0.17 (a.u. / min)	0.01	0.4
37 °C	10 ng	4	0.43 (a.u. / min)	0.03	1
21 °C	100 ng	4	0.17 (a.u. / min)	0.01	0.15
27 °C	100 ng	3	0.31 (a.u. / min)	0.02	0.29
37 °C	100 ng	3	1.07 (a.u. / min)	0.01	1

Table 7.2: Summary of fitted values of the expression time ($T_{plateau}$) at different temperatures.

T	input DNA	#N curves	$T_{plateau}$ (min)	SEM (min)	relative to 37 °C
21 °C	10 ng	5	1116	85	7
27 °C	10 ng	4	761	59	5
37 °C	10 ng	4	163	11	1
21 °C	100 ng	4	1126	89	10
27 °C	100 ng	3	501	24	4
37 °C	100 ng	3	112	4	1

Table 7.3: Summary of fitted values of the final protein yield at different temperatures.

T	input DNA	#N curves	Protein yield (a.u.)	SEM (a.u.)	relative to 37 °C
21 °C	10 ng	5	708	69	0.6
27 °C	10 ng	4	1051	37	1
37 °C	10 ng	4	1099	54	1
21 °C	100 ng	4	898	17	0.8
27 °C	100 ng	3	981	105	0.9
37 °C	100 ng	3	1060	26	1

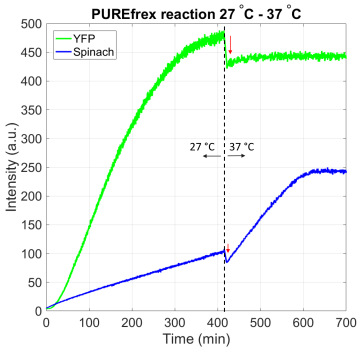


Figure 7.17: The fluorescence signal of the YFP and Spinach measured at the spectrophotometer in a PURE system reaction is temperature dependent. At 27 °C (and 21 °C), the fluorescence signal is 1.1 and 1.2 times higher for the YFP and Spinach, respectively. This means that the conversion factors, from a.u. to nM concentrations, are lowered by the same factor for 27 °C compared to the determined conversion at 37 °C. The data in Figure 7.2 are not corrected for this factor, thus the actual concentration of expressed YFP at 27 °C and 21 °C is overestimated by about 10% when comparing merely a.u. of fluorescence.

REFERENCES

[1] J. J. Dunn, F. W. Studier, and M. Gottesman, *Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements,* Journal of Molecular Biology **166**, 477 (1983).

[2] F. Chizzolini, M. Forlin, D. Cecchi, and S. S. Mansy, *Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength.* ACS synthetic biology (2013), 10.1021/sb4000977.

[3] T. H. Segall-Shapiro, A. J. Meyer, A. D. Ellington, E. D. Sontag, and C. a. Voigt, *A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase.* Molecular systems biology **10**, 742 (2014).

[4] D. K. Karig, S. Iyer, M. L. Simpson, and M. J. Doktycz, *Expression optimization and synthetic gene networks in cell-free systems.* Nucleic acids research **40**, 3763 (2012).

[5] H. Niederholtmeyer, L. Xu, and S. J. Maerkl, *Real-Time mRNA Measurement during an in Vitro Transcription and Translation Reaction Using Binary Probes,* ACS synthetic biology **2**, 411 (2013).

[6] H. Niederholtmeyer, V. Stepanova, and S. J. Maerkl, *Implementation of cell-free biological networks at steady state,* Proceedings of the National Academy of Sciences of the United States of America **110**, 15985 (2013).

[7] S. Kobori, N. Ichihashi, Y. Kazuta, and T. Yomo, *A controllable gene expression system in liposomes that includes a positive feedback loop.* Molecular bioSystems **9**, 1282 (2013).

[8] L. L. D. Maddalena, H. Niederholtmeyer, M. Turtola, Z. N. Swank, G. A. Belogurov, and S. J. Maerkl, *GreA and GreB Enhance Expression of Escherichia coli RNA Polymerase Promoters in a Reconstituted Transcription-Translation System,* ACS Synthetic Biology **5**, 929 (2016).

[9] M. K. Takahashi, J. Chappell, C. A. Hayes, Z. Z. Sun, J. Kim, V. Singhal, K. J. Spring, S. Al-Khabouri, C. P. Fall, V. Noireaux, R. M. Murray, and J. B. Lucks, *Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems,* ACS Synthetic Biology **4**, 503 (2015).

[10] Z. Z. Sun, E. Yeung, C. A. Hayes, V. Noireaux, and R. M. Murray, *Linear DNA for rapid prototyping of synthetic biological circuits in an escherichia coli based TX-TL cell-free system,* ACS Synthetic Biology **3**, 387 (2014), arXiv:arXiv:1011.1669v3 .

[11] J. Chappell, K. Jensen, and P. S. Freemont, *Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology,* Nucleic Acids Research **41**, 3471 (2013).

[12] D. D. Lewis, F. D. Villarreal, F. Wu, and C. Tan, *Synthetic biology outside the cell: linking computational tools to cell-free systems.* Frontiers in bioengineering and biotechnology **2**, 66 (2014).

[13] T. Niwa, B.-W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, and H. Taguchi, *Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins.* Proceedings of the National Academy of Sciences of the United States of America **106**, 4201 (2009).

[14] K. Fujiwara, T. Katayama, and S.-I. M. Nomura, *Cooperative working of bacterial chromosome replication proteins generated by a reconstituted protein expression system.* Nucleic acids research , 1 (2013).

[15] M. B. Iskakova, W. Szaflarski, M. Dreyfus, J. Remme, and K. H. Nierhaus, *Troubleshooting coupled in vitro transcription-translation system derived from Escherichia coli cells: synthesis of high-yield fully active proteins.* Nucleic acids research **34**, e135 (2006).

[16] L. Blanco, A. Bernads, J. M. Lazaro, G. Martin, C. Garmendia, and M. Salas, *Highly Efficient DNA Synthesis by the Phage phi 29 DNA Polymerase,* The Journal of biological chemistry **264**, 8935 (1989).

[17] Y. Sakatani, N. Ichihashi, Y. Kazuta, and T. Yomo, *A transcription and translation-coupled DNA replication system using rolling-circle replication,* Scientific Reports **5**, 10404 (2015).

[18] N. Y. Yao, J. W. Schroeder, O. Yurieva, L. A. Simmons, and M. E. O'Donnell, *Cost of rNTP dNTP pool imbalance at the replication fork.* Proceedings of the National Academy of Sciences of the United States of America **110**, 12942 (2013).

[19] D. Dulin, I. D. Vilfan, B. A. Berghuis, S. Hage, D. H. Bamford, M. M. Poranen, M. Depken, and N. H. Dekker, *Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase,* Cell Reports **10**, 983 (2015).

[20] M. Mencía, P. Gella, A. Camacho, M. de Vega, and M. Salas, *Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage Phi29.* Proceedings of the National Academy of Sciences of the United States of America **108**, 18655 (2011).

[21] M. L. DePamphilis and S. D. Bell, *Replicons,* in *Genome duplication* (Garland Science, 2011) pp. 169–189.

[22] J. Esteban and L. Blanco, *In vitro evolution of terminal protein-containing genomes,* Proceedings of the National Academy of Sciences of the United States of America **94**, 2921 (1997).

[23] C. Ducani, G. Bernardinelli, and B. Hogberg, *Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA,* Nucleic Acids Research , 1 (2014).

[24] V. Murthy, W. J. J. Meijer, L. Blanco, and M. Salas, *DNA polymerase template switching at specific sites on the phi29 genome causes the in vivo accumulation of subgenomic phi29 DNA molecules,* Molecular Microbiology **29**, 787 (1998).

[25] S. French, *Consequences of replication fork movement through transcription units in vivo.* Science (New York, N.Y.) **258**, 1362 (1992).

[26] M. Elías-Arnanz and M. Salas, *Resolution of head-on collisions between the transcription machinery and bacteriophage Φ29 DNA polymerase is dependent on RNA polymerase translocation,* EMBO Journal **18**, 5675 (1999).

[27] M. Elías-Arnanz and M. Salas, *Bacteriophage ϕ29 DNA replication arrest caused by codirectional collisions with the transcription machinery,* EMBO Journal **16**, 5775 (1997).

[28] B. Liu and B. M. Alberts, *Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex.* Science (New York, N.Y.) **267**, 1131 (1995).

[29] C. N. Merrikh, B. J. Brewer, and H. Merrikh, *The B. subtilis Accessory Helicase PcrA Facilitates DNA Replication through Transcription Units,* PLOS Genetics **11** (2015), 10.1371/journal.pgen.1005289.

[30] T. S. Sankar, B. D. Wastuwidyaningtyas, Y. Dong, S. A. Lewis, and J. D. Wang, *The nature of mutations induced by replication-transcription collisions,* Nature , 1 (2016).

[31] T. Maniatis, A. Jeffrey, and H. Van deSande, *Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis,* Biochemistry **14**, 3787 (1975).

[32] R. Ferrari, C. Rivetti, and G. Dieci, *Transcription reinitiation properties of bacteriophage T7 RNA polymerase,* Biochemical and Biophysical Research Communications **315**, 376 (2004).

[33] N. Ma and W. T. McAllister, *In a Head-on Collision, Two RNA Polymerases Approaching One Another on the Same DNA May Pass by One Another,* Journal of Molecular Biology **391**, 808 (2009).

[34] P. A. Pavco and D. A. Steege, *Characterization of elongating T7 and SP6 RNA polymerases and their response to a roadblock generated by a site-specific DNA binding protein.* Nucleic acids research **19**, 4639 (1991).

SUMMARY

This thesis contributes to the minimal cell project by gaining a better quantitative understanding of the transcription and translation processes in the minimal gene expression system and by defining and investigating the DNA replication module.

Life has evolved on Earth for more than three billion years, and today we are here to admire its extreme variety and beautiful complexity. The question: *How did life originate?* is a most captivating one for humanity as it leaves plenty of room for imagination and speculation. It can be subdivided in two parts: one concerns the specific environment and molecules involved, in other words, the transition from non-living to living matter, or so to say from chemistry to biology. The second part relates closely to the question *What is life?* Many different answers can be found in literature and definitions can differ from person to person. A slightly different question has emerged from the origins of life field: *How would a cell that contains the minimal and sufficient number of components to perform the basic functions of cellular life look like?* In this question, cellular life is defined as the capability to display the properties of self-maintenance, reproduction and evolution. The search will gradually reveal an answer to the reduction of complexity of modern cells and will thus lead to the construction of a minimal cell. The construction of such an elementary cell will provide fundamental insights into the basic principles of life and their concerted functioning in a cell.

Is there an organism out there that closely resembles a minimal cell? This question prompts researchers to investigate our contemporary organisms that have very small genomes, as the concept of a minimal cell is closely related to the minimal-gene-set concept. There are two distinct approaches to find a minimal genome, the top-down and bottom-up approach. The top-down approach is inspired by the genomes of contemporary small organisms and aims to reduce them even further. The bottom-up approach is inspired by synthetic biology and has its roots in biochemistry. This approach aims to identify the gene products essential for the reconstitution of necessary reactions in a living cell. The main advantage of this bottom-up synthetic biology approach is that it enables the combination of genes from different organisms, prokaryotes and eukaryotes, as well as viral genes. It is thus not restricted to the evolution of one species.

The long-term goal of our lab is to construct a minimal cell using a bottom-up synthetic biology approach. Our aim is to achieve the functions of self-maintenance, reproduction and evolution starting with an extant minimal gene expression system that is compartmentalized inside a container defined by a lipid boundary. The gene expression is based upon the following core elements: DNA (heritable information), RNA (messenger) and protein (enzyme function or structure). Gene expression is defined in this thesis as the combination of the subsequent processes of transcription from DNA that

generates RNA molecules, and translation of the RNA molecules by the ribosome and co-factors producing proteins. The flow of information from DNA to RNA to protein is generally known as the central dogma of molecular biology. The machinery of these systems is itself made of proteins and RNA, so if we want the cell to be an autonomous system, we also need to incorporate all the information for the production of these constituents in the cell. Our first -more 'modest'- goals are to *expand the gene expression system* with the following modules: *DNA replication*, lipid biosynthesis and division. To find suitable components and mechanisms that can accomplish these tasks, we search for minimal systems of living cells such as bacteria, but also strategies from viruses. Once these are defined, DNA templates are designed for gene expression and we study the activity of the modules seperately. Finally, all successful models can be combined that will hopefully lead to a minimal cell.
Our platform for gene expression is a commercially available cell-free system that is reconstituted from purified components, termed 'protein synthesis using recombinant elements' (PURE) system. The PURE system consists only of these components that are absolutely necessary for transcription and translation to occur. The encapsulation of this gene expression system in a phospholipid vesicle to build cell-like bioreactors serves as a core architecture in constructing the minimal cell.

After explaining the context of the minimal cel as described above in Chapter 1, this thesis is divided in two parts: the quantification of gene expression processes and the study of the DNA replication module.
We started in Chapter 2 with combining the read-out of two orthogonal fluorescence labeling tools that report the amounts of produced mRNA and protein simultaneously and in real-time. The Spinach RNA aptamer and its fluorogenic probe were used for mRNA detection, while a fluorescent protein was expressed that was detected as well. Applying this dual-reporter assay to the analysis of transcript and protein production inside lipid vesicles after 3 h of expression revealed that their levels are uncorrelated, most probably a consequence of the low-copy number of some components in liposome-confined reactions.
In Chapter 3, we continued to improve the Spinach reporter and achieved an enhancement in fluorescence signal, which allowed us to monitor transcription with unprecedented sensitivity over longer time periods. The quantification of gene expression kinetics in individual vesicles over 5 h revealed that there are large differences between individual liposomes, corroborating our previous results. An interesting finding was that a subset of expressing liposomes produced larger amounts of mRNA and/or proteins than measured in bulk experiments. The quantitative descriptions of gene expression obtained in this research provide a valuable starting point to study factors that can enhance gene expression in bulk and in liposomes.

The second part of this thesis investigates the DNA replication module we set out to implement in the minimal gene expression system. DNA replication is at the core of information transfer, vital to the propagation of life. Nature has invented numerous different DNA replication mechanisms, and in Chapter 4 a brief overview of some of the least complex systems is given. DNA replication systems of prokaryotes and viruses pass

in review, as well as minimal *in vitro* DNA amplification systems. Because it is hypothesized that different DNA replication machineries first originated in the viral world before being transferred to the cellular world, and since these systems are less complex, viral DNA replication strategies serve as our prime inspiration source for the minimal cell.
In Chapter 5, the investigation of the viral strategy to use RNA as an intermediate information carrier is described. Herein, we propose how a nucleic acid amplification cycle can be coupled with transcription and translation processes. Although this system turned out to be a poor candidate for the minimal cell, we did learn something new about viral RNAPs. We report on the properties of an intermediate species of this cycle, the DNA:RNA hybrid template, that can be transcribed by the T7 RNAP, although with much lower efficiency than the double stranded DNA.
In Chapter 6, we propose the DNA replication mechanism by the bacteriophage phi29 proteins, DNA polymerase (DNAP), terminal protein (TP), single-stranded binding protein (SSB) and double-stranded binding protein (DSB), as a potential candidate for the DNA replication module in the minimal cell. We have designed templates encoding for two replication proteins flanked by ca. 200 bp of the right and left end of the phi29 genome containing the origins of replication. The replication activity of the PURE system synthesized phi29 DNAP and TP is confirmed by imaging the samples after a purification protocol on a nucleic acid gel and by analyzing the band intensities of the DNA of interest at different time-points. We showed that the DNA that encodes for these proteins is replicated, completing one round of the central dogma in biology in a minimal gene expression system. We found that the two synthesized proteins are also capable of amplifying the ca. 20-kb phi29 genome. This minimal replication system generates large amounts of side products in absence of the accessory proteins SSB ans DSB. We showed that co-expression of the four replication proteins reduces the side product formation in the amplification reaction of the phi29-genome. This study laid out the groundwork for further investigations.
Future work, as discussed in Chapter 7, includes the regulation of gene expression in the PURE system, as well as optimizing the conditions for DNA replication. Applying the knowledge of the first part of this thesis, where we found that a subset of expressing liposomes produced larger amounts of proteins than measured in bulk experiments, might in the near future prove beneficial for the efficiency of the DNA replication reaction. As a first proof of principle, we have shown that the synthesized phi29 DNAP and TP replication proteins are active and can replicate dsDNA inside liposomes.

In summary, this thesis represents small stepping stones in the work required to construct a minimal cell by gaining new insights on a minimal gene expression system, its expression in liposomes, and by identifying the phi29 DNA replication mechanism as a promising candidate for genome replication and evolution. We hope that this will aid the next generation of researchers in their journey to understand and reconstruct the basic principles of life.

"The known is finite, the unknown infinite; intellectually we stand on an islet in the midst of an illimitable ocean of inexplicability. Our business in every generation is to reclaim a little more land." T.H. Huxley, 1887

SAMENVATTING

Dit proefschrift draagt bij aan het minimale cel project door een beter kwantitatief begrip te krijgen van de transcriptie en translatie processen die plaatsvinden in het minimaal genexpressie systeem, en door het definiëren en bestuderen van de DNA replicatie module.

Het leven op aarde is meer dan drie miljard jaar geleden ontstaan, en vandaag de dag kunnen wij, als mens, genieten van de enorme variatie en bewonderingswaardige complexiteit van al het leven om ons heen. De vraag: *Hoe is het leven ontstaan?* is daarom ook altijd een fascinerende vraag geweest voor de mensheid die genoeg ruimte laat voor onze fantasie en diepe overpeinzingen. Deze vraag kan opgesplitst worden in twee delen: het eerste deel gaat over de specifieke omgeving en moleculen die betrokken waren bij de belangrijke transitie van niet-levend naar levend materiaal, ofwel, wanneer en hoe werd scheikunde opeens biologie? Het tweede deel is natuurlijk gerelateerd aan de vraag: *Wat is leven?* In de wetenschappelijke literatuur zijn veel verschillende antwoorden op die vraag te vinden en definities van 'leven' verschillen van persoon tot persoon. Een enigszins andere maar gerelateerde vraag is naar voren gekomen uit het onderzoek naar het ontstaan van het leven: *Hoe zou een cel eruit zien die alleen het minimale aantal benodigde componenten heeft om de basisfuncties van het cellulaire leven uit te kunnen voeren?* In deze vraag is cellulair leven gedefinieerd als het vermogen van een cel om in haar eigen levensbehoeften en onderhoud te kunnen voorzien; een cel die kan reproduceren en vatbaar is voor evolutie. De zoektocht naar zo een cel zal op den duur een antwoord kunnen geven op de vraag in hoeverre de complexiteit van moderne cellen gereduceerd kan worden, oftewel, het zal leiden tot de constructie van een minimale cel. De constructie van deze elementaire cel zal ons fundamentele inzichten kunnen geven over de basisprincipes van het leven en hoe deze samenwerken in de totstandkoming van een cel.

Bestaat in de natuur een organisme dat erg lijkt op een minimale cel? Deze vraag heeft onderzoekers gestimuleerd om hedendaagse organismen te bestuderen die kleine genomen bezitten, omdat het concept van de minimale cel gerelateerd is aan het concept van een cel met 'het minimale aantal' genen. Er zijn twee verschillende methodes om een minimaal genoom te vinden, te weten de *top-down* ('van bovenaf') methode en *bottom-up* ('vanaf de bodem') methode. De top-down methode probeert de kleinste genomen die we kennen van hedendaagse organismen nog verder te verkleinen. De bottom-up methode vindt haar inspiratie in de synthetische biologie en heeft wortels in de biochemie. Deze methode heeft als doel te identificeren welke producten van genen essentieel zijn voor de reconstructie van de noodzakelijke processen in een levende cel. Het voornaamste voordeel van de bottom-up, synthetische biologie methode, is dat zij ons in staat stelt om genen van verschillende organismen, zoals prokaryoten, eukaryoten en dus ook virussen, te combineren. Hierdoor is de constructie van een minimale

cel dus niet beperkt door de evolutie van een specifieke soort.

Het langetermijndoel van onze onderzoeksgroep is de constructie van een minimale cel waarbij we gebruik maken van de bottom-up synthetische biologie methode. Ons streven is gericht op het tot stand brengen van een cel die in haar eigen levensbehoeften en onderhoud kan voorzien, kan reproduceren en vatbaar is voor evolutie. Ons beginpunt is een bestaand minimaal genexpressie systeem dat zich binnenin een compartiment, gedefinieerd door een afbakening van een lipide membraan, bevindt. Genexpressie is gebaseerd op de volgende kernelementen: DNA (erfelijke informatie), RNA (boodschapper van informatie, mRNA) en eiwit (met enzymfunctie of belangrijk voor de (infra-) structuur van een cel). Genexpressie is hierin verder gedefinieerd als de combinatie van transcriptie van het DNA dat RNA moleculen genereerd en het opeenvolgende proces van translatie van de RNA moleculen door het ribosoom en cofactoren dat leidt tot de productie van eiwitten. De informatiestroom van DNA naar RNA naar eiwit is algemeen bekend als het centrale dogma van de moleculaire biologie. De machines die de informatie van DNA naar eiwitten omzetten bestaan zelf ook uit eiwitten, dus als de cel een autonoom systeem zou zijn, zouden we ook de informatie voor de productie van al deze eiwitten moeten integreren. Onze eerste -meer bescheiden- doelen zijn om *het genexpressie systeem uit te breiden met* de volgende modules: *DNA replicatie*, biosynthese van lipiden en een delingsmechanisme. Om geschikte mechanismen en componenten te vinden die deze taken uit kunnen voeren, zoeken we naar minimale systemen in levende cellen als bacteriën, en ook strategieën van virussen passeren de revue. Wanneer we tevreden zijn met de keuze van een systeem, ontwerpen we DNA templates voor genexpressie en bestuderen we de activiteit van de modules afzonderlijk. Uiteindelijk kunnen alle succesvolle modules gecombineerd worden en zal dit hopelijk resulteren in een minimale cel.

Ons platform voor genexpressie is een commercieel verkrijgbaar celvrij systeem dat is samengesteld uit gezuiverde componenten, genaamd 'protein synthesis using recombinant elements'(PURE) systeem. Het PURE systeem bestaat uit slechts die componenten die noodzakelijk zijn voor het laten plaatsvinden van de transcriptie en translatie processen. Het omvatten van dit genexpressie systeem door een fosfolipide membraan (ook wel inkapseling genoemd) om een celachtige bioreactor te bouwen, dient als de kernarchitectuur in de constructie van een minimale cel.

Na de uitleg van de context van de minimale cel als hierboven beschreven in hoofdstuk 1, is het proefschrift opgedeeld in twee delen: de kwantificatie van genexpressie processen en het onderzoeken van de DNA replicatie module.
We zijn in hoofdstuk 2 begonnen met het toepassen van twee orthogonale fluorescente labeling tools, die in real time kunnen rapporteren hoeveel mRNA en eiwit er simultaan wordt geproduceerd. Voor mRNA detectie gebruikten we de 'Spinach' RNA aptamer en het bijbehorende molecuul dat gedurende de binding hieraan fluoresceert, terwijl een fluorescent eiwit dat tot expressie werd gebracht ook gedetecteerd kon worden. Het duo-reporter assay hebben we toegepast op de analyse van mRNA en eiwit productie in liposomen. We ontdekten dat na drie uur expressie de niveaus geen correlatie met

elkaar hebben. Dit is zeer waarschijnlijk een gevolg van het lage aantal moleculen dat van sommige componenten aanwezig is in de liposomen.
In hoofdstuk 3, hebben we geprobeerd de Spinach reporter te verbeteren en dat heeft geleid tot een toename van het fluorescente signaal, waardoor het mogelijk wordt transcriptie te volgen over langere tijdperiodes met ongekende gevoeligheid. De kwantificatie van de kinetiek van genexpressie in individuele liposomen gedurende vijf uur liet zien dat er grote verschillen zijn tussen liposomen, wat overeenkomt met ons eerdere resultaat. Een interessante observatie was dat enkele liposomen veel meer mRNA en/of eiwit produceerden dan we voorheen in reacties met grote volumes (een miljoen keer groter dan het volume in een liposoom) hebben waargenomen. De kwantitatieve beschrijvingen van genexpressie, die we in dit onderzoek hebben verkregen, biedt een waardevol uitgangspunt om factoren te bestuderen die genexpressie kunnen verbeteren in grote volumes en in liposomen.

Het tweede deel van dit proefschrift onderzoekt de DNA replicatie module die we willen implementeren in het minimale genexpressie systeem. DNA replicatie vormt de basis voor het overbrengen van informatie, een proces dat vitaal is voor de voortplanting van het leven. De natuur heeft een scala aan verschillende DNA replicatie mechanismen gecreëerd, waarvan in hoofdstuk 4 een klein overzicht van de minst complexe is opgenomen. Zowel mechanismen van prokaryoten, virussen en ook minimale *in vitro* DNA amplificatie systemen komen aan bod. De strategieën van virussen zijn onze belangrijkste inspiratiebron voor de minimale cel omdat de hypothese bestaat dat verscheidene DNA replicatie eiwitten eerst zijn ontstaan in virussen, voordat ze werden overgenomen door cellen en omdat deze vaak minder complex zijn.
In hoofdstuk 5 beschrijven we de studie naar een strategie van virussen om RNA te gebruiken als een tussentijdse informatiedrager. We doen een voorstel hoe een amplificatiecyclus van nucleïnezuur kan worden gekoppeld met transcriptie en translatie processen. Hoewel dit systeem een gebrekkige kandidaat is voor een minimale cel, hebben we toch iets nieuws geleerd over RNA polymerases van virussen. We rapporteren over de eigenschappen van een RNA:DNA hybride nucleïnezuurketen, een tussenvorm van de cyclus, waarop transcriptie door de T7 RNAP kan worden uitgevoerd. De transcriptie heeft echter wel een veel lagere efficiëntie dan die vanaf dubbelstrengs DNA.
In hoodstuk 6 stellen we het DNA replicatie mechanisme van het phi29 virus voor als potentiële kandidaat voor het DNA replicatie mechanisme voor de minimale cel. Dit mechanisme bestaat uit een DNA polymerase (DNAP), een terminal eiwit (TP), een enkelstrengs DNA bindingseiwit (SSB) en een dubbelstrengs DNA bindingseiwit (DSB). We hebben DNA templates ontworpen die de genen coderen van twee replicatie eiwitten, geflankeerd door ongeveer 200 baseparen van het rechter- en linker uiteinde van het phi29 genoom waarin de replicatie-oorsprongen zich bevinden. De replicatie activiteit van de door het PURE systeem gesynthetiseerde phi29 DNAP en TP is bevestigd door het zichtbaar maken van de samples op een nucleïnezuur gel en door het analyseren van de intensiteit van de DNA bandjes op verschillende tijdpunten. We hebben aangetoond dat het DNA dat codeert voor deze eiwitten is gerepliceerd; dat houdt in dat een volledige cyclus van het centrale dogma van de moleculaire biologie is uitgevoerd in het minimale genexpressie systeem. We hebben ook aangetoond dat de gesynthetiseerde eiwitten in

staat waren het ongeveer 20 kilobasenparen lange phi29 genoom te repliceren. Echter, het minimale systeem met slechts de twee eiwitten genereert een grote hoeveelheid van kleinere DNA bijproducten in de afwezigheid van SSB en DSB. Wanneer we de vier replicatie eiwitten tot co-expressie brengen is de formatie van bijproducten tijdens de amplificatie reactie van het phi29 genoom aanzienlijk verminderd. Dit onderzoek legt een stevige basis voor toekomstige studies betreffende phi29 DNA replicatie in een minimaal genexpressie systeem.
Bespreking van toekomstig werk is opgenomen in hoofdstuk 7, met onder meer de regulatie van genexpressie in het PURE system en hoe we de condities voor DNA replicatie kunnen optimaliseren. De toepassing van de vergaarde kennis van het eerste deel van dit proefschrift, waarin we in experimenten constateerden dat kleine hoeveelheden van liposomen een hogere expressie van eiwitten vertonen dan bulk hoeveelheden, kan in de nabije toekomst voordelig uitpakken voor de efficiëntie van de DNA replicatie reactie. Als een eerste 'proof of principle' hebben we laten zien dat de gesynthetiseerde phi29 DNAP en TP replicatie eiwitten actief zijn en dsDNA kunnen repliceren in liposomen.

Kort samengevat beschrijft dit proefschrift kleine 'stepping stones' in het werk dat nodig is om een minimale cel te construeren, door nieuwe inzichten te verkrijgen over een minimaal genexpressie systeem, de expressie ervan in liposomen, en door het identificeren van het phi29 DNA replicatie mechanisme als veelbelovende kandidaat voor genoom replicatie en evolutie. We hopen dat dit de volgende generatie van onderzoekers zal helpen in hun missie om de basisprincipes van het leven te begrijpen en te reconstrueren.

ACKNOWLEDGEMENTS

These pages do not contain any scientific content; yet they are most valuable for so many people. No one is ever really honest on these pages in describing difficult periods during the course of the PhD time. While some of the encountered struggles originate from the consequences of the current academic system, some of them are just very personal. For many of my friends, and for me, the PhD is an extremely valuable experience not only because we enjoy to work on our passion, but also because we learn to deal with and overcome all kind of obstacles. We get to know ourselves quite well and we learn; gradually we become more balanced persons. This is thanks to all the support, feedback and advice, friendship and love we received from the people surrounding us. I am grateful to all of you, and am very happy if you are here to share this special day and moments of happiness with me.

Dear Christophe, I have never forgotten the moment I first walked into your office. Before I could nervously finish 'Can I ask you a question, about...' you answered 'Anything I can do to help', with a welcoming arm gesture, moving from your desk to the table where I could join you. Your genuine kindness struck me and this has never changed since. I have enormous respect for you, you were great and always respectful even in the times I struggled most. Till this day you amaze me. I truly appreciated the balance of support and freedom you gave me in research, the open scientific atmosphere in the lab you created and the many opportunities you gave me to attend conferences, workshops, to meet other scientists and to learn more. It feels weird to leave your lab after all those years we spent together on interesting science, discussing ideas, having fun on lab-trips and making jokes. I will cherish and carry the good memories with me, and whole-heartedly wish you and your lab all the joy and success for the future. Thank you for everything.
Zohreh 'joon', I also remember our first conversation ('boos') where I immediately had the pleasure to enjoy your merry laughter that is exemplary for all the positivism you brought in the office and in the lab. I enjoyed the many discussions on science, culture, food, family and thousand other things, and I am grateful for the support you gave me and for teaching the art of making liposomes. Thanks for all the quality time we shared outside the lab, also with Davini and Soheil of course. I am happy to announce we are soon officially life-long friends according to Ellinor. Ilja, thanks for introducing me to the bench with your ever-lasting enthusiasm. I really appreciate all your input and work on our projects in the last years and the pipetting together, with some rock(y) or push-up intermezzos. You are a sunshine inside and outside the lab and your altruism is admirable. I feel honoured to have two of such knowledgeable, funny, sweet and strong persons as paranymphs. Andrew, I miss our walks, random conversations, discussions on life, hugs, dinners, and you doing another sketch that made us all laugh. You are one of a kind and knowing you for so many years brings me a lot of happiness. David, though I

did not choose to do a master project with you, we soon ended up playing fussball and football together, discussing about the beauty of viruses and transcription, having dinners at the sports centre, listening and complaining, and making plenty of jokes. Thanks for always being my incredibly sweet support and for giving advice or just a hug. Magnus and Fabai, two excellent scientists and company, those were some good years with dumplings, music, BBQs, doeraks en kloosters. Mathia, Benjamin, Carsten, Magda and Lisa, what a great joy it was to share the executive office with you guys the last half year! You cheered me up and calmed me down with laughter and always interesting conversations about sometimes quite intimate topics. Julia and Tim, I am very happy to have had you as contemporaries throughout the physics and PhD years, in the TN hallways and outside. The tea-moments, little chats, lunches, hugs, weirdest birthday presents, cakes, sweet notes and messages were these moments that really sparked the day. You are my successful toppers. Fabrizio, thanks for always sending interesting papers, for your kindness and for being honest without being direct. Good luck with your next adventures, and keep enjoying running. Alicia, your enthusiasm and energy are contagious. Essie, thanks for your liveliness and always taking care of us. Sweet Jonas, remember those good moments. Brave Johannes, thanks for introducing me to Bavaria and for your humour, go well. Huong, you were always so kind and funny, thanks for collaborating on the T7 project and not to forget thanks for your delicious loempias. Maurits, you are an interesting character, it was nice to work with you and to discuss over drinks. Sabine, Emma, Katy, Roeland, Duco and Anne, thank you for the fruitful research and personal interactions. Thanks to everyone else that contributed to the good atmosphere in the lab over the past years: Mischa, Elise, Paul, Sanne, Margreet, Alex, Just, Ivo, Maryse, Kasper, Karel, Mashid and Eduardo. Helena and Nicole, you are cool, keep up the combination of good science and fun always. Mo and family, thanks for the lovely get togethers and best of luck wherever you will go. Eugen, you are the most unconventional and funny guy in the department, thanks for so many good conversations. Louis, you rock, curious about your future plans. Tim, Meghan and Lucy, thanks for the quality time in France, we should definitely do that again. Ammeret, you have a gift of bringing people together and talking about the beautiful things in life that is endearing. Jacob, you are so talented and always so modest, thanks for the nice interactions over all these years. Ruben, captain of Real RKC, thanks for the soccer, help with LateX, and good conversations.
Sander, thank you for being my promotor, I appreciated your valuable input at our yearly meetings a lot. Cees, thank you for creating such a lively and dynamic environment in MB and early BN times and for introducing me to Stefan and the world of experiments as a bachelor student. Marileen, it was a great experience to perform an internship in your lab with Olga, I am honoured that you accepted to be in my committee. Bertus, those few 'geef jezelf de ruimte' conversations were vital for me, my appreciation is enormous. Martin and Olya, thanks for sharing that special day with us.
So many other kind people I enjoyed interacting with in BN, at the coffee corner, fussball table, christmas parties, in the good old times: the spectacular Felix and Nienke and Melina, Tessa and Calin, Stefan and Tomas, Francesca, central D Maarten, Francesco, Jan, Kuba, Michela and Ciro, Vlad, Stanley, Adi (good luck with the nups), Stephanie, Mahipal, Jelmer (thanks for sharing Delft-Zoetermeer memories and making fun of David), Bojkie, Cleo, Jetty, Suzanne and Serge, Natalia, Laura D., Bronwen, Yaron, Mariette (best

security lady ever), Regis and Aafke. And fun new times with more of the above: Victoria and Thijs, it was great to share the office with you, breaking-news Daniel, Laura, Luuk, Vanessa, Victor, Pawel, Marek thanks for good discussions, Malwina, Federico, Anthony, Jorine, Nuria, Siddarth, Misha, Yoones, Kim, Eve, Jaco, Theo, Anna, Sergii, Afshin, Jochem, Mehran, Werner and Jelle (your work ethic is inspirational). Dijana, Emmylou, Jolijn, Amanda, Esther and Marije, thanks for all your help and friendly chats. Laurens, Ana and Young Mi, it was very energizing to interact with you and to learn about design thinking and coaching skills, you are inspiring persons. Bas en Luuk, the eye-opener you gave me with the Health Coach Program was a most welcoming and positive change in my life, thanks a lot.
Susan, the time we spent on adventures in the Netherlands past years was great fun and relaxing. Thanks that I can discuss just everything with you, and for then ending up laughing both. Ellinor, those backpacking Swedish hikes were jättebra. You are a beautiful person and have a special gift of making people think differently and more positive, thanks for being awesome. Elisa, I am so looking forward to your visit! Patrizia and Marco, thanks for the good times in Switzerland and in the Netherlands. Julia, Robert, Dane, Wouter en Koen, Swiss family dinners for already six years, 'LEUK'! Dorine en Sietske, the good memories started 10 years ago at the intro-weekend of physics and I am happy that 'eten en spelletjes' turned into 'skype and spelletjes'. I look forward to our next mountainbike adventure, or just rummikub. Guen, which country will we meet next? Liesbeth, I never forget you, thank you for bringing calmness back in my speech, and in life. Thanks to everyone that shared their enthousiasm for running, and for the joy of playing futsal in the teams of Tutor and Mazzelstars. Inge, Liesbeth, Nanda and Emma those were fun years becoming champion, celebrating Sinterklaas and the trip to Antwerpen. Feetje and family, thanks for being warm, welcoming and gezellig for more than 20 (!) years now. Eva, so good to meet old friends again. Also thanks to my friends from chess, even after leaving the scene. My old teammates, thanks for those enjoyable Saturdays. Bianca, Biaina, Jan, Tjerk, Matthew, Anne, Lisa, Jan-Pieter, Martine, we saw each other too little but when it happened I enjoyed those times a lot! Ted, your laugh remains as contagious, it was always good to see you again, thanks for the entertaining and helpful conversations during lunches.
My housemates, Joris, Emma, Angelique en Evie, Bart-Jeroen, Piet and Teun, it felt so good to have a place to come home to and know that you were there for so many good conversations, dinners, coffees, or doing random stuff such as playing hide-and-seek ('because of Evie'). I will miss the great time at the Voorstraat with you guys. Special thanks to Emma for drawing the beautiful cover just like that.
My family, aunts, uncles, cousins, thanks for giving affection and accepting people as they are, and for giving me the feeling of geborgenheid. My sister Jessica, you always paved the way for me, thank you for your care and protection. Now you are a doctor and have the sweetest little family, I wish you all the happiness together with Bart, dancing Thalia and cute Aedan. Dingeman en Brigitte, pp en mm, I could not wish for better parents; your endless love and support always reassures me that everything will be all right.
David, the best thing that happened to me during the PhD was that after those years of close friendship, we eventually both fell in love with each other. Now, after 2.5 years, we finally live together! I am so happy to be with you, you make me laugh everyday.

CURRICULUM VITÆ

Pauline Yvonne Brigitte VAN NIES

31-01-1989 Born in Zoetermeer, the Netherlands.

EDUCATION

2000-2006 Secondary Education
Erasmus College Zoetermeer

2006-2010 B. Sc. Applied Physics
Delft University of Technology

2010-2011 Exchange study
ETH Zürich

2010-2012 M. Sc. Applied Physics
Delft University of Technology

2012-2016 Ph. D. research
Department of Bionanoscience, Delft University of Technology
Thesis: Virus-inspired DNA replication coupled with gene expression in a minimal cell framework
Supervisor: Dr. C.J.A. Danelon
Promotor: Prof. S. Tans

LIST OF PUBLICATIONS

P. van Nies, I. Westerlaken, D. Blanken, M. Mencia, M. Salas and C. Danelon **ds-DNA replication by synthesized proteins of the phi29 virus reconstituted in liposomes**, *in preparation* .

P. van Nies, I. Westerlaken, S. van Schie and C. Danelon **DNA replication via RNA intermediates and transcription-translation of hybrid DNA:RNA templates**, *in preparation.*

P. van Nies, A. Soler Canton, Z. Nourian and C. Danelon, **Monitoring mRNA and Protein Levels in Bulk and in Model Vesicle- Based Artificial Cells**, Methods in Enzymology, Volume 550: 197-214 (2015).

P. van Nies, Z. Nourian, M. Kok, R. van Wijk, J. Moeskops, I. Westerlaken, J. Poolman, R. Eelkema, J. van Esch, Y. Kuruma, T. Ueda and C. Danelon, **Unbiased Tracking of the Progression of mRNA and Protein Synthesis in Bulk and inside Lipid Vesicles**, ChemBioChem, 14, 1963-1966, (2013).

S.Kowalczyk, L. Kapinos, T. Blosser, T. Magalhaes, P. van Nies, R. Lim, C. Dekker, **Single-molecule transport across an individual biomimetic nuclear pore complex**, Nature Nanotechnology, 6, 433–438 (2011)